

MECOR

Basic descriptive statistics: Measures of association



Philip Hopewell, M.D.
Curry International Tuberculosis Center
University of California, San Francisco

Overview of Presentation

Points covered

- Prevalence
- Incidence



Basic descriptive statistics

- Relative risk
- Odds ratio



Measures of association

These are fundamental tools that make information useful.
This is a starting point

Using Information Sensibly

© Original Artist
Reproduction rights obtainable from
www.CartoonStock.com



"There's a 50% chance of rain, so I only watered half the lawn."

Question

You are the Director of the Division of Communicable Diseases in the Ministry of Health and the Minister wants to be briefed on tuberculosis (so he doesn't look stupid in an important international meeting).

How would you describe the magnitude of tuberculosis in the country?

Determining Prevalence

- The standard statistic for conveying the magnitude of a condition in a population at a specified point in time is its prevalence.
- What data do you need to determine the prevalence?
 - Numerator (Case definition and method of identification)
 - Denominator (accurate count of population)

FOR BONUS POINTS!

- What study design would you use?
 - **Cross-sectional**

Prevalence

Number of persons in a population with a **disease** at a specified time

Number of persons in the **population** at the specified time

- Prevalence is a proportion of a population with the condition or disease at that time.
- Describes disease burden at the specified time. Presents a snapshot or slice-in-time. It does not describe events.
- Knowing the prevalence is useful for public health program planning

Numerator and Denominator

Both numerator and denominator are determined by the specific question asked

■ Numerator

- How will you define tuberculosis (Only pulmonary vs. all types? Only smear positive?)
- Will all groups be included (adults, children)?

■ Denominator

- Define at risk population
- Determine exclusions (? previous TB)

TB Prevalence Survey

You find that you can't answer the Minister's question and that you will need to collect data to provide the answer.

How will you go about it (if you want to keep your job)?

Determine the numerator (case definition)

- All persons with a positive sputum smear
 - Identification of those to evaluate
 - Timing of specimen
 - Direct vs. concentrated smear/fluorescent vs. conventional microscopy
 - Training and quality control
- All persons with a positive culture
 - Culture media
 - Number of colonies
- All persons with abnormal chest radiograph
 - Defining the abnormalities

Determine the denominator

- Total population
- A sample of the population
 - Ensuring the sample is representative (sample selection)
 - Adequate size
 - Stratify by
 - Race/ethnicity
 - Rural-urban
 - Age structure
 - Economic status
 - Method for random selection

Population Selection



Random:
Select row
and
seat number
at random

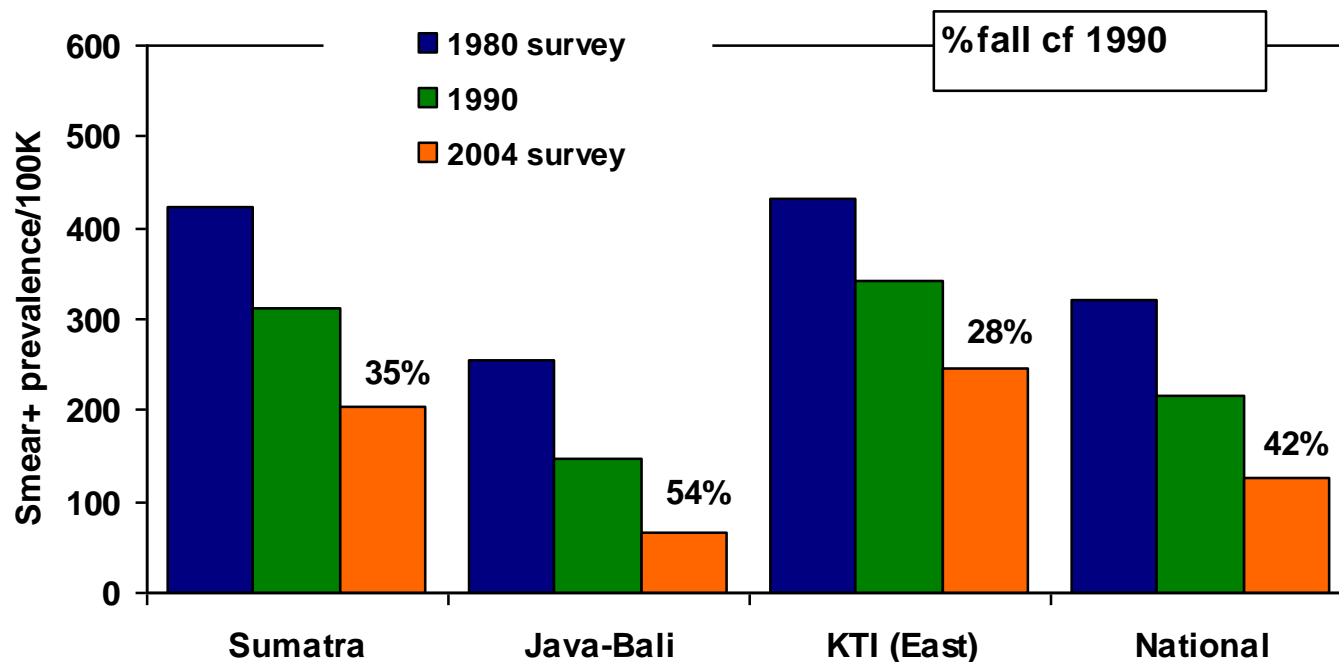
Stratified random:
Select from those
With white hats

