

DOI: <https://doi.org/10.5554/22562087.e1076>

# Epidemiological measures: application and interpretation in real-life scenarios according to epidemiological study designs

## *Medidas de frecuencia, asociación e impacto: aplicación e interpretación en escenarios reales según el diseño de investigación*

Paula Angélica Suárez-Álvarez<sup>a</sup> , Claudia Marcela Granados-Rugeles<sup>a</sup> ,  
Carlos Javier Rincón-Rodríguez<sup>a</sup> 

<sup>a</sup> Clinical Epidemiology and Biostatistics Department, Medical School, Pontificia Universidad Javeriana. Bogotá, Colombia.

**Correspondence:** Departamento de Epidemiología Clínica y Bioestadística, Facultad de Medicina, Pontificia Universidad Javeriana. Carrera 7 No. 40-62. Bogotá, Colombia. **Email:** [suarezpaula@javeriana.edu.co](mailto:suarezpaula@javeriana.edu.co)

**How to cite this article:** Suárez-Álvarez PA, Granados-Rugeles CM, Rincón-Rodríguez CJ. Epidemiological measures: application and interpretation in real-life scenarios according to epidemiological study designs. Colombian Journal of Anesthesiology. 2023;51:e1076.

### Abstract

Frequency, association and impact measures are key concepts in clinical epidemiology; however, it has been found that a considerable proportion of health students and professionals have no knowledge of how to use or interpret them when reading a scientific paper or conducting research. This article aims to explain the main epidemiological measures, how they are used, derived and interpreted. They are approached from the perspective of each of the most frequently used types of primary quantitative research studies (randomized clinical trials, cohort studies, case-control studies and cross-sectional studies) in order to provide the reader with the context in which they are used. Moreover, the process for calculating and interpreting each result in a real setting is explained using clinical examples for a better understanding of these concepts and in order to prevent their use from becoming just a mechanical or repetitive exercise.

### Key words

Epidemiological studies; Measures of frequency; Measures of association; Exposure; Risk or outcome; Medical education; Anesthesiology.

### Resumen

Las medidas de frecuencia, asociación e impacto son conceptos fundamentales de la epidemiología clínica; sin embargo, se ha encontrado que una parte considerable de los estudiantes y de los profesionales en el área de la salud no sabe cómo usarlas ni cómo interpretarlas al leer un texto científico o al hacer una investigación. Este artículo busca explicar las principales medidas epidemiológicas, cuándo se usan, cómo se obtienen y cómo se interpretan. Se abordan desde cada tipo de estudios primarios más frecuentemente utilizados cuando se realizan investigaciones cuantitativas (ensayos clínicos aleatorizados, estudios de cohorte, casos y controles y estudios de corte transversal), con el fin de darle al lector el contexto en el cual se usan. Además, mediante ejemplos clínicos, se explica el proceso para calcular e interpretar cada resultado en un escenario real, con el fin de lograr una mayor comprensión de estos conceptos y de que su uso no sea un ejercicio mecánico o de repetición.

### Palabras clave

Estudios epidemiológicos; Medidas de ocurrencia de enfermedades; Medidas de asociación; Exposición; Riesgo o desenlace; Educación en salud; Anestesiología.

Lea la versión en español de este artículo en [www.revcolanest.com.co](http://www.revcolanest.com.co)

Copyright © 2023 Sociedad Colombiana de Anestesiología y Reanimación (S.C.A.R.E.).

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## INTRODUCTION

Clinical epidemiology aims to gain insights into the distribution of health outcomes and identify associated factors in a given population (1). Although epidemiological measures are widely known, it has been found that healthcare professionals misuse them, use inadequate terms or misinterpret them out of lack of knowledge (2-5).

This article explains the main measures applied in epidemiology, organized in terms of frequency, association and impact, by epidemiological design, from the perspective of a positivist paradigm. The aim is to improve their understanding, linking the originating mathematical expressions with their interpretation. The types of epidemiological studies will be first explained, followed by the measures used for each design which will be presented with step-by-step examples of how the equation is developed, providing the reader with the tools for adequate interpretation. Instead of mechanical formulas becoming a mechanical action, the idea is that their application should be guided by an understanding of the subject and of what they represent. This article is designed to be a teaching tool for professionals and students of different academic ranks that will help them apply their knowledge of epidemiology and biostatistics when reading a scientific paper or planning a research study.

