Universidad del Cauca

ERASMUS UNIVERSITEIT ROTTERDAM

MEASURES OF OCCURRENCE, ASSOCIATION AND EFFECT

Jose Andres Calvache MD MSc PhD

Department of Anesthesiology, Universidad del Cauca, Colombia

Department of Anesthesiology, Easmus University Medical Centre, Rotterdam, The Netherlands

@jacalvache

Measures of occurrence, association and effect

- To describe the occurrence of an event of interest (X)
- When we have two variables (*X, Y*) there is an interest in their degree of "**association**"

  - Measures of occurrence:
    - Risks, rates (tasas), Incidence, Prevalence

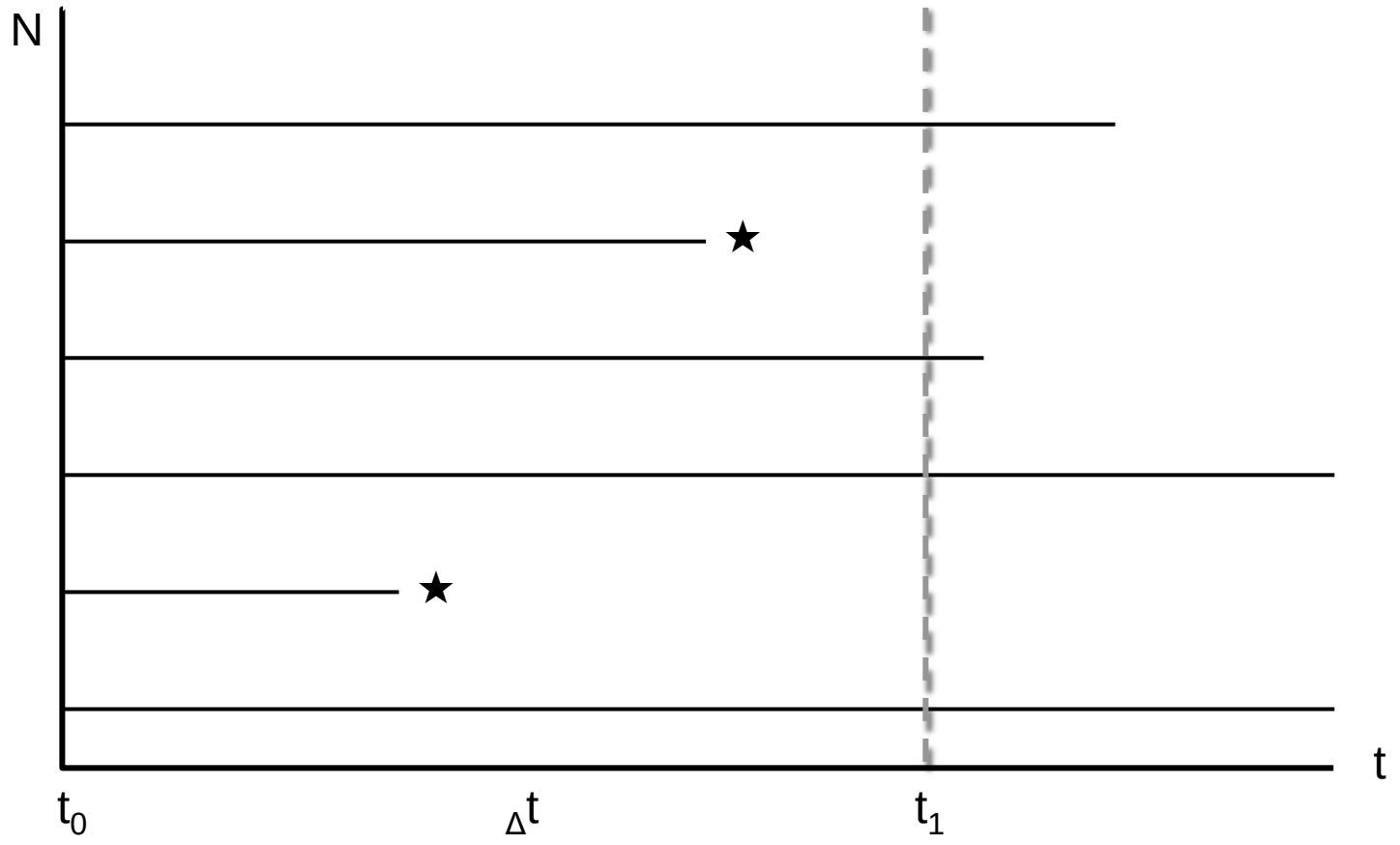
  - Measures of association:
    - Absolute: Absolute risk and Absolute risk reduction (ARR)
    - Relatives: Odds, Relative risk, Odds ratio, Hazard ratio