Systematic:
Select every
10th person

Cluster: Select
groups of
several
rows
and seats

Convenience:
Select those
near entrance

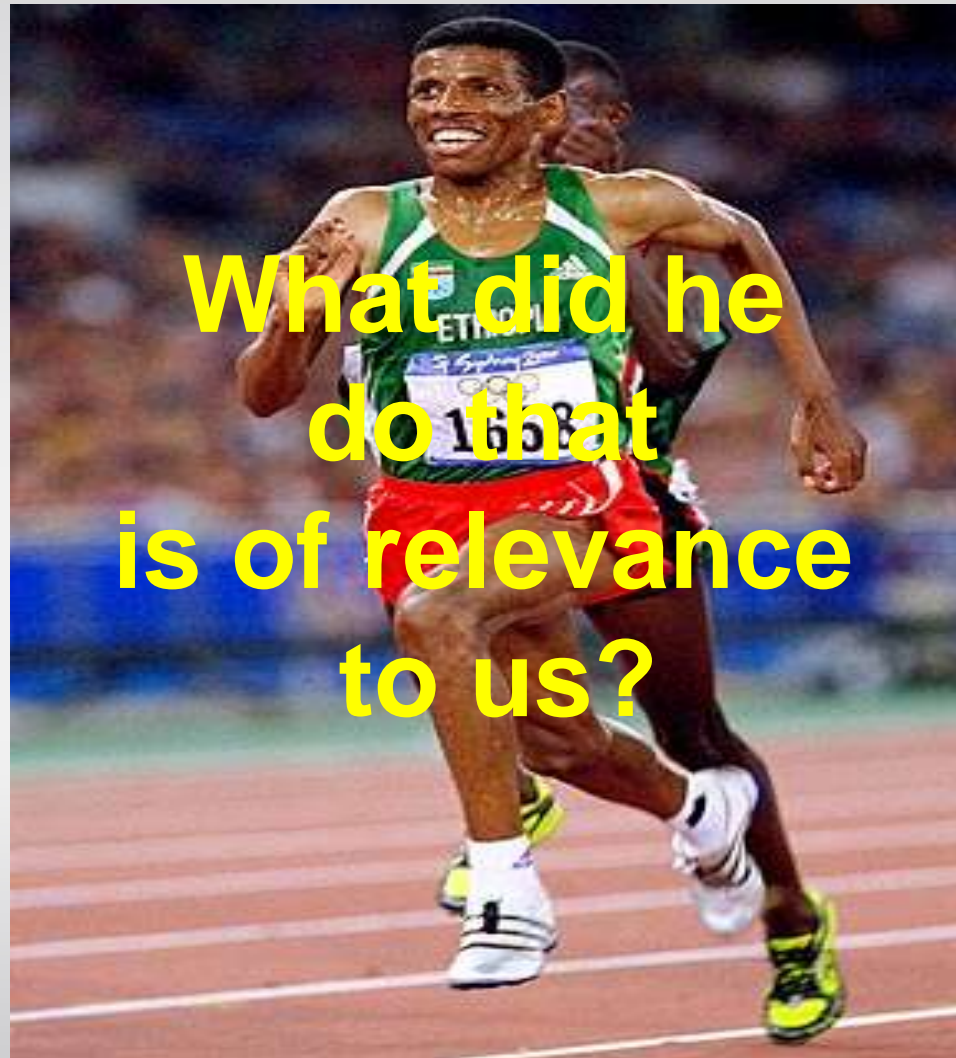
TB Prevalence Survey, Indonesia



Prevalence: Summary

- Useful to estimate disease burden (impact)
- Does not discriminate between new “events” vs people who already have the disease
- Dependent on the inflow (new cases) as well as the outflow of cases (death, or cure)
- Often requires an active process (survey)
- Not very useful when the objective is to understand risk factors

Who is this?



What did he
do that
is of relevance
to us?

Question

You are the Director of the Division of Communicable Diseases in the Ministry of Health and the Minister wants to be briefed on tuberculosis. How would you describe the magnitude of tuberculosis in the country?