## EPIDEMIOLOGICAL STUDIES

Study designs depend on factors such as objectives, limitations, and available resources. In the quantitative realm, there are two large groups of primary studies: experimental and non-experimental or observational, each with their own subtypes (6).

Experimental studies include laboratory studies, community interventions and randomized clinical trials (RCT). RCTs are conducted under controlled conditions in which subjects are randomly assigned to the study intervention, or to a control group in which they receive either the

standard of care or placebo. The purpose of randomization is to ensure that the groups are very similar at the baseline so that their characteristics will not become confounding factors when the outcomes for both groups are compared. Subjects are followed for a period of time, after which their response to management is assessed and outcomes are compared; differences between the groups suggest either a beneficial or adverse effect from the new treatment. To prevent outcomes from being influenced by information on what subjects are receiving, studies should ideally be masked so that neither the investigator nor the participants know what the subjects were assigned to. In RCTs, generalization of some of the findings could be affected by participant selection; if the characteristics of the subjects are not similar to those of the target population and their condition is modified by the efficacy of the treatment, the study conclusions will only apply to the observed population (6,7). On the other hand, there are studies similar to the experimental studies, called quasi-experimental, the main difference being that they do not use randomization to assign the participants to a specific study factor. Policy studies and natural experiments, among others, are part of this group (6).

In turn, non-experimental or observational studies can be descriptive or analytical. Descriptive studies are selected when there is little information about the occurrence of the health outcome or the natural history of the disease, and they are designed to estimate the extent of occurrence and generate etiologic hypotheses. Analytical studies are performed when there is enough knowledge about the disease and the aim is to test specific etiologic hypotheses, identifying risk factors or estimating the effects of exposure in order to suggest potential interventions. In these studies, given that participants are not randomly selected, there is a higher risk of initial imbalance in terms of the characteristics of the exposed and non-exposed subjects. This may result in potential differences attributable to baseline characteristics but

not to the exposure (6,7). These designs require the use of different strategies to minimize this potential error; however, it is important to bear in mind that this does not apply in all scenarios, although it is specifically true when assessing interventions, particularly pharmacological interventions.

The main subtypes are cohort, case-control (analytical) and cross-sectional (descriptive) studies. Regarding cohort studies, the investigator selects a group of people in which the outcome has not occurred and then assigns them either to an exposure group or index cohort and or a non-exposed group or reference cohort. The investigator identifies the outcomes that occur during a follow-up period in each cohort and makes a comparison to assess the exposure effect. These are forward-direction studies, which means that they start from the presence or absence of the exposure until the outcome occurs. In case-control studies, cases are defined based on the presence of the disease or outcome, and controls are defined on the basis of its absence. Exposure distribution is observed in both groups in order to determine a difference that can point to a relationship. These are a backward-direction studies where the researcher begins by determining whether the exposure is present or absent. Finally, cross-sectional studies consider the entire target population or a representative sample thereof at a specific point in time, and it is selected regardless of whether the subjects have been exposed or not, or whether a disease is present or not. They are designed to assess the state of exposure or disease within a specific period of time, and compare the occurrence of the event of interest between groups of different exposures in order to determine potential associations or generate hypotheses. Unlike the previous studies, these have no directionality (1,6,7).

## MEASURES IN EPIDEMIOLOGY

There are three types of measures that serve different purposes. First, frequency

measures assess the magnitude of occurrence of an outcome or study factor; association measures, on the other hand, assess potential changes in outcome occurrence among the varying degrees of exposure to the study factor. Finally, impact measures which indicate to what extent the occurrence of the event is attributable to the study factor. In epidemiological designs, when the outcome of interest and the study factor are dichotomous, they can be summarized in a 2 x 2 table which includes, in each cell, the number of subjects classified by exposure and outcome occurrence (Table 1). Therefore,  $a$  is the number of subjects with exposure to the study factor in whom the outcome is present,  $b$  is the number of exposed subjects who do not present the outcome,  $c$  is the number of individuals without exposure in whom the outcome is present, and  $d$  is the number of individuals without exposure who do not show the outcome (1).