  - Pearson linear “correlation” coefficient

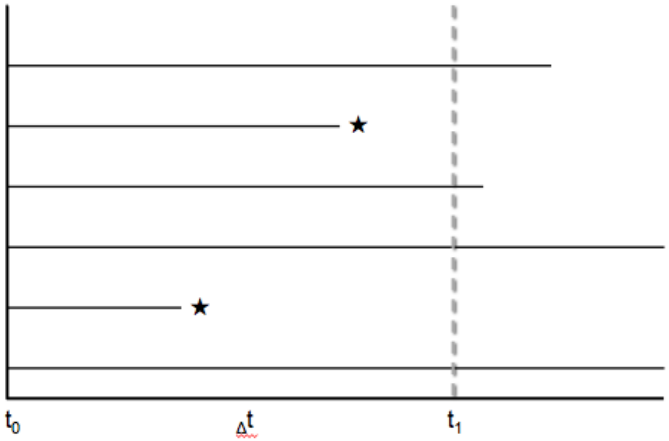
Measures of occurrence, association and effect

- Risk
  - Probability of an event occurring (illness, death…)
  - Measure of frequency
  - Measure of association
- Basic measures of frequency
  - Incidence (incident or new cases)
    - Proportion or rate
  - Prevalence (prevalent or existing cases)
    - Point or period prevalence

Risk - Incidence



Risk - Incidence



- Risk
  - Cumulative incidence

$$CI = \frac{\text{Number of new events}}{\text{Total population at risk}}$$

Risk – Cumulative incidence

- Cumulative incidence
- It is a proportion (cases / total) – fraction (%)
- Represents risk (from 0 to 1)
- Requires a **time specification**
- It requires defining "population at risk"
- Absence of losses (complete follow-up)

Risk – Cumulative incidence

- Examples,
  - 5-year mortality of patients diagnosed with lung cancer
  - Mortality within 30 days of ICU admission
  - 10-year incidence of gastric cancer in general practitioners

- Uses in health sciences
  - (Absolute) risk
    - Fatality, lethality, mortality
    - Survival to 5 years of…

Risk – "Rate“ (tasa)

- It's a measure of "how fast something happens."

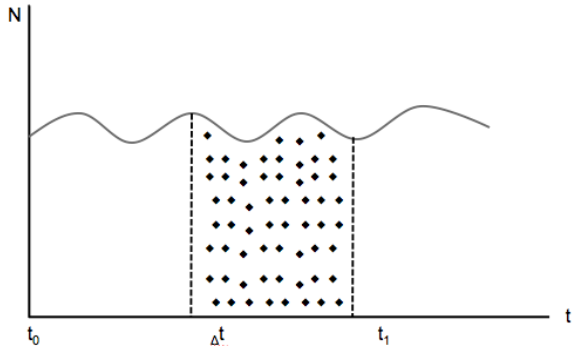
1 hour

60 km

Average rate = 60 km/h

Risk – Incidence rate

- Incidence rate(velocity)
- Event Rate
- Incidence rate
- It is a density / NOT a ratio or proportion
- Allows risk estimation only indirectly

