Residência Pediátrica
Publicação Oficial da Sociedade Brasileira de Pediatria

INTRODUCTION TO SCIENTIFIC METHODOLOGY

Measures of association: Which to choose and how to interpret

Márcia Alves Galvão[1], Marilene Augusta Crispino Santos[1]

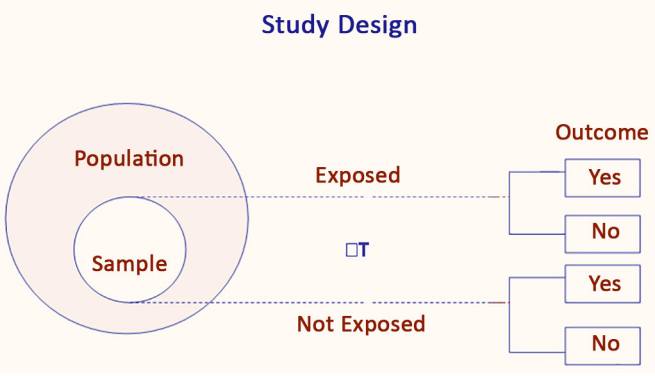
When conducting a scientific study, we often want to know whether a certain factor (e.g., medicine, vaccine, environmental factor) is related to a particular condition (e.g., symptom relief, occurrence of disease).

A number of measures to assess and quantify the association between exposure to a factor and the occurrence of an outcome can be used. One of the most frequently used measures in scientific articles is the relative risk (RR).

This article aims to briefly describe the definition and uses of RR.

RELATIVE RISK

To evaluate the effectiveness of a drug, a vaccine, or a healthcare service, a study design can be used that analyzes individuals who comprise the sample according to their exposure to the factor of interest and the development of a particular outcome:



This is known as a cohort study. For conducting such a study, it is necessary to select a group of individuals (a cohort) in which none of them has experienced the outcome of interest, but they are all are able to do so. Thus, for example, if the study aims to assess whether radiation exposure in the first year of life is a risk factor for the development of gynecologic cancer, the cohort should comprise only girls.

Knowing this information, how can the responses to the exposures be measured?

One way to do so is by assessing whether a particular exposure (medication, vaccine, environmental factor) decreases or increases the risk of a particular outcome. For that, the relative risk (RR) is used.

For better understanding of this measure, some definitions should be considered:

- Risk: proportion of participants experiencing an event of interest.
- Incidence: the number of new cases or outcomes in an established population over a given period of time.

The RR is the ratio of risk in both groups: intervention and control. It is defined as the ratio of the risk of the occurrence of the outcome in the group that received the intervention (medication, vaccine, environmental factor) and risk of the occurrence of the same outcome in the control group, which received no intervention.

The RR is therefore a ratio of incidence: the ratio between the incidence of the outcome among the exposed and the incidence of the outcome among the unexposed.

[1] Associated Editors of Residencia Pediátrica.

Correspondence to:
Marilene Augusta Crispino Santos.
E-mail: crispinomarilene06@gmail.com

Residência Pediátrica 2016;6(2):59-62.

The fundamental question is, how many times is it more likely for exposed people to become ill in comparison with those who were not exposed?

Table 1. Relative risk calculation

	With outcome	No outcome	Total
Exposed	a	b	a+b
Untreated	c	d	c+d
Total	a+c	b+d	a+b+c+d

Calculation of RR:

RR: incidence among those who were exposed (a/a+b)
incidence among those who were not exposed (c/c+d)

Interpretation:

= 1: the exposure does not interfere in the development of outcome

> 1: the exposure is a risk factor for the development of outcome

< 1: the exposure is a protective factor against the development of outcome.

In a simplified manner, it is considered that the association is:

Strong: RR > 3

Moderate: RR around 2

Low: RR = 1.5

Nonexistent: RR = 1

Confidence interval:

When reading studies that assess a risk or evaluate the efficacy of a treatment, it can be observed that RR is followed by a confidence interval (CI), commonly 95%. What is the CI?

When a study evaluates the entire population of interest, after careful removal of biases, the calculated RR is precise. However, this is not feasible in most situations. Researchers work with population samples. When selecting a sample, it is possible to find, at random, a result that is different from what happens in the entire population.