The frequency, association and impact measures are described below by type of epidemiological study, including their calculation based on Table 1 and their interpretation based on real studies. The focus of this article is to describe point estimates of each measure and their interpretation; however, it is important to remember that these must be accompanied by their respective confidence intervals which consider the random error associated with the selection of the sample participating in the study.

## MEASURES BY EPIDEMIOLOGICAL DESIGN

### RCTs and cohort studies

The measures described below are obtained in cohort studies and RCTs. In terms of frequency measures, incidence is expressed as a risk or probability measure (cumulative incidence -  $CI$ ), or as rate or instant change measure (incidence density -  $ID$ ). The former refers to the number of people who show the outcome of interest during a specific

**Table 1.** Number of subjects classified by study factor and outcome. General notation in a 2 x 2 table.

|                  | Outcome (D) | No outcome (ND) | Total   |
|------------------|-------------|-----------------|---------|
| Exposed (E)      | $a$         | $b$             | $a + b$ |
| Non-exposed (NE) | $c$         | $d$             | $c + d$ |
| Total            | $a + c$     | $b + d$         | $n$     |

Source: Authors.

**Table 2.** Presence of COVID-19 symptoms after 21 days of ivermectin vs. placebo.

|                | With symptoms (D) | Without symptoms (ND) | Total |
|----------------|-------------------|-----------------------|-------|
| Ivermectin (E) | 36                | 164                   | 200   |
| Placebo (NE)   | 44                | 156                   | 200   |
| Total          | 80                | 320                   | 400   |

Source: Authors from (8).

time period or a specific geographical area determined in a population that is initially free of the event and has a risk of presenting it. It is estimated as  $CI = (a + c) / n$  and indicates the odds of the outcome occurring during the observation time defined in the study. Cumulative incidence can be obtained in both exposed and non-exposed subjects by replacing  $CI_E = a / (a + b)$  and  $CI_{NE} = c / (c + d)$ , respectively. Incidence density measures the occurrence of an event in a specified population in relation to time-person units (TP) at risk of the risk presenting itself; it is an expression of the "speed" of occurrence of the disease. It is obtained by dividing the number of outcomes in the numerator by the sum of the time during which the subjects were under observation, free of the disease, i.e.,  $ID = (a + c) / \sum_{i=1}^n TP_i$ . Incidence density can be derived both for the exposed group as well as for the non-exposed, replacing  $ID_E = a / \sum_{i=1}^{a+b} TP_i'$  and  $ID_{NE} = c / \sum_{i=1}^{c+d} TP_i''$ , respectively (1).

### Cumulative incidence

One RCT assessed the effect of ivermectin in adult patients with mild COVID-19 vs.

placebo in the presence of symptoms on day 21 (8) (Table 2).

$CI = 80/400 = 0.2$  which means that the risk of having symptoms on day 21 is 0.2, or that 20% have symptoms on follow-up day 21.  $CI_E = \frac{36}{200} = 0.18$  and  $CI_{NE} = \frac{44}{200} = 0.22$ , indicate that the risk of having symptoms in those who received ivermectin for 21 days was 0.18 and 0.22 in those who received placebo.

### Incidence density

A study conducted in Lombardy collected cases of upper respiratory infection (URI) in children under 6 years of age to assess the incidence of this condition (9); 111 children between 1 and 2 years of age and 49 children between 5 and 6 were followed during a four-month period before the pandemic (November 2019 to February 2020), recording the number of URI cases and the follow-up time contributed by each subject (Table 3).

$ID = \frac{236}{160 \times 4} \times 100 = 36.8$  cases for every 100 exposed children per month. Taking age groups as exposure and no exposure,

**Table 3.** URI incidence in children before the COVID-19 pandemic in Lombardy, Italy.

|                   | URI cases<br>(D) | Exposure Time<br>(ET) |
|-------------------|------------------|-----------------------|
| 1 to 2 years (E)  | 183              | 111 × 4 months        |
| 5 to 6 years (NE) | 53               | 49 × 4 months         |
| Total             | 236              | 160 × 4 months        |

Source: Authors from (9).

$ID_E = \left(\frac{183}{111 \times 4}\right) \times 100 = 41.2$  exposed cases for every 100 children 1-2 years of age, and  $ID_{NE} = \left(\frac{53}{49 \times 4}\right) \times 100 = 27$  non-exposed cases for every 100 children 5-6 years of age.