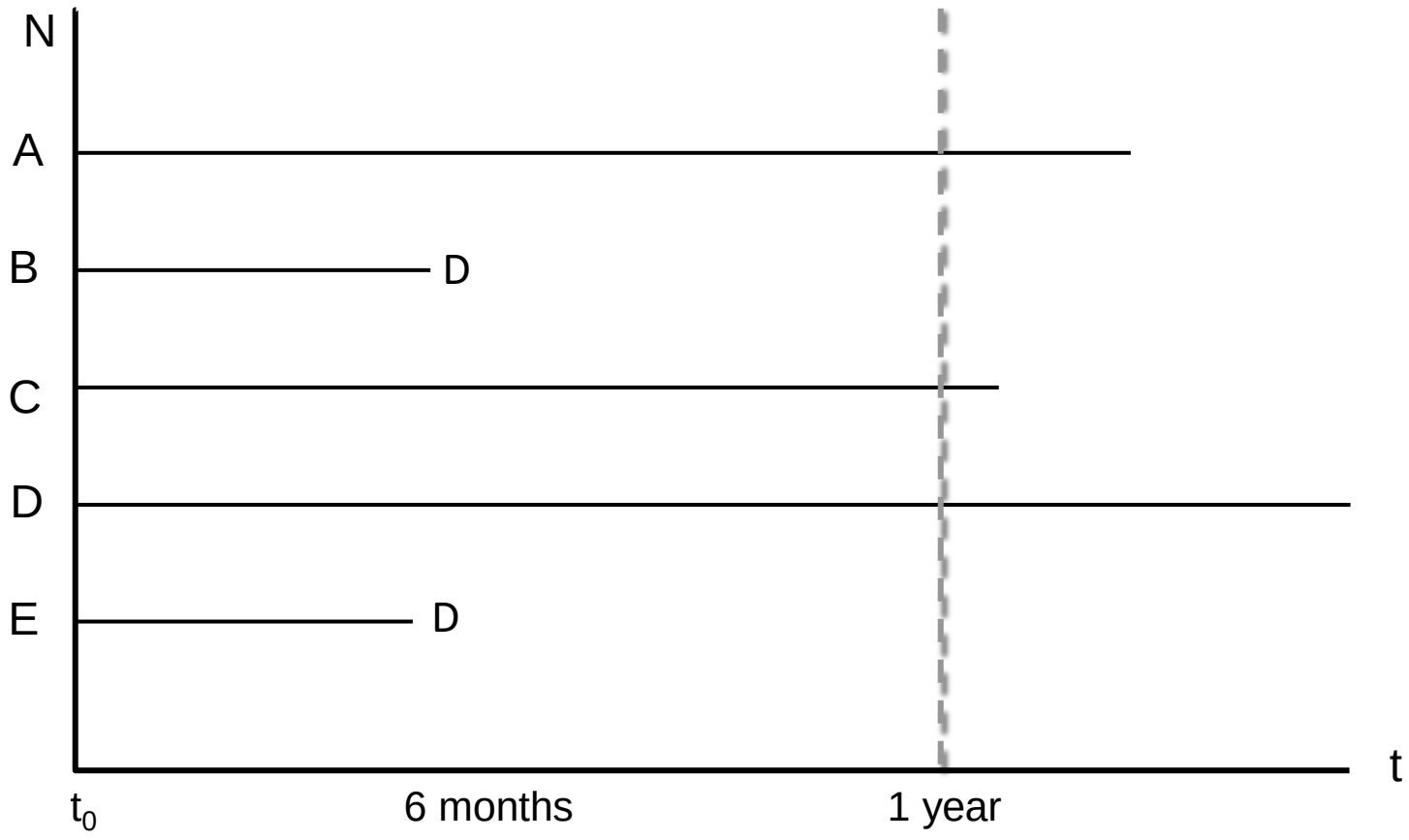


Definition,

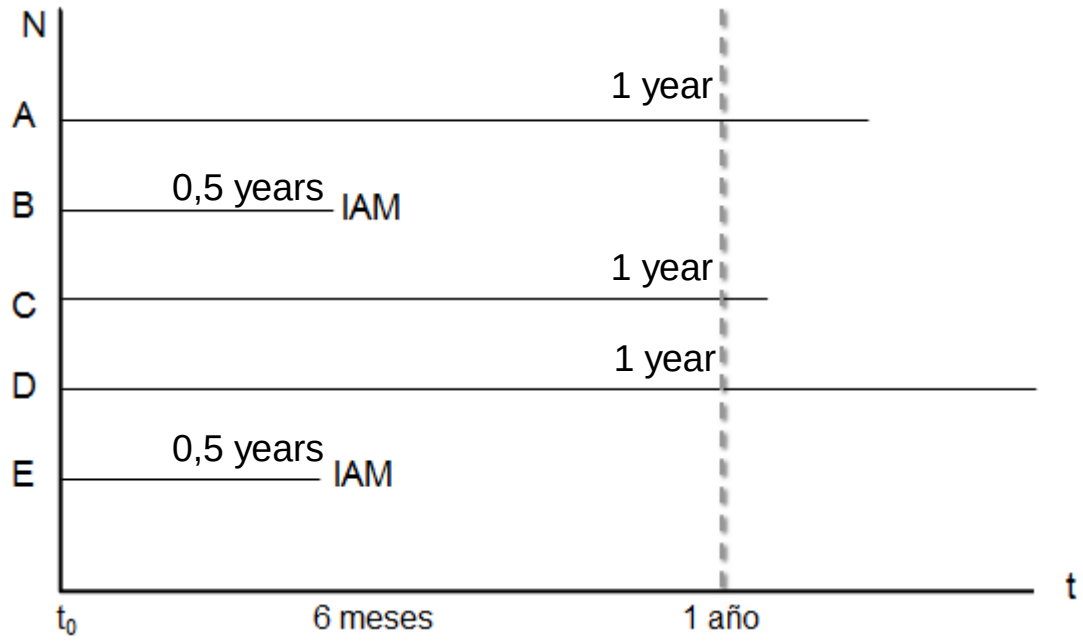
Number of subjects presenting the event per unit of subject-follow-up time at risk of the particular event

$$IR = \frac{\text{Number of new events}}{\text{Person-time free of the event}}$$

