A Textbook of Pharmacoepidemiology and Pharmacovigilance for Pharmacy Students

Lesson 4:

Measures of risk and association

Fernando Fernandez-Llimos

ISBN-13: 978-84-124195-5-9

Porto, Portugal.

Published by CIPF, S.L.

Suggested citation: Fernandez-Llimos F. Lesson 4: Measures of risk and association. In: A Textbook of Pharmacoepidemiology and Pharmacovigilance for Pharmacy Students. Porto: CIPF; 2025.

Available at http://www.cipf-es.org

Lesson 4. Measures of risk and association

Contents

4.1 Ratios, Proportions, and Rates

Ratios, **rates**, and **proportions** are basic tools in pharmacoepidemiology, enabling researchers and clinicians to quantify relationships between drug exposures and health outcomes, assess disease burden, and evaluate public health interventions. While these terms are often used interchangeably in casual discourse, they represent distinct mathematical constructs with unique applications. This section clarifies their definitions, formulas, and practical utility in drug safety and epidemiological research.

4.1.1 Ratios: Comparing two quantities

A **ratio** is a simple comparison of two quantities, where neither value is necessarily a subset of the other. It is calculated by dividing one measure (*a*) by another (*b*):

$$\text{Ratio} = \frac{a}{b}$$

Ratios are not dimensionless and often expressed as *a:b* or "a to b." They are particularly useful for comparing groups or exposures in studies where the relationship between numerator and denominator is not hierarchical. For example, a study might calculate the **male-to-female ratio** of adverse drug reaction (ADR) reports to identify sex-based differences in drug safety. If 120 males and 80 females experienced ADRs, the ratio would be:

$$\frac{120}{80} = 1.5 \quad (\text{or } 3{:}2)$$

This indicates that for every 2 females with ADRs, there are 3 males. Ratios are also employed in **disproportionality analyses** for signal detection in pharmacovigilance. For instance, the

reporting odds ratio (ROR) compares the odds of a specific ADR being reported for a drug versus all other drugs in a database. A ratio greater than 1 suggests a potential safety signal.

However, ratios can be misleading if not contextualized. A high ratio of ADRs in elderly patients might reflect greater medication use in this population rather than inherent drug risk. Thus, ratios are most informative when paired with absolute measures like rates or proportions.

4.1.2 Proportions: Parts of a whole

A **proportion** is a type of ratio where the numerator is a subset of the denominator, representing the fraction of a population with a specific characteristic. It is calculated as:

$$\text{Proportion} = \frac{\text{Number of cases}}{\text{Total population}} \times 10^n$$

Proportions are dimensionless and are expressed as percentages (0–100%) or decimals (0–1) and are ideal for describing **prevalence** or **static characteristics** of populations. For example, in a clinical trial of 1,000 participants, if 150 develop mild hypertension, the proportion affected is:

$$\frac{150}{1000} \times 100 = 15\%$$

This measure helps quantify the burden of a condition at a specific time. Proportions are widely used in **drug utilization studies** to estimate the percentage of patients adhering to a medication regimen or the proportion of prescriptions off-label. However, they lack a temporal dimension, making them less suitable for studying dynamic processes like disease incidence.

A common pitfall arises when proportions are misinterpreted as causal measures. For instance, a high proportion of stroke patients taking anticoagulants does not imply the drug caused the strokes; confounding factors like age or comorbidities must be considered.

4.1.3 Rates: Accounting for time and risk

A **rate** measures the frequency of events over time in a population at risk. Unlike proportions, rates incorporate a **time component**, making them essential for studying disease dynamics, drug safety, and intervention efficacy. The general formula is:

$$\text{Rate} = \frac{\text{Number of new events}}{\text{Total person-time at risk}} \times 10^n$$

Here, *person-time* accounts for varying follow-up durations, such as in cohort studies where participants enter or exit the study at different times. Rates are typically expressed per 1,000 or 100,000 person-years. For example, if 50 new cases of hepatotoxicity occur in a cohort of 10,000 patients followed for 2 years, the incidence rate is:

$$\frac{50}{10000 \times 2} \times 100{,}000 = 250 \text{ cases per } 100{,}000 \text{ person-years}$$

Rates are indispensable in **pharmacovigilance** for quantifying ADR risks. Consider a drug linked to 12 cases of anaphylaxis in 50,000 users over one year. The rate of anaphylaxis would be:

$$\frac{12}{50{,}000} \times 100{,}000 = 24 \text{ cases per } 100{,}000 \text{ person-years}$$

This allows direct comparison with background rates in the general population to assess excess risk. Rates also underpin **mortality studies**; for instance, the mortality rate from opioid overdose in a region might be calculated as deaths per 100,000 residents annually.

A key strength of rates is their ability to accommodate dynamic populations, such as patients entering or leaving a health system. However, accurate calculation requires precise data on both event counts and person-time at risk, which can be challenging in real-world settings.