Association measures are obtained from a comparison of frequency measures between exposed and non-exposed subjects, by means of quotient or subtraction. Based on cumulative incidences, relative risk (RR) is expressed as follows:  $RR = [a/(a+b)] / [c/(c+d)]$  and indicates that the risk of the outcome being present in the exposed group is that number of times (higher or lower) than the risk of the non-exposed group. The absolute risk difference (ARD) is expressed as follows:  $ARD = [a/(a+b)] - [c/(c+d)]$  and indicates risk increase or decrease in absolute terms in the exposed group, as compared to the non-exposed. If  $RR > 1$  or  $ARD > 0$ , the risk of presenting the outcome is greater in the exposed vs. the non-exposed group, and if  $RR < 1$  or  $ARD < 0$ , the risk of the outcome being present is greater in the non-exposed vs. the exposed group. Similarly, based on incidence densities in the exposed and non-exposed groups, there is an incidence density ratio (IDR) or incidence density difference (IDD) whose interpretations are equivalent to those mentioned previously, but this time in reference to the rate or instantaneous risk of presenting the outcome (1).

In the ivermectin study for COVID-19, the relative risk is equal to  $RR = 0.18/0.22 = 0.82$ , i.e., the risk of those who received ivermectin having symptoms on day 21 is a fraction of 0.82 times the risk of symptoms being present in the placebo group. The  $ARD = 0.18 -$

$0.22 = -0.04$ , i.e., a negative 4% difference that indicates that this is the magnitude by which the risk of symptoms being present in the placebo group is higher.

When the outcome of interest is an unfavorable event such as death or adverse events, impact measures indicate how much of the event is "attributable" or "preventable" by the exposure. If  $RR > 1$ , the attributable fraction in the exposed group ( $AF_E$ ) is  $AF_E = (CI_E - CI_{NE}) / CI_E = (RR - 1) / RR$  and represents the proportion of unfavorable outcomes in the exposed group that are attributable to this exposure. The attributable fraction for the population ( $AF_p$ ), which is  $AF_p = (CI - CI_{NE}) / CI$ , is also obtained and represents the percentage of unfavorable outcomes attributable to exposure in the population. Finally, the number needed to harm (NNH) which is  $NNH = 1 / (CI_E - CI_{NE}) = 1 / ARD$  is also obtained and represents the number of subjects to expose in order to cause an additional outcome vs. the non-exposed. If  $RR < 1$ , the impact measure is called preventable fraction in the exposed ( $PF_E$ ), which is  $PF_E = (CI_{NE} - CI_E) / CI_{NE} = 1 - RR$  and expresses the percentage of potential unfavorable outcomes prevented in the exposed group as a result of the exposure. The preventable fraction for the population ( $PF_p$ ), which is  $PF_p = (IA_{NE} - IA) / IA_{NE}$  expresses the proportion of potential unfavorable outcomes prevented in the population as a result of the exposure. Finally, the number needed to treat (NNT), which is  $NNT = 1 / (CI_{NE} - CI_E)$ , represents the number of subjects that

need to be exposed in order to avoid one outcome as compared to the non-exposed (1,6).

Therefore, in the ivermectin study  $RR < 1$ , then  $PF_p = (0.22 - 0.2) / 0.22 = 0.09$ , which means that 19% of potential patients with symptoms on day 21 were prevented in the exposed group due to the medication.  $PF_p = \frac{0.2 - 0.18}{0.2} = 0.1$ , which means that 9.09% of the potential patients with symptoms on day 21 were prevented in the population due to the medication. Finally,  $NNT = 1 / (0.22 - 0.18) = 25$ , indicating that 25 subjects must be treated with ivermectin to prevent one additional case of a patient with symptoms on day 21, compared to placebo.

### Case-control studies

In this type of study, the number of cases and controls is determined by the researchers and, for that reason, no frequency measures for the outcome can be obtained. Case-control studies allow to determine exposure frequency in the cases and compare it with exposure frequency in the controls. Based on this comparison, the association measure is the odds ratio or OR (1).