CI versus IR



CI versus IR



CI = 2 / 5
CI = 0,4

Cumulative incidence at 1 year of AMI is 40 %

IR = 2 / (3 + 0,5 + 0,5)
IR = 2 / 4

Incidence rate of 0.5 person-years
Incidence rate of 5 in 10 person-years

IR Incidence Rate - Example

- 5 new cases per 10 person-years
- Interpretation
  - An average of 5 new cases occur, for every 10 years of observation in a cohort of disease-free subjects.
  - Its interpretation is not easy to be used in the prediction of risk at the individual level
  - It becomes useful at the population level or in research

Risk vs Rate

- Cumulative incidence
- Proportion (0 – 1)
- Probability that an individual will develop the disease in a certain time
- Use for individual prognosis
- Does not take into account situations such as loss of follow-up or risk competence
- Easy to quantify in small populations without major losses

- Incidence rate
- Describes how quickly new events occur in a population
- It is useful in etiological comparisons (causality)
- Low use at the individual level
- Can handle situations such as loss of follow-up or risk competence
- Can be calculated in large populations or with losses during follow-up

Mortality

- It can be expressed as cumulative incidence (proportion) or as rate
- Classification as a rate is used in vital statistics

$$\text{Mortality rate} = \frac{\text{Number of all-cause deaths}}{\text{Population at risk at the same period*}} \times 100.000$$

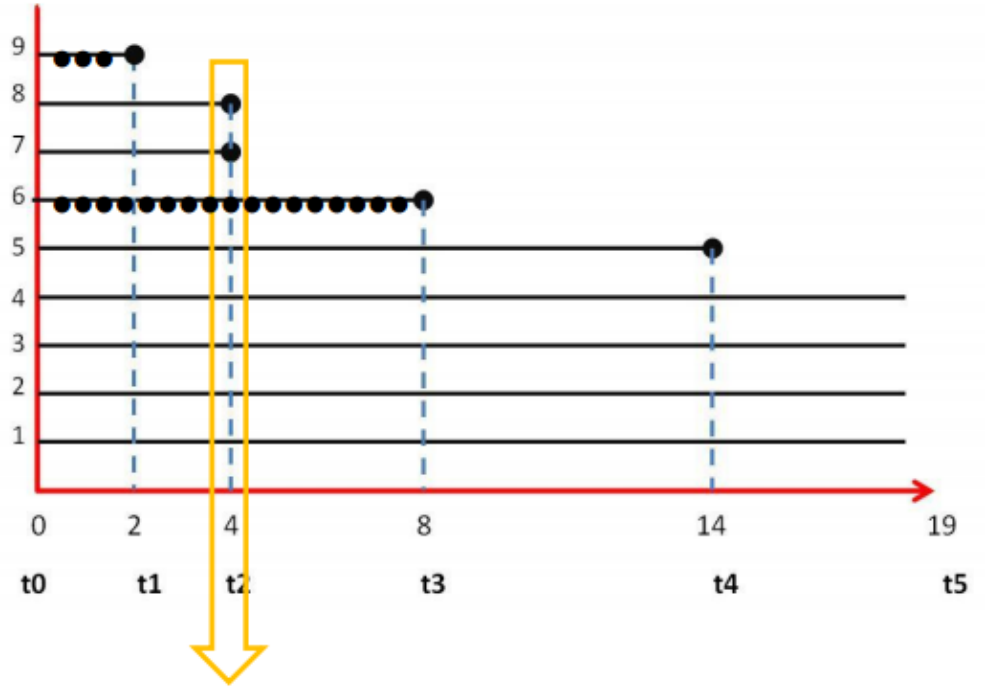
*Population at risk at the middle of the period (year)

Risk - Prevalence

- Existing cases / total population
- Cases that are present
- Proportion (%). Ranges from 0 to 1
- Point or period