4.1 Prevalence and Incidence

The concepts of prevalence and incidence are crucial in pharmacoepidemiology, providing essential tools for understanding the patterns of diseases and health-related events within populations. These measures not only help describe the frequency of diseases but also guide public health interventions, resource allocation, and the evaluation of drug safety and effectiveness. Although both prevalence and incidence are measures of disease frequency, they capture different aspects of disease occurrence and are used for distinct purposes in research and clinical practice.

Prevalence refers to the proportion of individuals in a population who have a particular disease or health condition at a specific point in time or over a specified period. It is essentially a measure of how widespread a disease is within a population at a given moment. Prevalence can be further classified as point prevalence, which measures the proportion of individuals with the disease at a single point in time, and period prevalence, which considers the proportion of individuals who have the disease at any time during a defined period, such as a month or a year. For example, if a survey conducted on January 1st finds that 500 out of 10,000 people in a town have hypertension, the point prevalence of hypertension in that population is five percent. If, over the course of a year, 800 individuals report having hypertension at any time, the period prevalence for that year would be eight percent.

$$Prevalence = \frac{Existing\ cases\ (new + old)}{total\ population}$$

For example, if 150 individuals in a population of 5,000 have diabetes, the prevalence is:

$$\frac{150}{5{,}000} \times 100 = 3\%$$

Prevalence is particularly useful for assessing the burden of chronic diseases, such as diabetes, asthma, or depression, where individuals may live with the condition for many years. By quantifying the total number of people affected at a given time, prevalence provides valuable

information for healthcare planning and resource allocation. For instance, knowing the prevalence of asthma in a community can help public health officials determine how many inhalers and educational programs are needed. In the context of pharmacoepidemiology, understanding the prevalence of a disease can also inform decisions about drug procurement, distribution, and the need for ongoing monitoring of medication safety and effectiveness.

However, prevalence has its limitations. Because it includes both new and existing cases, it does not distinguish between recent and long-standing diseases. This means that prevalence can be influenced not only by the rate at which new cases occur but also by the duration of the disease and the rate at which individuals recover or die. For example, a chronic disease with a low incidence but a long duration, such as rheumatoid arthritis, may have a high prevalence, while a rapidly fatal disease with a high incidence, such as Ebola, may have a low prevalence because affected individuals do not remain in the population for long.

In contrast to prevalence, incidence measures the occurrence of new cases of a disease or health event in a population over a specified period. Incidence provides information about the risk of developing a disease and is a key measure for studying disease causation, identifying risk factors, and evaluating the impact of interventions. There are two main ways to express incidence: incidence proportion, also known as cumulative incidence, and incidence rate, sometimes called incidence density.

Incidence (cumulative incidence or incidence proportion)[1] is calculated by dividing the number of new cases that develop during a specified period by the number of individuals at risk at the start of that period. For example, if a cohort of 1,000 individuals without diabetes is followed for one year and 30 of them are diagnosed with diabetes during that time, the incidence proportion is three percent per year. This measure is straightforward and easy to interpret, but it assumes that all individuals are followed for the same length of time and that the population is stable, with no losses to follow-up.

$$Incidence = \frac{New\ cases\ (in\ a\ period)}{Risk\ population} * 100$$

For example, if 45 new cases of tuberculosis emerge in a population of 30,000 over one year, the incidence proportion is:

$$\frac{45}{30{,}000} \times 1000 = 1.5 \text{ per } 1000 \text{ person-years}$$

The incidence rate (or incidence density), on the other hand, considers the amount of time each individual is at risk and is expressed as the number of new cases per unit of person-time, such as cases per 1,000 person-years. This measure is particularly useful in studies where individuals are followed for varying lengths of time or where the population is dynamic. For example, in a study of adverse drug reactions, if 5,000 people are followed for a total of 10,000 person-years and 50 new cases of a specific reaction occur, the incidence rate would be five cases per 1,000 person-years. Incidence rates are especially valuable in pharmacoepidemiology for monitoring the safety of new medications, as they allow for the comparison of risk across different populations and time periods.

[1] Cumulative incidence is also called incidence proportion to differentiate from the incidence density, also called incidence rate. But this is a confusing terminology because incidence is always a rate.

$$\text{Incidence Rate} = \frac{\text{Number of new cases}}{\text{Total person-years at risk}} \times 1000$$

If 120 individuals develop lung cancer during 50,000 person-years of observation, the incidence rate is:

$$\frac{120}{50{,}000} \times 1000 = 2.4 \text{ per } 1000 \text{ person-years}$$

The applications of incidence in pharmacoepidemiology are numerous. Incidence is crucial for identifying outbreaks of infectious diseases, such as influenza or COVID-19, where the rapid appearance of new cases signals the need for urgent public health action. It is also essential for evaluating the effectiveness of preventive measures, such as vaccines or screening programs, by comparing the incidence of disease before and after their implementation. In the context of drug safety, incidence data help quantify the risk of adverse drug reactions, enabling clinicians and regulators to make informed decisions about the benefits and risks of specific therapies. For example, if a new anticoagulant is associated with an increased incidence of serious bleeding events compared to standard therapy, this information can guide prescribing practices and regulatory actions.