The exposure odds for cases and controls are required in order to obtain the OR. In general, the odds represent the ratio between the probability of an event happening, divided by the probability that it will not occur. The odds of exposure in the cases are  $O_E(\text{casos}) = (a / (a+c)) / [1 - (a / (a+c))] = a/c$  and the odds of exposure in the control are  $O_E(\text{controles}) = (b / (b+d)) / [1 - (b / (b+d))] = b/d$ . The ratio between these two odds is the OR, i.e.,  $OR = (a/c) / (b/d) = (axd) / (cxb)$  (6,7,10).

The OR takes values between zero and infinite and is interpreted as the number of times that the odds of being exposed occur in cases in relation to the odds of controls being exposed.  $OR < 1$  indicates that exposure is associated with a reduction of the risk of occurrence of the event of interest,  $OR = 1$  indicates that there is no association between exposure and the risk

of presenting the event, and  $OR > 1$  means that exposure increases the risk of the event occurring (1).

The effectiveness of vaccination for the SARS-CoV-2 delta variant was assessed in the United States (11) (Table 4). Cases were people with the delta variant infection and controls were individuals with a negative SARS-CoV-2 test. Exposure was having two doses of the Moderna mRNA-1273 vaccine.

Of the 2,027 infected cases, 232 (11.4%) had two doses of the vaccine, and of the 10,135 controls, 4,588 (45.3%) had received the two doses.  $OR = (232/1,795) / (4,588/5,547) = 0.13/0.83 = 0.16$ , and it means that the odds of being vaccinated in the group of infected individuals are a fraction equal to 0.16 of the odds of being vaccinated in the group of non-infected individuals.

The odds of exposure in cases are 0.13 and 0.83 in controls, which means that, in cases, for every 100 non-vaccinated individuals, there are 13 who are vaccinated, while in the controls, for every 100 non-vaccinated individuals, there are 83 who are. This indicates that vaccination with the Moderna mRNA-1273 biologic protects against infection by the SARS-CoV-2 delta variant.

OR can be found in the literature as indirect relative risk, given that it is close to the IDR or the RR obtained from a cohort study or from a RCT if the assumption is that the study population is stable and that the control sample is a representation of the population at risk of presenting the outcome. For the case of RR, the outcome cumulative incidence must be low ( $CI < 0,1$ ). If  $CI > 0,1$ , as the incidence of the outcome of interest increases, the OR overestimates RR (if  $RR > 1$ ), or underestimates RR (if  $RR < 1$ ) (1,6,7,10).

### Cross-sectional studies

Frequency measures are point prevalence and period prevalence, and are interpreted as the probability of an individual having the outcome at a specific point in time or period. Overall

**Table 4.** Effectiveness of two doses of the mRNA-1273 vaccine against infection by the SARS-CoV-2 delta variant.

|                  | Cases (D) | Controls (ND) |
|------------------|-----------|---------------|
| Exposed (E)      | 232       | 4,588         |
| Non-exposed (NE) | 1,795     | 5,547         |

Source: Authors from (11).

prevalence ( $Prev$ ) is equal to the number of subjects who presented the outcome over the total number of subjects, i.e.,  $Prev = (a + c) / n$ . Prevalences for the exposed group ( $Prev_E$ ) and for the non-exposed group ( $Prev_{NE}$ ) can also be obtained, as follows:  $Prev_E = a / (a + b)$  and  $Prev_{NE} = c / (c + d)$  (1,6).

Another way to express outcome frequency is by means of prevalence odds ( $PO$ ), which represents the times an outcome occurs for every instance in which it did not occur, i.e.,  $PO = (a + c) / (b + d)$ . Similarly, prevalence odds can be derived in the exposed ( $PO_E$ ) and the non-exposed ( $PO_{NE}$ ) using the following expressions:  $PO_E = a / b$  and  $PO_{NE} = c / d$  (1,6).

Association measures are derived from outcome prevalence comparison between exposed and non-exposed, either by subtraction or quotient. First, the prevalence difference ( $PD$ ) is  $PD = [a / (a + b)] - [c / (c + d)]$  and is interpreted as the difference in the possibility of the outcome being present in the exposed vs. the non-exposed; and second, the prevalence ratio ( $PR$ ) is  $PR = [a / (a + b)] / [c / (c + d)]$  and is interpreted as the relationship between the possibility of the outcome being present in the exposed vs. the same possibility in the non-exposed. Another association measure can also be built from the prevalence odds ratio ( $POR$ ) between exposed and non-exposed, i.e.,  $POR = (axd) / (bxc)$  and is interpreted as the relationship of prevalence odds in the exposed vs. the prevalence odds in the non-exposed (1,6).