Point prevalence

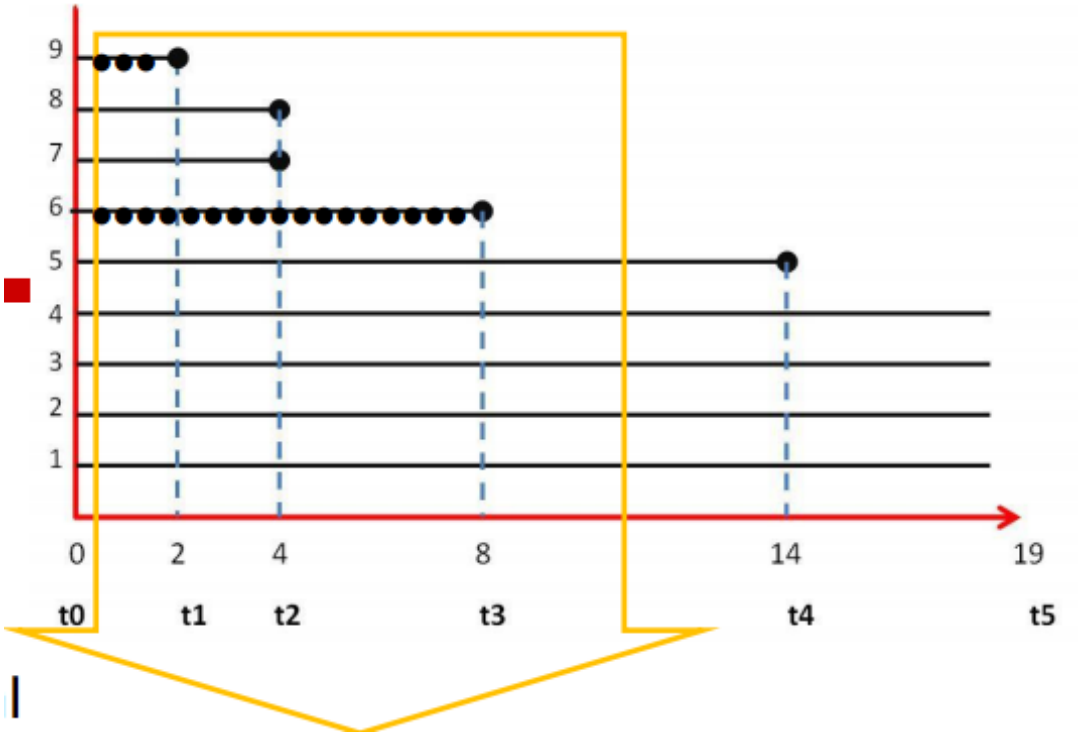
- Prevalence of an event at a point in time
- Proportion of the population with the event P(A)t
- Synonym: P. Punctual, P. Proportion, P. Rate



P(p)= 3/8 = 37%

Period prevalence

- Prevalence of a state in a defined period of time
- Total number of people in the defined period



P(per) [1 y 4 year]
P(per) = 4/9 = 44%

Point / Period

Proportion of people experiencing the clinical event (Case), at a given point in time. It is a cut at a given moment, in which I count the CASES

Proportion of cases. It is considered those who fell ill in previous periods and who are still CASES (Existing and New). Each person represented in the Numerator had the disease at some point during the specified period

Prevalence and incidence relationship

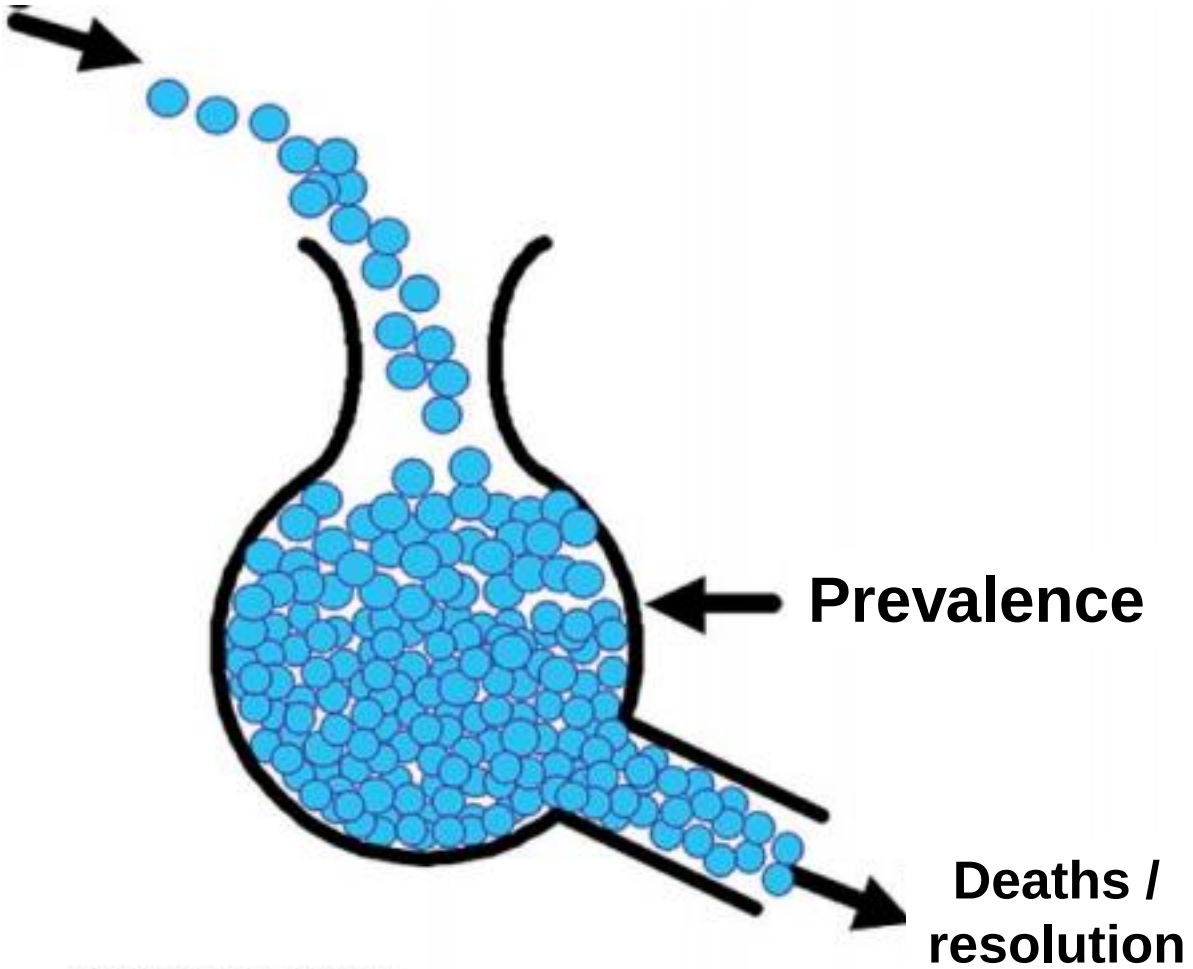
- Relationship between prevalence and incidence rate (IR)