Despite their similarities, prevalence and incidence differ in important ways, and understanding these differences is key to their appropriate use and interpretation. Prevalence is a measure of the total disease burden in a population at a given time, reflecting both new and existing cases. It is influenced by both the incidence of the disease and its duration. In contrast, incidence focuses exclusively on new cases, providing a direct measure of disease risk and the effectiveness of interventions aimed at reducing that risk. The relationship between prevalence and incidence can be described by the equation: prevalence is approximately equal to incidence multiplied by the average duration of the disease. This means that a disease with a high incidence but short duration, such as the common cold, may have a low prevalence, while a disease with a low incidence but long duration, such as HIV infection, may have a high prevalence.

In practical terms, prevalence is most useful for planning healthcare services and allocating resources, as it indicates how many people need ongoing care or treatment at a given time. Incidence, on the other hand, is more informative for identifying causes of disease, evaluating the impact of preventive measures, and monitoring trends in disease occurrence over time. For example, a decrease in the incidence of myocardial infarction following the introduction of a smoking cessation program would suggest that the intervention is effective in reducing new cases, even if the prevalence of heart disease remains high due to the long survival of affected individuals.

In summary, both prevalence and incidence are essential measures in pharmacoepidemiology, each providing unique and complementary information about disease patterns in populations. Prevalence offers a snapshot of the total disease burden, informing healthcare planning and resource allocation, while incidence measures the risk of developing disease and the impact of interventions. A clear understanding of these concepts, their calculation, and their appropriate application is fundamental for pharmacy students, clinicians, and public health professionals engaged in the study and management of medication use and safety.

4.2.1 Key Differences Between Prevalence and Incidence

While both metrics describe disease frequency, their distinctions are critical for accurate interpretation:

Temporal Focus:

- **Prevalence** reflects the total disease burden at a specific time, irrespective of when cases began.
- **Incidence** captures the rate of new cases within a defined period, emphasizing disease onset

Impact of Disease Duration:

Prevalence is influenced by both incidence and disease duration. Chronic conditions with low mortality (e.g., arthritis) often exhibit high prevalence despite stable incidence, as patients live with the disease for years. This relationship is summarized as:

$$\text{Prevalence} \approx \text{Incidence} \times \text{Disease Duration}$$

For example, a disease with an annual incidence of 100/100,000 and a 10-year duration would have a prevalence of approximately 1,000/100,000

Public Health Implications:

- **High Incidence, Low Prevalence**: Seen in rapidly fatal or quickly resolved conditions (e.g., Ebola). New cases arise frequently, but deaths or recoveries limit total cases
- **Low Incidence, High Prevalence**: Common in chronic diseases (e.g., diabetes). Few new cases emerge annually, but long survival leads to accumulation

Data Interpretation:

- Prevalence is skewed by diagnostic improvements or increased survival. A rise in cancer prevalence may reflect better treatments rather than more cases.
- Incidence is less affected by these factors, making it more reliable for identifying risk factors

4.3 Risk and Association Measures

Understanding the fundamental concepts of risk and association is essential for interpreting the findings of pharmacoepidemiologic research and for applying these findings to clinical and public health decision-making. These concepts form the backbone of any investigation into the safety and effectiveness of medications in real-world populations. Before exploring the specific measures used to quantify risk and association, it is important to clarify the terms that underpin these analyses: **risk**, **association**, **exposure**, and **outcome**.

Risk refers to the probability that an individual exposed to a certain drug or intervention will experience a specific outcome or health event within a defined period. Risk is a **forward-looking concept**, capturing the chance that something such will occur as a result of the exposure. For example, if a clinical trial finds that out of 1,000 patients taking a new antihypertensive medication, 20 experience a serious side effect within one year, the risk of that side effect is two percent over the study period. Risk can be expressed as a proportion, a percentage, or as a rate *per* a defined number of person-years, depending on the study design and the nature of the data.

Association describes the statistical relationship between an exposure and an outcome. In pharmacoepidemiology, association is used to determine whether the use of a particular drug is related to an increased or decreased likelihood of a specific health event. **Importantly, association does not necessarily imply causation; it merely indicates that two variables occur together more or less frequently than would be expected by chance**. For example, an observational study may find an association between the use of a certain antibiotic and the development of *Clostridioides difficile* infection, but this does not prove that the antibiotic causes the infection. Other factors, such as underlying health conditions, concomitant medications, or healthcare practices, may also play a role. Establishing causality requires careful study design, control of confounding variables, and, often, evidence from multiple sources.

Exposure in pharmacoepidemiology refers to the factor or intervention under investigation. In most cases, exposure relates to the use of a specific drug, but it can also encompass doses, duration of therapy, route of administration, or even non-drug factors such as environmental exposures or lifestyle behaviors. For example, exposure can be defined as taking at least one dose of a statin medication, receiving a high cumulative dose of corticosteroids, or being prescribed a particular class of antibiotics. Accurate measurement of exposure is crucial for valid risk assessment, as misclassification can lead to biased estimates of risk and association.