In the COVID-19 vaccination report entered in OurWorldData (12), 4,451,584,807 people completed the initial scheme out of a total world population of 7,874,965,730, that is to say that 56.5% ( $Prev = (4,451,584,807 / 7,874,965,730) \times 100$ ) of the population completed the vaccination scheme. This result, expressed as odds, indicates that for every individual who had not completed the scheme, 1.3 ( $PO = 0.565 / (1 - 0.565)$ ) people had already completed it. Table 5 shows the number of vaccinated individuals and the number of inhabitants in two regions of the world: the countries of the European Union (EU) and the African continent (AC).

Of the EU population, 73 % ( $Prev_{EU} = 324,864,451 / 447,189,915$ ) had completed the scheme, i.e., for every individual who had not completed the scheme, 2.66 ( $PO_{EU} = 0.73 / (1 - 0.73)$ ) had already done so. In the AC, only 14% ( $Prev_{AC} = 195,379,026 / 1,373,486,472$ ) had completed the vaccination scheme, which indicates that for every 100 non-vaccinated individuals, only 17 were vaccinated ( $PO_{AC} = 0.14 / (1 - 0.14)$ ).

The prevalence of vaccinated individuals increases by 58% ( $PD = 0.73 - 0.14$ ) in the EU when compared to the AC, and the prevalence of vaccinated individuals in the EU is 5.11 ( $PR = 0.73 / 0.14$ ) times the prevalence of vaccinated individuals in the AC. Finally, the odds ratio between the EU and the AF indicates that the odds of vaccinated individuals in the EU is 16.01 ( $POR = 2.66 / 0.17$ ) times the odds of vaccinated individuals in the AC.

**Table 5.** Complete initial COVID-19 vaccination scheme in EU and AC countries.

| Regions                | Complete scheme |               |               |
|------------------------|-----------------|---------------|---------------|
|                        | Yes             | No            | Population    |
| European Union (EU)    | 324,864,451     | 122,325,464   | 447,189,915   |
| African Continent (AC) | 195,379,026     | 1,178,107,446 | 1,373,486,472 |

**Source:** Our World in Data. Cut-off date: March 10, 2022.

## CONCLUSIONS

This article describes basic concepts in epidemiology such as study types and their frequency, association and impact measures. It highlights the importance of understanding study designs as the basis for the adequate interpretation of epidemiological measures and for correct decision-making in clinical practice. Given that mathematical expressions are the basis for deriving the measures and understanding what they each mean, it is important to know how to apply them without risking initial rejection on the grounds of the nomenclature used. This terminology is important and concept-unifying, and hence the imperative need to understand it and use it correctly. Understanding frequency, association and impact measures, their aim and their interpretation is essential to avoid mechanical use of the formulas, thus strengthening a clearer understanding of the subject.

## ACKNOWLEDGEMENTS

### Author's contribution

**PASA, CMGR, CJRR:** conceptualization, literature review, writing, drafting and editing of the manuscript.

### Conflict of interest

The authors declare having no conflict of interest.

### Funding

The authors declare having no funding.