$$Prevalence = IR \times duration\ of\ the\ disease$$

**Only useful in cases of low IR

- Prevalence depends on INCIDENCE
  - A high incidence produces a high prevalence if the duration of cases does not change
  - Prevalence depends on the DURATION of the disease (recovered, dead, migrated)

Incidence



Gordis: Epidemiology, 4th Edition.

Relationship between incidence and prevalence: IV.

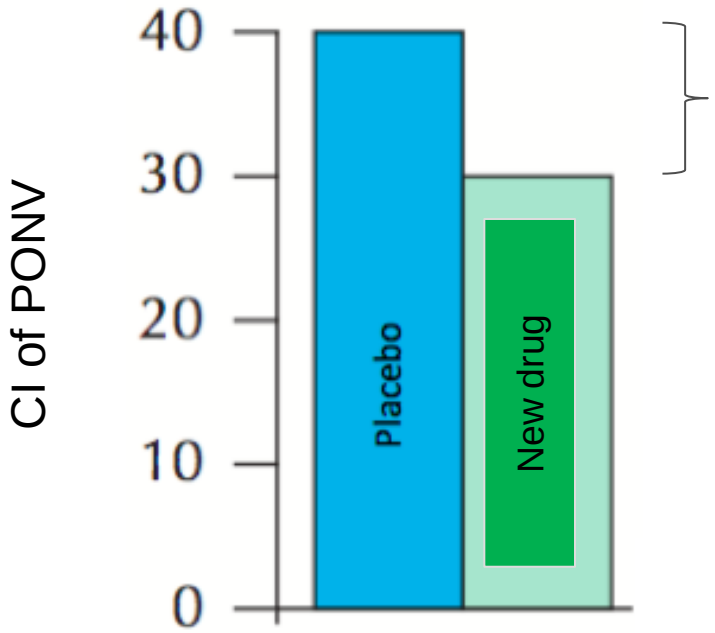
2 x 2 table

		Outcome / Disease		
		Yes	**No**	**Total**
Exposure	**Exposed**	a	b	**a+b**
	Non exposed	c	d	**c+d**
	Total	**a+c**	**b+d**	**a+b+c+d**

- Frequency measurements can be extracted
  - CI (exp) = a / a+b
  - CI (non exp) = c / c+d
  - CI = a+c / a+b+c+d
  - Prevalence = a+c / a+b+c+d

Absolute association measures

- **Absolute risk** is the probability of occurrence of an event such as an outcome or disease (% = CI)



- **ARR** is a direct derivative of the absolute risks in each group (difference) and represents a direct estimate of population risk

Relative association measures
Relative risk / CIR

- It is a ratio of cumulative incidences
- Cumulative incidence ratio (CIR)

		Outcome / Disease		
Exposure		**Yes**	**No**	**Total**
	Exposed	a	b	**a+b**
	Non Exposed	c	d	**c+d**
	Total	**a+c**	**b+d**	**a+b+c+d**

CI (exp) = a/a+b

CI (non exp) = c/c+d

$$(CIR) = \frac{CI\ en\ expuestos}{CI\ en\ no\ expuestos} = \frac{\left(\frac{a}{a+b}\right)}{\left(\frac{c}{c+d}\right)}$$