Outcome refers to the health event or endpoint of interest in a study. Outcomes in pharmacoepidemiology are often adverse drug reactions (ADRs), hospitalizations, disease recurrence, or mortality, but also include positive therapeutic effects. The choice of outcome depends on the research question and the clinical context. For example, a study might investigate the outcome of gastrointestinal bleeding in patients taking nonsteroidal anti-inflammatory drugs (NSAIDs), the occurrence of hypoglycemia in individuals using insulin, or the reduction in cardiovascular events among patients treated with lipid-lowering agents. Outcomes must be clearly defined and consistently measured to ensure the reliability and comparability of study results.

Together, these concepts provide the foundation for the calculation and interpretation of risk and association measures. By quantifying the probability of adverse outcomes in exposed and unexposed groups, researchers can estimate the magnitude of risk associated with drug use and

compare it to alternative therapies or to the background risk in the general population. Measures of association, such as **relative risk**, **odds ratio**, and **hazard ratio**, enable the comparison of risk between groups and help identify factors that may modify or confound the observed relationships.

4.3.1 Odds Ratio

The **odds ratio**[2] **(OR)** is one of the most widely used measures of association in pharmacoepidemiology and observational research. It quantifies the relationship between an **exposure**, such as a drug, genetic factor, or environmental agent, and an **outcome**, which could be an adverse drug reaction, disease, or therapeutic benefit.

The **odds ratio compares the odds of an outcome occurring in a group exposed to a particular factor to the odds of the same outcome in an unexposed group**. To understand this concept, it is essential to **distinguish between *odds* and *probability***. **Probability** refers to the likelihood of an event occurring, expressed as the number of events divided by the total number of possible outcomes. **Odds represent the ratio of the probability that an event occurs to the probability that it does not**. For example, if the probability of developing a rash after taking a new antibiotic is 20%, the odds of developing the rash are 0.2 divided by 0.8, or 0.25 (often expressed as 1:4).

	Outcome Present	Outcome Absent
Exposed	a	b
Unexposed	c	d

The formula for the **odds ratio** is derived from a 2×2 contingency table, which categorizes individuals based on their exposure status and the presence or absence of the outcome. In such a table, the cells are labeled as follows:

- *a*: **Exposed** individuals **with** the outcome.
- *b*: **Exposed** individuals **without** the outcome.
- *c*: **Unexposed** individuals **with** the outcome.
- *d*: **Unexposed** individuals **without** the outcome.

The odds ratio is calculated as:

$$\text{OR} = \frac{\text{Odds of outcome in exposed}}{\text{Odds of outcome in unexposed}} = \frac{a/b}{c/d} = \frac{a \cdot d}{b \cdot c}$$

This formula compares the odds of the outcome in the exposed group to the odds in the unexposed group.

To illustrate this concept, consider a hypothetical **case-control study** investigating the association between antidepressant use and suicidal ideation in adolescents. Suppose

[2] Use this term always in English. Do not translate into Portuguese. It is usually translated as "*razão de probabilidades*", which is incorrect. It would be more accurate "*razão de desproporção*" or "*razão de vantage*", but English term is always preferred.

researchers recruit 150 adolescents with a history of suicidal ideation (cases) and 300 adolescents without such a history (controls). Among the cases, 60 reported using a specific antidepressant, while 90 did not. Among the controls, 80 reported using the antidepressant, and 220 did not. The data can be organized as follows:

- Exposed (antidepressant users) **with** suicidal ideation: 60 (a)
- Exposed (antidepressant users) **without** suicidal ideation: 80 (b)
- Unexposed (non-users) **with** suicidal ideation: 90 (c)
- Unexposed (non-users) **without** suicidal ideation: 220 (d)

		Outcome	
		Y	N
Exposed	Y	60	80
	N	90	220

Using the formula, the odds ratio is calculated as:

$$\text{OR} = \frac{60 \cdot 220}{80 \cdot 90} = \frac{13{,}200}{7{,}200} \approx 1.83$$

This hypothetical result suggests that **adolescents using the antidepressant have 1.83 times higher odds of suicidal ideation compared to non-users**. However, this does not necessarily mean the antidepressant *causes* suicidal ideation. The observed association could be **confounded** by factors such as the severity of depression, which may influence both the likelihood of being prescribed the drug and the risk of suicidal thoughts.

Interpreting the OR around the **null effect line**[3]:

- OR = 1: No association between exposure and outcome.
- OR > 1: Exposure increases the odds of the outcome.
- OR < 1: Exposure decreases the odds of the outcome (protective effect).

The odds ratio has several unique properties that make it valuable in pharmacoepidemiology. First, it is **mathematically symmetrical**, meaning that the odds ratio for the outcome occurring in the exposed group is the reciprocal of the odds ratio for the outcome *not* occurring. For instance, if the odds ratio for suicidal ideation in antidepressant users is 1.83, the odds ratio for *not* experiencing suicidal ideation in non-users is 1/1.83≈0.55.

Second, the odds ratio is **invariant to the direction of sampling**. In case-control studies, where participants are selected based on their outcome status, the odds ratio remains a valid measure of association even though the overall risk cannot be calculated.