## REFERENCES

- Ruiz Morales A, Gómez Restrepo C. Medidas de frecuencia, de asociación y de impacto. En: *Epidemiología Clínica*. 2nd ed. Bogotá: Editorial Médica Panamericana; 2015. pp. 113-128.
- Windish D, Huot S, Green M. Medicine residents' understanding of the biostatistics and results in the medical literature. *JAMA*. 2007;298(9):1010. doi: <https://doi.org/10.1001/jama.298.9.1010>
- Fardales-Macías V, Fábregas-Tejeda J, Carrazana-Rodríguez R. Insuficiencias en la preparación estadística del estudiante de medicina. *Medisur*. 2017;15(4).
- Harris A, Reeder R, Hyun J. Survey of editors and reviewers of high-impact psychology journals: Statistical and research design problems in submitted manuscripts. *J Psychol*. 2011;145(3):195-209. doi: <https://doi.org/10.1080/00223980.2011.555431>
- Susarla S, Redett R. Plastic surgery residents' attitudes and understanding of biostatistics: A pilot study. *J Surg Educat*. 2014;71(4):574-9. doi: <https://doi.org/10.1016/j.jsurg.2013.12.003>
- Kleinbaum D, Kupper L, Morgenstern H. *Epidemiologic research. Principles and quantitative methods*. New York: Van Nostrand Reinhold; 1996. Chapters 3, 6-9, p. 40-50, 96-180.
- Rothman K, Greenland S, Lash T. *Modern epidemiology*. 4th ed. Philadelphia: Lippincott-Raven Publishers; 2021. Chapters 4-8, p. 53-182.
- López-Medina E, López P, Hurtado IC, Dávalos DM, Ramírez O, Martínez E, et al. Effect of ivermectin on time to resolution of symptoms among adults with mild covid-19: A randomized clinical trial. *J Am Med Assoc*. 2021;325(14):1426-35. doi: <https://doi.org/10.1001/jama.2021.3071>
- Mameli C, Picca M, Buzzetti R, Pace ME, Badolato R, Cravidi C, et al. Incidence of acute respiratory infections in preschool children in an outpatient setting before and during covid-19 pandemic in Lombardy Region, Italy. *Ital J Pediatr*. 2022;48(1):18. doi: <https://doi.org/10.1186/s13052-022-01221-w>
- Sistrom C, Garvan C. Proportions, odds, and risk. *Radiology*. 2004;230(1):12-9. doi: <https://doi.org/10.1148/radiol.2301031028>
- Bruxvoort K, Sy L, Qian L, Ackerson B, Luo Y, Lee G, et al. Effectiveness of mRNA-1273 against delta, mu, and other emerging variants of SARS-CoV-2: test negative case-control study. *BMJ*. 2021(375). doi: <https://doi.org/10.1136/bmj-2021-068848>
- Ritchie H, Mathieu E, Rodés-Guirao L, Appel C, Giattino C, Ortiz-Ospina E et al. *Coronavirus pandemic (covid-19)*. Our World in Data. [Internet]. 2022. Available at: <https://ourworldindata.org/coronavirus>