Relative risk / CIR

- RR = 1 no association between categories and outcome
- RR > 1 positive association or risk
- RR < 1 negative association or less risk (protective)

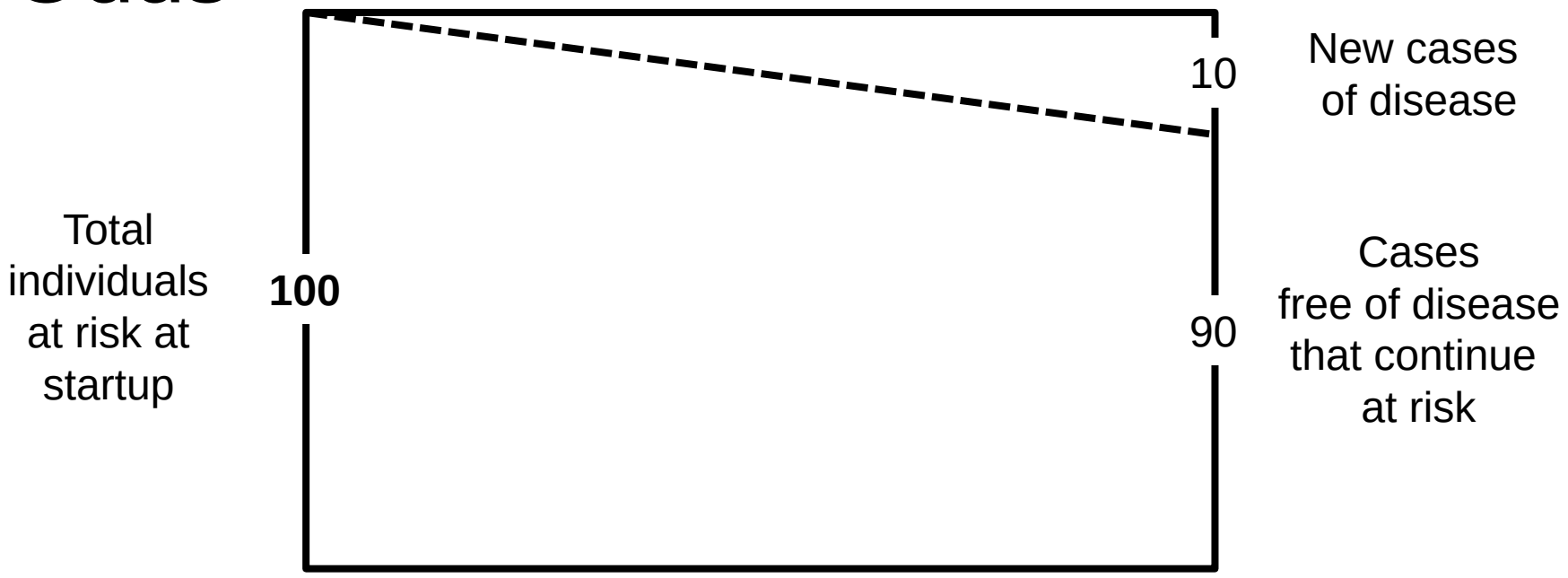
(Honolulu Heart Programme, Abbot, 1980)

Cohort	Cerebral thrombosis after 12-year FU: Yes	No	Total
Smokers	171	3264	3435
Non-smokers	117	4320	4437
Total	228	7584	7872

CI (smok) = 171 / 3435 = **0,049** en 12 years of follow-up
CI (non smok) = 117 / 4437 = **0,026** en 12 years of follow-up

$$Riesgo\ Relativo\ (CIR) = \frac{CI\ en\ expuestos}{CI\ en\ no\ expuestos} = {}^{0,049}/_{0,026} = \mathbf{1,89}$$

Odds



- Risk (CI) = 10 / 100 = 0,1 = 10%
- Odds = 10 / 90 = 0,11 = 11%

Odds : It is the relationship between the probability of occurrence of a event and the non-occurrence of the same event

Odds ratio

- Useful association measure when the entire exposed population is not available
- Useful in case-control studies but applicable to any study data
- Flexible mathematical properties (0 - Infinite) / numerator not included

Exposure		Cases	Controls
	Exposed	a	b
	Non Exposed	c	d
	Total	**a+c**	**b+d**

Odds (exp) = a / b

Odds (non exp) = c / d

$$Odds\ Ratio = \frac{a \times d}{c \times b}$$

(MacMahon et al., NEJM, 1981)

Cups of coffee/day	Pancreatic cancer: Cases	Controls	Total
≥ 1	347	555	902
0	20	88	108
Total	367	643	1010

$$\widehat{OR} = \frac{347 \times 88}{555 \times 20} = 2.75$$

RR and OR Interpretation

- Both indicate magnitudes of association

- The RR is the reason for two cumulative incidences (risks)
  - As they are the direct risk assessment it is appropriate to refer to RR as "risk"
  - RR = 1.8. Compared to the control group, the intervention group had a **1.8-fold increased risk** of developing side effects.