Unlike the **relative risk**, which requires knowledge of the incidence of an outcome in exposed and unexposed groups, the **odds ratio** is particularly useful in studies where the outcome is rare or where the study design does not allow for direct calculation of risk, such as case-control studies.

The odds ratio is particularly advantageous in **case-control studies**, which are commonly used in pharmacovigilance to investigate rare adverse events. For example, if researchers want to

[3] This concept will be highly relevant when studying meta-analysis.

study the association between a newly marketed analgesic and acute liver failure, they might identify 50 cases of liver failure and 200 controls without liver failure. By comparing past exposure to the analgesic in these groups, they can calculate an odds ratio to assess whether the drug is associated with higher odds of the outcome. Because the incidence of acute liver failure is low, a **cohort study** would require an impractically large sample size, making the case-control design more efficient.

Despite its utility, the odds ratio has limitations. **It is often misinterpreted as a measure of risk**, leading to overestimation of clinical significance, especially when outcomes are common. Additionally, the odds ratio **does not provide information about the absolute risk** of an outcome, which is critical for clinical decision-making. For example, an odds ratio of 2.0 for a serious adverse event might correspond to an absolute risk increase from 1% to 2% (a clinically meaningful difference) or from 20% to 40% (a substantial public health concern). Therefore, the odds ratio should ideally be reported alongside absolute measures, such as the **attributable risk**, to provide a complete picture of the exposure-outcome relationship.

4.3.2 Relative Risk (or Risk Ratio)

Relative risk (RR), also known as the **risk ratio** quantifies the **association** between an **exposure**, such as a medication, environmental factor, or behavioral intervention, and a health **outcome (beneficial or adverse)**. This measure compares the **probability** of an outcome occurring in a group exposed to a specific factor to the probability in an unexposed group, providing a direct and intuitive assessment of risk. RR is used in **cohort studies** and **randomized controlled trials** (RCTs), where researchers can observe populations over time to measure the incidence of outcomes.

Relative risk is mathematically defined as the ratio of the risk (probability) of an outcome in the exposed group to the risk in the unexposed group. To calculate RR, data are typically organized into a **2×2 contingency table**:

	Outcome Present	Outcome Absent
Exposed	a	b
Unexposed	c	d

Letters represent:

- *a*: **Exposed** individuals **with** the outcome.
- *b*: **Exposed** individuals **without** the outcome.
- *c*: **Unexposed** individuals **with** the outcome.
- *d*: **Unexposed** individuals **without** the outcome.

The formula for RR is:

$$RR = \frac{\text{Risk in Exposed}}{\text{Risk in Unexposed}} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$$

This formula calculates the ratio of two probabilities: the risk in the exposed group (numerator) and the risk in the unexposed group (denominator).

Consider a hypothetical **cohort study** investigating the association between a new statin (a cholesterol-lowering drug) and myopathy, a muscle disorder. Researchers follow 2,000 patients prescribed the statin (exposed group) and 3,000 patients prescribed a non-statin lipid-lowering therapy (unexposed group) over five years. By the study's conclusion, 60 patients in the statin group and 15 patients in the non-statin group develop myopathy.

		Outcome		
		Y	N	Total
Exposed	Y	60	1940	2000
	N	15	2985	3000

Risk in Exposed Group:

$$\text{Risk}_{\text{exposed}} = \frac{60}{2000} = 0.03 \quad (3\%)$$

Risk in Unexposed Group:

$$\text{Risk}_{\text{unexposed}} = \frac{15}{3000} = 0.005 \quad (0.5\%)$$

Relative Risk:

$$RR = \frac{0.03}{0.005} = 6.0$$

This hypothetical result indicates that patients taking the new statin have **six times the risk** of developing myopathy compared to those on non-statin therapy.

The interpretation of RR relates to its magnitude:

- **RR = 1**: The risk of the outcome is identical in both groups, suggesting no association between exposure and outcome.
- **RR > 1**: The exposure is associated with an increased risk of the outcome. For example, an RR of 3.0 implies that exposed individuals are three times as likely to experience the outcome as unexposed individuals.
- **RR < 1**: The exposure is associated with a reduced risk of the outcome, indicating a potential protective effect. An RR of 0.4 suggests a 60% reduction in risk.

In the statin example, the RR of 6.0 signals a strong positive association between the drug and myopathy. However, **RR alone does not establish causation**. **Confounding** factors-such as age, pre-existing muscle conditions, or concomitant medications-must be examined to rule out alternative explanations.

Also, **Absolute Risk** context is necessary to establish the clinical relevance of a RR result. RR does not convey the baseline risk of an outcome, which is critical for clinical decision-making. For example, an RR of 3.0 for a rare adverse event (e.g., 0.1% to 0.3% risk) represents a small absolute risk increase (0.2%), while the same RR for a common outcome (e.g., 10% to 30% risk) reflects a substantial public health burden.