**Complementary material.** Summary of frequency, association and impact measures by study type.

|                  | RCT and cohort  | Case-Control   | Cross-sectional   | List  |  |       |             |          |          |              |                  |          |          |              |       |              |              |          |  |  |
|------------------|---|--|---|---|--|-------|-------------|----------|----------|--------------|------------------|----------|----------|--------------|-------|--------------|--------------|----------|--|--|
| Frequency        | $CI = \frac{a+c}{n}$<br>$CI_E = \frac{a}{a+b}$<br>$CI_{NE} = \frac{c}{c+d}$   | $O_E(casos) = \frac{a}{c}$<br>$O_E(controles) = \frac{b}{d}$   | $Prev = \frac{a+c}{n}$<br>$Prev_E = \frac{a}{a+b}$<br>$Prev_{NE} = \frac{c}{c+d}$ | $PO = \frac{a+c}{b+d}$<br>$PO_E = \frac{a}{b}$<br>$PO_{NE} = \frac{c}{d}$ | <b>RCT and cohort studies</b><br><i>CI</i> : Incidencia acumulada<br><i>CI<sub>E</sub></i> : Cumulative incidence in the exposed<br><i>CI<sub>NE</sub></i> : Cumulative incidence in the non-exposed<br><i>ID</i> : Incidence density<br><i>DI<sub>E</sub></i> : Incidence density in the exposed<br><i>ID<sub>NE</sub></i> : Incidence density in the non-exposed<br><i>IDR</i> : Incidence density ratios<br><i>IDD</i> : Incidence density differences<br><i>RR</i> : Relative risk<br><i>ARD</i> : Absolute risk difference<br><i>AF<sub>E</sub></i> : Attributable fraction in the exposed<br><i>AF<sub>p</sub></i> : Population attributable fraction.<br><i>NNH</i> : Number needed to harm<br><i>PF<sub>E</sub></i> : Preventable fraction in the exposed<br><i>PF<sub>p</sub></i> : Population preventable fraction<br><i>NNT</i> : Number needed to treat. |       |             |          |          |              |                  |          |          |              |       |              |              |          |  |  |
| Association      | $RR = \frac{a/(a+b)}{c/(c+d)}$<br>$ARD = \frac{a}{a+b} - \frac{c}{c+d}$   | $OR = \frac{a/c}{b/d} = \frac{ad}{cb}$   | $PD = \frac{a}{a+b} - \frac{c}{c+d}$<br>$PR = \frac{a/(a+b)}{c/(c+d)}$            | $POR = \frac{a/c}{b/d} = \frac{ad}{bc}$                                   | (Continuation of list from previous row)   |       |             |          |          |              |                  |          |          |              |       |              |              |          |  |  |
| Impact           | $SiRR > 1$<br>$AF_E = \frac{(CI_E - CI_{NE})}{CI_E} = \frac{(RR - 1)}{RR}$<br>$AF_p = \frac{(CI - CI_{NE})}{CI}$<br>$NNH = \frac{1}{(CI_E - CI_{NE})} = \frac{1}{ARD}$<br>$SiRR < 1$<br>$PF_E = \frac{(CI_{NE} - CI_E)}{CI_{NE}} = 1 - RR$<br>$PF_p = (IA_{NE} - IA) / IA_{NE}$<br>$NNT = \frac{1}{(CI_{NE} - CI_E)}$ | <b>Table 1.</b> Number of subjects classified by study factor and outcome. General notation in a 2 x 2 table. <table border="1" style="margin-top: 10px;"> <thead> <tr> <th></th> <th>Outcome (D)</th> <th>No outcome (ND)</th> <th>Total</th> </tr> </thead> <tbody> <tr> <th>Exposed (E)</th> <td style="text-align: center;"><b>a</b></td> <td style="text-align: center;"><b>b</b></td> <td style="text-align: center;"><b>a + b</b></td> </tr> <tr> <th>Non-exposed (NE)</th> <td style="text-align: center;"><b>c</b></td> <td style="text-align: center;"><b>d</b></td> <td style="text-align: center;"><b>c + d</b></td> </tr> <tr> <th>Total</th> <td style="text-align: center;"><b>a + c</b></td> <td style="text-align: center;"><b>b + d</b></td> <td style="text-align: center;"><b>n</b></td> </tr> </tbody> </table> |   | Outcome (D)   | No outcome (ND)  | Total | Exposed (E) | <b>a</b> | <b>b</b> | <b>a + b</b> | Non-exposed (NE) | <b>c</b> | <b>d</b> | <b>c + d</b> | Total | <b>a + c</b> | <b>b + d</b> | <b>n</b> | <b>Case-Control studies</b><br><i>O<sub>E</sub></i> (casos): Odds of exposure in cases.<br><i>O<sub>E</sub></i> (controles): Odds of exposure in controls<br><i>OR</i> : Exposure odds ratio of between cases and controls | <b>Cross-sectional studies</b><br><i>Prev</i> : Overall prevalence<br><i>Prev<sub>E</sub></i> : Prevalence in exposed<br><i>Prev<sub>NE</sub></i> : Prevalence in non-exposed<br><i>PD</i> : Prevalence difference<br><i>PR</i> : Prevalence ratio<br><i>PO</i> : Prevalence odds<br><i>PO<sub>E</sub></i> : Prevalence odds in exposed<br><i>PO<sub>NE</sub></i> : Prevalence odds in non-exposed<br><i>POR</i> : Prevalence odds ratio |
|                  | Outcome (D)   | No outcome (ND)  | Total   |   |  |       |             |          |          |              |                  |          |          |              |       |              |              |          |  |  |
| Exposed (E)      | <b>a</b>  | <b>b</b>   | <b>a + b</b>  |   |  |       |             |          |          |              |                  |          |          |              |       |              |              |          |  |  |
| Non-exposed (NE) | <b>c</b>  | <b>d</b>   | <b>c + d</b>  |   |  |       |             |          |          |              |                  |          |          |              |       |              |              |          |  |  |
| Total            | <b>a + c</b>  | <b>b + d</b>   | <b>n</b>  |   |  |       |             |          |          |              |                  |          |          |              |       |              |              |          |  |  |

$\sum_{i=1}^n x_i$ : Sum of x values from subject 1 to subject n.  
 $PT_i$ : Time to risk of the outcome being present in subject i.  
**Source:** Authors.