At first, the solution could be to repeat the same study several times, using different samples of the same population, until the accuracy of results could be ensured. To answer this question, confidence intervals, which are a measure of uncertainty of the statistical results found, can be used.

The 95% CI means that, if the study were repeated on other samples of the same population, 95% of the times the CI of these studies would include the true value of RR. Alternatively, 90% or 99% CI is sometimes used. Larger CIs indicate lower accuracy; narrower ranges, higher accuracy.

Interpretation:

CI includes 1: exposure does not interfere in the development of outcome.

CI does not include 1: there is no statistically significant difference between the treatment and control groups.

- If the CI is all > 1: the exposure is a risk factor
- If the CI is all < 1: the exposure is protection factor

MOVING TO THE PRACTICAL PART:

The first step for epidemiological calculations is to draw a 2×2 table as follows:

	Outcome YES	Outcome NO	
Exposure YES	A	B	A+B
Exposure NO	C	D	C+D
	A+C	B+D	A+B+C+D

Thus, the cells will represent the following:

- A: the number of the subjects who have been exposed and present the outcome
- B: the number of individuals who were exposed but do not present the outcome
- C: the number of individuals who were not exposed but present the outcome
- D: the number of individuals who were not exposed and do not present the outcome

Read the hypothetical example:

Imagine a group of newborns (NB) followed up forpostnatal death notice where there were 1,000 infants with low birth weight and 1,000 infants with normal birth weight.

The first step will be to complete the part of the 2×2 table with the figures given in the example, knowing that the outcome is "Death" and the exposure is "Birth Weight".

The total sample consists of 2,000 infants, divided into two subgroups of 1,000 classified according to the adequacy of their birth weight.

	Death SIM	Death NÃO	
Low birth weight YES	A	B	1000
Low birth weight NO	C	D	1000

At the end of the study, it was observed that 300 infants with low birth weight and 100 newborns with normal birth weight at birth died.

	Death YES	Death NO	
Low birth weight YES	300	B	1000
Low birth weight NO	100	D	1000
	(A+C) 400	B+D	2000

We now have enough information to complete the 2×2 table:

	Death YES	Death NO	
Low birth weight YES	300	(1000-300) 700	1000
Low birth weight NO	100	(1000-100) 900	1000
	(A+C) 400	(B+D) 1600	2000

RR = A/(A+B)
C/(C+D)

Applying the formula above, we get:

RR = 300/1000
100/1000

RR = 3

If the results found were RR = 1, then there would be no difference between infants with low birth weight and those with adequate weight, in relation to the risk of death after birth.

Conversely, RR < 1 would lead to the assumption that the risk of death would be lower among those with low birth weight.

The RR value of 3 found in the example indicates that the risk of death is three times higher among those with low birth weight compared to those with normal weight.

We shall now interpret the 95% CI:

Note the following 95% CI and the respective mean RR value. Answer the following questions:
RR = 10 (95% CI: 4.3-25)

I) What is the interpretation of the 95% CI value?
II) Was this result statistically significant?
III) What can be concluded about the accuracy of RR?

Answers:

I) This 95% confidence interval (95% CI) means that the result will be within that range in 95 out of 100 hypothetical studies performed, *i.e.*, the RR is between 4.3 and 25. Thus, a correct reading of the confidence interval is to state that *in 95 out of 100 hypothetical samples, the result will be within this range*.
II) This result was significant ($p < 0.05$), as the value 1 was not included between the upper and lower limits of the CI.
III) The result of the RR shows an IC with a considerable width-about 20 units-which indicates low accuracy.

Now note the following 95% CI and the respective mean RR value. Answer the following questions:
RR = 0.5 (95% CI: 0.4-1.1)

I) What is the interpretation of the 95% CI?
II) Was this result significant from a statistical point of view?

Answers:

I) This 95% confidence interval (95% CI) means that the RR will be within the range of 0.4 to 1.1 in 95 out of 100 hypothetical studies performed.

II) This result was not significant ($p > 0.05$), as the value 1 was included between the upper and lower limits of the CI.

IMPORTANT NOTE

Risk estimates can only be made when we start from an exposure and observe the event (outcome).