4.3.3. Differences Between Odds Ratio and Relative Risk

While OR and RR share similarities, their conceptual foundations, mathematical properties, and applications differ significantly. Understanding these distinctions is critical for interpreting study results accurately and selecting the appropriate measure based on study design and outcome frequency.

Conceptual Differences

The **relative risk (RR)** compares the *probability* of an outcome in two groups, answering the question: ***How many times more likely is the outcome in the exposed group compared to the unexposed group?*** For example, if smokers have a 20% risk of lung cancer and non-smokers a 5% risk, the RR is 4.0, indicating smokers are four times as likely to develop the disease. RR is intuitive for clinicians and patients because it directly relates to individual risk.

The **odds ratio (OR)**, in contrast, compares the *odds* of an outcome, which is the ratio of the probability of occurrence to non-occurrence. It answers: ***How many times higher are the odds of the outcome in the exposed group compared to the unexposed group?*** Odds are less intuitive than probabilities because they represent a **ratio of successes to failures** rather than a proportion of the total population. For instance, if the odds of lung cancer in smokers are 0.25 (20% risk) and 0.05 in non-smokers (5% risk), the OR is 5.0. While the OR also indicates higher risk in smokers, its interpretation as a multiplier of odds is less straightforward than RR.

Mathematical Differences

The formulas for RR and OR highlight their mathematical distinctions. Consider a 2×2 contingency table:

- **Relative Risk**:

$$\mathrm{RR} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$$

RR compares the risk of the outcome in the exposed group ($a/(a + b)$) to the unexposed group ($c/(c + d)$).

- **Odds Ratio**:

$$\mathrm{OR} = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{a \cdot d}{b \cdot c}$$

OR compares the ratio of outcomes to non-outcomes in the exposed group (a/b) to the same ratio in the unexposed group (c/d).

The key mathematical difference lies in the denominators: RR uses the total number of individuals in each group, while OR uses the number of individuals without the outcome. This distinction becomes consequential when outcomes are common.

Example: Divergence in common outcomes

Consider a study of a medication linked to a frequent side effect, such as headache. Suppose 200 patients are exposed to the drug, and 160 develop headaches, while 200 unexposed patients have 80 headaches:

	Headache	No Headache
Exposed	160 (a)	40 (b)
Unexposed	80 (c)	120 (d)

- **RR Calculation**:

$$RR = \frac{\frac{160}{200}}{\frac{80}{200}} = \frac{0.8}{0.4} = 2.0$$

- **OR Calculation**:

$$OR = \frac{160 \cdot 120}{40 \cdot 80} = \frac{19200}{3200} = 6.0$$

Here, the RR of 2.0 indicates that exposed patients are twice as likely to experience headaches. However, the OR of 6.0 suggests the odds are six times higher. This discrepancy arises because the OR exaggerates the association when outcomes are common. The RR, grounded in probabilities, provides a more accurate reflection of risk in such scenarios.

Rare Disease Assumption

When outcomes are rare (prevalence <10%), the OR approximates the RR. For example, if a medication is associated with a 0.2% risk of anaphylaxis in exposed individuals versus 0.1% in unexposed individuals:

- RR: $\frac{0.002}{0.001} = 2.0$
- OR: $\frac{0.002/0.998}{0.001/0.999} \approx 2.0$

Here, the OR and RR are nearly identical. However, for common outcomes, the OR diverges from the RR, as seen in the headache example.

4.3.4. Dispersion around OR and RR

When presenting measures of association such as the odds ratio (OR) and relative risk (RR), it is crucial not only to report the point estimates but also to include a measure of the dispersion around these estimates, most commonly the 95% confidence interval (95% CI). The confidence interval provides essential information about the precision and reliability of the estimated association and allows for proper interpretation of whether an effect is likely to be real or attributable to random variation in the data. Several reasons justify the need to always present the dispersion of the two estimates:

- **Precision**: A 95% CI quantifies how confident we can be that the true value of OR or RR lies within a certain range. If the CI is wide, the estimate is less precise, potentially due to a small sample size or high variability in the data. A narrow CI indicates a more precise estimate.
- **Statistical significance**[4]: For measures such as OR and RR, if the 95% CI includes the value 1, there is no statistically significant association between exposure and outcome at the 5% level. If the interval does not include 1, the association is considered statistically significant.
- **Clinical importance**: A CI provides insight into the possible range of effect sizes. For example, even if an OR is large, a wide CI that includes 1 means we cannot exclude the possibility of no association.
- **Model comparison**: In research, CIs can be used to compare the fit and reliability of different models or estimations. Models yielding narrower CIs are generally preferred because they indicate less dispersion and more certainty about the estimated effect.

The calculation of 95% CIs around OR and RR commonly relies on the natural logarithm (ln) transformation, because OR and RR are ratio measures and their distribution is asymmetric.

