Research methodology topics: Cohort studies or prospective and retrospective cohort studies

Luís Marcelo Aranha Camargo¹,²,³,⁴,⁵, Romeu Paulo Martins Silva⁶,⁷, Dionatas Ulises de Oliveira Meneguetti⁵,⁷

Abstract

In health sciences, the epidemiological method can be divided into descriptive and analytical epidemiology and the latter being divided into observational (cross-sectional study, case-control study and cohort study) and experimental studies. Cohort studies may be retrospective or prospective, and both assume that the researcher will follow a population over time to seek a possible association between exposure(s) and outcome(s). These types of studies have as advantages the possibility of measuring several exposure factors and outcomes, both primary and secondary, for both relatively frequent outcomes and rare exposure factors. However, they are often long and therefore expensive studies. They have as main biases those of selection, memory and information. These are studies that may point to statistical associations between exposure and outcome that need other models to prove the casualty of these associations.

Keywords: cohort, longitudinal study, follow-up.

The epidemiological method, in the area of Health Sciences, can be divided into two main areas: descriptive and analytical epidemiology. In the first part, indicators of morbidity, mortality, demographic, socioeconomic, quality of access health services and quality of life are used, among others. The second part, on other hand, uses data from descriptive epidemiology to analyze and seek explanations and/or associations for the descriptive data. These (analytical) studies, on other hand, are divided into observational (cross-sectional study, case-control study and cohort study) and experimental studies, best known as clinical trials. They all have different advantages, disadvantages, costs, runtime and accuracy. While observational studies raise more hypothesis rather than showing causal association, experimental studies, considered the gold standard of epidemiological studies, are expensive, time-consuming but more accurate and may show a causal relationship between exposure and outcome.

INTRODUCTION

The epidemiological method, in the area of Health Sciences, can be divided into two main areas: descriptive and analytical epidemiology. In the first part, indicators of morbidity, mortality, demographic, socioeconomic, quality of access health services and quality of life are used, among others. The second part, on other hand, uses data from descriptive epidemiology to analyze and seek explanations and/or associations for the descriptive data. These (analytical) studies, on other hand, are divided into observational (cross-sectional study, case-control study and cohort study) and experimental studies, best known as clinical trials. They all have different advantages, disadvantages, costs, runtime and accuracy. While observational studies raise more hypothesis rather than showing causal association, experimental studies, considered the gold standard of epidemiological studies, are expensive, time-consuming but more accurate and may show a causal relationship between exposure and outcome.

COHORT STUDY

These studies (also called longitudinal or “follow-up”) assume that the researcher will follow a population over time to seek a possible (at least statistical) association between exposure and outcome. Roughly this study can be divided into 2 subtypes: a retrospective cohort study and a prospective cohort study (Figure 1).

In the prospective study the researcher is present at the time of exposure of one or more factors and accompanies for a period of time to observe one or more outcomes. After some time of monitoring (months, years or decades) the researcher can relate the exposure to the outcome(s) using the Relative Risk (RR), which is nothing more than the incidence of the outcome(s) in the cohort. At the end of the study the investigator analyzes the incidence of the outcome in the group of exposed and non-exposed to the risk factor(s) in a contingency table (Table 1). The RR will be calculated using the following formula:

\[
RR = \frac{I_e}{I_u}
\]

This formula, in summary, represents the ratio of the Incidence between exposed (I_e) and Incidence between unexposed (I_u), therefore RR = I_e / I_u.

The calculation of the sample size for a cohort study is given by the following formula:

\[
n = \frac{N \cdot Z^2 \cdot p \cdot (1 - p)}{Z^2 \cdot p \cdot (1 - p) + e^2 \cdot (N - 1)}
\]

(n - Calculated sample; N - Population; Z - Normal standardized variable associated with the confidence level; p - True probability of the event; e - Sample error)

Figure 1: Prospective and retrospective cohort study.

---

Authors summary

Why was this study done? The present study was conducted to show the advantages and biases of a prospective and retrospective cohort study, demonstrating its application and in which situations it is indicated.

What did the researchers do and find? It was found that the cohort study may be retrospective and prospective. In the retrospective study the researcher collects previous information on exposure factors and over time the in individuals. In the prospective study, the researcher is present at the exposure of one or more factors and followed for a period of time to observe one or more outcomes.