- In the interpretation of the OR it is preferable to speak of "times more probability" and avoid the term "risk"
  - OR = 1.8. Compared to the control group, the intervention group was 1.8 times more likely to develop side effects.

  - When the outcome event is infrequent, the OR and RR approach (< 10%)
  - When interpreting the clinical significance of an outcome expressed as OR, it should be treated as an RR (most of the times)

Risk “vs” Odds

CHARACTERISTIC	PROBABILITY	ODDS
Ratio	$\frac{\text{occurrence}}{\text{whole}}$	$\frac{\text{occurrence}}{\text{nonoccurrence}}$
Range	0 to 1	0 to ∞
Transformation to other measure	$\text{odds} = \frac{\text{probability}}{1 - \text{probability}}$	$\text{probability} = \frac{\text{odds}}{1 + \text{odds}}$

Effective Clinical Practice May/June 2000 Volume 3 Number 3

Clarifying some concepts

- What (exactly) is a **risk factor** ? *

*Think in prediction versus explanation

Is caviar a risk factor for being a millionaire?

Anders Huitfeldt argues that the answer depends on your definition of "risk factor" and calls for greater clarity in research

BMJ 2016;355:i6536 doi: 10.1136/bmj.i6536

Anders Huitfeldt *postdoctoral scholar*

- The **risk factor approach** to epidemiology was introduced by the Framingham Heart Study investigators, who first alluded to the idea in 1951 and introduced in 1961
- The semantic confusion has hindered precise communication about study design and data analysis

- You have a secretive friend, and, among other questions, you are interested in knowing whether he is a millionaire. You are aware that there are some attributes, or risk factors, that are thought to be linked to being a millionaire

- Clinical research can generally be divided into four broad objectives based on the intended use of the information obtained by the study:

  - **diagnosis,**
  - **prognosis,**
  - **treatment effects, and**
  - **aetiology**

Table 1| **Objectives of clinical research and associated definitions of risk factor**

Research objective	Definition of risk factor*	Suggested term	Example of application	Preferred data analysis or study design	Relevant biases and shortcomings
Diagnosis	Any personal attribute that can be used to make a diagnosis more reliable	Diagnostic factor	Serum cholesterol in people presenting with chest pain[5]	Prediction model with binary outcome variable (measured at the same time as the diagnostic factor)†	Ascertainment of outcome may have imperfect sensitivity and specificity. Model may be overfit to training dataset
Prognosis	Any personal attribute that can be used to make more reliable predictions about future risk of medical conditions	Prognostic factor	Serum cholesterol predicts future cardiovascular disease[6]	Prediction model with time-to-event outcome variable	As above
Treatment effects	An action that may be taken to increase or decrease the probability of the outcome	Treatment effect	Cardiac risk is reduced by lowering serum cholesterol levels[7]	Randomised controlled trials. Observational studies with explicit causal models[8]	Confounding, selection bias, etc
Aetiology	A phenomenon, action, or substance that has a role in the aetiological mechanism	Aetiological factor	Cholesterol is involved in the mechanism behind atherosclerosis[9]	Some aetiological questions can be examined using the same methods as for treatment effects (eg, mendelian randomisation).[10] For others, there is no consensus on preferred study design. Relevant concepts include reverse causal inference,[11] excess fraction,[12] aetiological fraction,[13] and sufficient component cause models[14]	Imprecisely stated research questions because of current state of statistical methods

*Note that not all commonly accepted risk factors for cardiovascular disease meet all four definitions. For example, family history is valid both as a prognostic factor and as a diagnostic factor, but if you attempt to reduce your patient's coronary risk by starting their parents on primary prevention, you are likely to be struck from the register. Some variables even have opposite effects depending on whether we are interested in prediction or causation. For example, if the patient's clinical history shows that he has had a coronary artery bypass graft, your risk estimate increases for the purposes of both diagnosis and prognosis, although the procedure itself almost certainly reduced his risk. †Such models are often termed "detection models" in the data mining literature, where they are used to detect fraud.

- Therefore, when conducting observational studies, data analysis needs to be designed to match the particular definition that is being considered.

- The definition of “risk factor” will vary depending o whether a research question is exploring diagnosis, prognosis, treatment effects, or aetiology

- Unless a definition is specified, it is not possible for readers of research papers to understand what the investigators attempted to learn or evaluate whether they succeeded in their objectives