Calculating the 95% confidence interval (95%CI) of OR (SE=Standard error):

$$\text{Lower Bound} = e^{\ln(OR) - 1.96 \times SE[\ln(OR)]}$$

$$\text{Upper Bound} = e^{\ln(OR) + 1.96 \times SE[\ln(OR)]}$$

Calculating the 95% confidence interval (95%CI) of RR (SE=Standard error):

$$RR_{lower} = \exp\left(\ln(RR) - 1.96 \times SE[\ln(RR)]\right)$$

$$RR_{upper} = \exp\left(\ln(RR) + 1.96 \times SE[\ln(RR)]\right)$$

Since the calculation of both dispersion measures is complex, statistical software of online calculators are commonly used[5,6].

[4] This will be highly relevant when studying meta-analysis.
[5] https://www.medcalc.org/calc/odds_ratio.php
[6] https://www.medcalc.org/calc/relative_risk.php

4.3.5. Absolute risk measures

Absolute measures provide critical insights into the real-world impact of exposures and interventions by quantifying differences in risk between groups. Unlike relative measures such as the **odds ratio** or **relative risk**, which express associations as ratios, absolute measures convey the actual magnitude of risk change, making them indispensable for clinical decision-making and public health planning. This section explores three key absolute measures- **attributable risk (AR)**, **absolute risk reduction (ARR)**, and **number needed to treat (NNT)**.

Absolute Risk Reduction (ARR)

Absolute risk reduction measures the **clinical benefit of an intervention** by comparing the risk of an adverse outcome in a control group to that in an intervention group. It is calculated as:

$$ARR = Risk(exposed) - Risk\ (unexposed)$$

For instance, in a hypothetical randomized trial of a new anticoagulant, suppose the annual risk of stroke is 8% in the control group (standard therapy) and 3% in the intervention group (new drug). The ARR is:

$$ARR = 0.03 - 0.08 = \ -0.05\ \approx 5\%$$

This hypothetical result means the new anticoagulant reduces the absolute risk of stroke by 5 percentage points. ARR is vital for understanding the practical impact of an intervention, as it reflects the actual proportion of patients who benefit.

Quite frequently, the ARR is presented in textbooks or classroom slides in the opposite order, but the common way of presenting the results in biomedical articles is as indicated before. For example, the pivotal RCT published in New England Journal of Medicine[7], reported "In the all-randomized modified intention-to-treat population, participants receiving molnupiravir had a lower risk of hospitalization or death through day 29: 6.8% (48 of 709 participants) in the molnupiravir group as compared with 9.7% (68 of 699 participants) in the placebo group (difference, 3.0 percentage points; 95% CI, −5.9 to −0.1)". This means:

$$ARR = Risk(molnupiravir) - Risk\ (placebo)$$

$$ARR = 0.068 - 0.097 = \ -0.029\ \approx 3\%$$

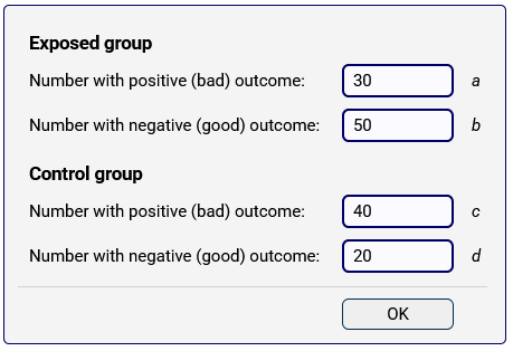
For clarity reasons, the risk is reported in the text of the article as an absolute value (3%), but as a negative value in the 95%CI (95% CI, −5.9 to −0.1), which is the correct one.

[7] Jayk Bernal A, Gomes da Silva MM, Musungaie DB, et al. Molnupiravir for Oral Treatment of Covid-19 in Nonhospitalized Patients. N Engl J Med. 2022;386(6):509-520. doi:10.1056/NEJMoa2116044

Risk of using online RR or OR calculators with no educated interpretation.

Pharmacoepidemiology aims to explore the association between exposure (usually to a drug) and effects. The latter can be positive effects (e.g., cure, improvement, remission) or negative effects (e.g., death, ADR, relapse). Although, rational and equations are applicable to any outcome, whatever the direction has, cognitive bias may pose difficulties in the understanding.

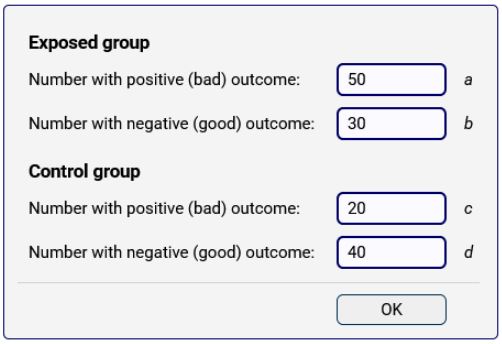
Different online OR and RR calculators may use different and confusing terminology. Terms like "positive (bad) outcome" or "disease" may not be accurate to some effects explored. In a hypothetical example of the association of an anticancer drug to cancer remission, with 50 (out of 80 patients under remission in the exposed group, but only 20 (out of 60) in the unexposed group, if one considers that the "bad outcome" of the absence of remission, tables could be interpreted as:



Results

Relative risk	0.5625
95% CI	0.4025 to 0.7861
z statistic	3.369
Significance level	P = 0.0008
NNT (Benefit)	3.429
95% CI	2.213 (Benefit) to 7.609 (Benefit)

This means that (RR<1) the exposure (the drug) is associated with a reduced risk of the outcome (reduced risk of remission), so the drug is not efficacious, which is incorrect. But, also in this result, the NNT is 3.369 (NNT≈ 4), which means that the drug is efficacious.