What do these findings mean? That the cohort study, even with some biases, is a method that can estimate the incidence of an outcome (or more) exposed to one or more factors, and verifies whether there is a statistical association between exposure and outcome, either primary and secondary.
As an example of a prospective cohort study, sets up the following situation: a cohort of 3,400 people followed for 10 years, of whom 1,600 were exposed to a chemical and another 1,800 who were not exposed. As the outcome, the incidence of cases of myeloid leukemia was measured. In the exposed group there were 100 cases and in the group without product exposure (25) cases in 10 years (prospective cohort study). In order to calculate the RR using the formula above, we elaborate the one of contingency below (Table 2).

<table>
<thead>
<tr>
<th>Exposure / Outcome Factor</th>
<th>Leukemia +</th>
<th>Leukemia -</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Product Exposure +</td>
<td>100</td>
<td>1,500</td>
<td>1,600</td>
</tr>
<tr>
<td>Product Exposure –</td>
<td>25</td>
<td>1,775</td>
<td>1,800</td>
</tr>
<tr>
<td>Total</td>
<td>125</td>
<td>3,275</td>
<td>3,400</td>
</tr>
</tbody>
</table>

Using the formula for calculating RR above:

\[
\text{RR} = \frac{AD}{BC} = 4.5 \quad (CI \ 95\% \ 2.918-6.939)
\]

* Interval Confidence 95\% (95\% IC) and p <0.05 (statistically significant).

With this data we can affirm that the occurrence of new cases of leukemia in the group exposed to the chemical has an incidence (Ie) of 62.55 new cases / 1,000 people, which in the group not exposed to incidence (Iu) is 14.1 /1,000 people and the statistically significant risk of leukemia in the exposed group is 4.5 higher than in the non-exposed group (RR) with high statistical significance (p = 0.001). Since p <0.05 and number 1 is not in the confidence interval, I accept the hypothesis of statistical association between outcome exposition.

In addition to the calculation of RR and incidence between exposed and non-exposed, this study model allows to calculate (Table 3).

<table>
<thead>
<tr>
<th>Total Cumulative Incidence (TCI)</th>
<th>TCI = [(a+c)/a+b+c+d]. 10^n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attributable Risk (AR)</td>
<td>AR = I_e - I_o for 10^n</td>
</tr>
<tr>
<td>Percentage of Attributable Risk (PAR)</td>
<td>PAR = \frac{(RR-1)}{RR} \times 100</td>
</tr>
<tr>
<td>Risk Attributable to the Population (RAP)</td>
<td>RAP = (IAT- I_o) \times 10^n</td>
</tr>
<tr>
<td>Risk Attributable to the Population in % RAP%</td>
<td>RAP% = \frac{(IAT- I_o)}{IAT} \times 100</td>
</tr>
</tbody>
</table>
CONCLUSÃO

Cohort studies present as advantages the ability of measuring several exposure factors and outcomes (primary and secondary). They can be applied to relatively frequent outcomes and rare exposure factors. In general, these studies are relatively expensive due to the cohort follow-up time. The errors of information and memory are the major biases of these studies. Often, in the presence of a statistical association between exposure and outcome, the Clinical Trial (experimental) is sought to confirm the hypothesis.

PALAVRAS-CHAVE: coorte/cohorte, estudo longitudinais, follow-up.

REFERENCES


Resumo

Na área de ciências da saúde, o método epidemiológico, pode ser dividido em epidemiologia descritiva e a analítica, essa última se divide em observacional (estudo de corte transversal, estudo caso-controle e estudo de coorte/cohorte) e experimentais. Os estudos de coorte ou cohorte, podem ser retrospectivos ou prospectivos, e ambos partem do pressuposto que o pesquisador irá acompanhar uma população ao longo do tempo para buscar possivel associação entre exposição e desfecho. Esses tipos de estudos apresentam como vantagens a possibilidade de se mensurar vários fatores de exposição e desfechos, tanto primários como secundários, aplicam-se tanto para desfechos relativamente frequentes e fatores de exposição raros. Porém, muitas vezes são estudos prolongados e, portanto, caros. Têm como principais viéses os de seleção, memória e informação. São estudos que podem apontar para associações estatísticas entre exposição e desfecho que necessitam de outros modelos para se comprovar há casualidade destas associações.

Palavras-chave: coorte/cohorte, estudo longitudinais, follow-up.