Results

Relative risk	1.8750
95% CI	1.2618 to 2.7862
z statistic	3.111
Significance level	P = 0.0019
NNT (Harm)	3.429
95% CI	7.609 (Harm) to 2.213 (Harm)

However, if one understands that "positive outcome" is not "bad", as described in the website, the RR results in 1.88 (RR>1), so the exposure (the drug) is associated with an increased risk of the outcome (cancer remission), meaning that the drug is efficacious, which is correct. But in this second case, the NNT is labeled as "Harm", which means that it should be read as a NNH (Number Needed to Harm), meaning that the drug is not efficacious, which is incorrect.

This is why a recommendation is using the traditional epidemiological table accompanied of human reasoning to have a judicious interpretation of the online calculator results.

Number Needed to Treat (NNT)

NNT and NNH translate ARR into a clinically actionable metric. They depict the average number of patients who must receive a particular intervention to prevent one additional adverse event (the NNH) or to secure one additional positive outcome (the NNT), as compared to a control or placebo group, over a specified period and for a defined endpoint.

Both are calculated as the inverse of ARR:

$$\mathrm{NNT} = \frac{1}{\mathrm{ARR}}$$

Using the anticoagulant example with an ARR of 5% (0.05):

$$NNH = \frac{1}{0.05} = 20$$

An NNH of 20 means 20 patients must be treated with the new anticoagulant for one year to prevent one stroke. NNT is a cornerstone of evidence-based medicine, helping clinicians weigh the benefits and costs of therapies. Lower NNT values indicate more effective interventions; for example, an NNT of 10 is preferable to an NNT of 50.

Dispersion should also be calculated in NNT and NNH point estimates. Also, their calculation is complex and software and online calculators are preferred[8].

4.3.5. Hazard Ratio

The **hazard ratio** is a key measure in pharmacoepidemiology, particularly in the context of time-to-event (survival) analyses. It quantifies the relative risk of an event occurring at any given point in time in one group compared to another, while accounting for the time at which events take place. Hazard ratios are widely used in studies assessing the effect of drugs, interventions, or exposures on outcomes such as mortality, disease recurrence, or adverse events.

Mathematically, the hazard ratio (HR) is defined as the ratio of the hazard rates between two groups. The hazard rate, h(t)h(t), is the instantaneous probability that an event will occur at time tt, given that the individual has not yet experienced the event up to that time. For two groups, such as a treatment group and a control group, the hazard ratio can be expressed as:

$$HR = \frac{h_1(t)}{h_0(t)}$$

where h1(t) is the hazard in the exposed or treatment group, and h0(t) is the hazard in the unexposed or control group.

Like OR and RR, HR relates to its magnitude:

[8] https://www.medcalc.org/calc/relative_risk.php

- **HR = 1**: The risk of the outcome is identical in both groups, suggesting no association between exposure and outcome.
- **HR > 1**: The exposure is associated with an increased risk of the outcome.
- **HR < 1**: The exposure is associated with a reduced risk of the outcome, indicating a potential protective effect.

Unlike RR or OR, which compare cumulative risk or odds over a fixed period, the hazard ratio **incorporates the element of time** and allows for comparison across groups throughout the duration of follow-up. This is particularly important in studies where the risk of an event is not constant over time or where some individuals are censored (i.e., lost to follow-up, or the event has not occurred by the end of the study). The Cox proportional hazards model is a common statistical method used to estimate hazard ratios, adjusting for potential confounding variables. When using this Cox proportional hazards model, the hazard in each group can be modeled as:

$$h_i(t) = h_0(t)\exp(\beta X_i)$$

where h0(t) is the baseline hazard at time t, Xi is a binary (or continuous) indicator for group assignment (e.g. 1 for treatment, 0 for control), and β is the estimated coefficient from the regression. The hazard ratio is then obtained by taking the exponential of the regression coefficient:

$$HR = e^{\beta}$$

Calculation of HR is complex, and dispersion should also be calculated for HR point estimates, so software applications are preferred for tis purpose (e.g., HR is the "Exp(B)" in the results of a Cox Regression analysis in SPSS).

If a drug is being tested and the hazard ratio for the drug versus a placebo is 0.75 with a 95% confidence interval of (0.60, 0.95), this suggests that, at any given point in time during the study period, patients receiving the drug have a 25% lower instantaneous risk of the event (e.g., death) compared to those receiving the placebo, and this reduction is statistically significant[9].

[9] Sashegyi A, Ferry D. On the Interpretation of the Hazard Ratio and Communication of Survival Benefit. Oncologist. 2017;22(4):484-486. doi:10.1634/theoncologist.2016-0198