How do you answer the Minister's question so that of tuberculosis so that she understands how frequently the disease develops and how well your program is performing?

Determining Incidence

- The standard statistic for conveying the frequency with which a disease develops in a population is its incidence.
- What data do you need to determine the incidence?
 - Numerator (number of new cases **in a given time**)
 - Denominator (defined population)

How will you determine the incidence?

- Numerator

- How will you define tuberculosis?
- Will all groups be included (adults, children)?

- Denominator

- Define at risk population
- Determine exclusions (? preexisting TB)

- Additional issues

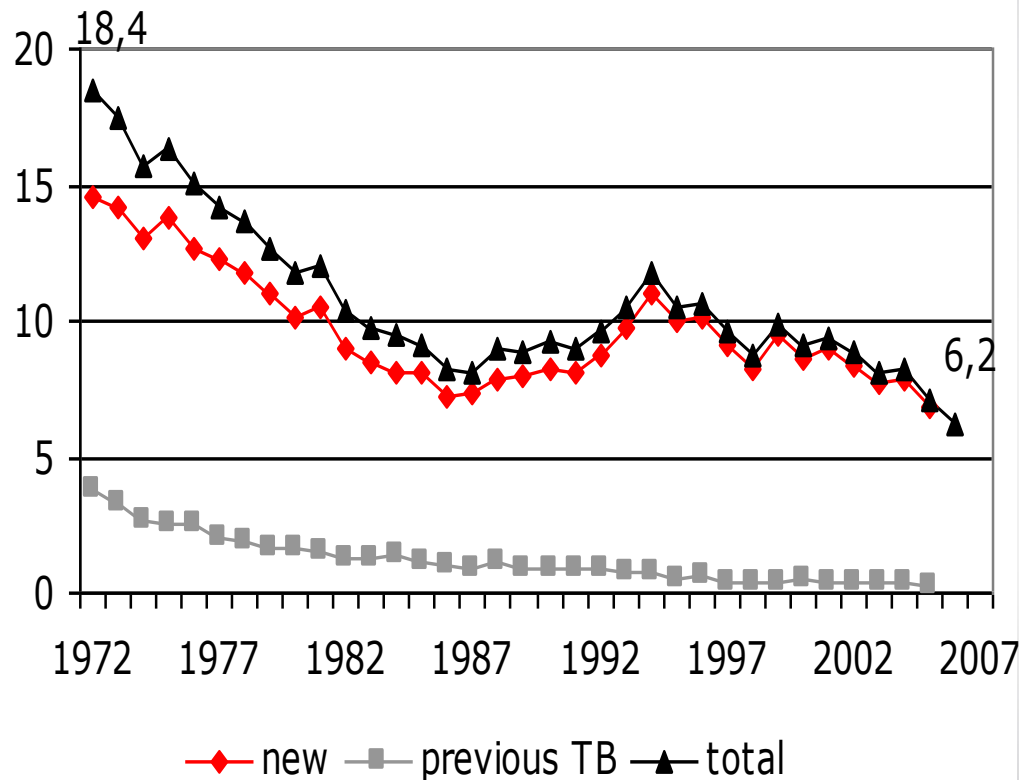
- Source of data
- Quality and completeness of data



Passive reporting

TB incidence: The Netherlands

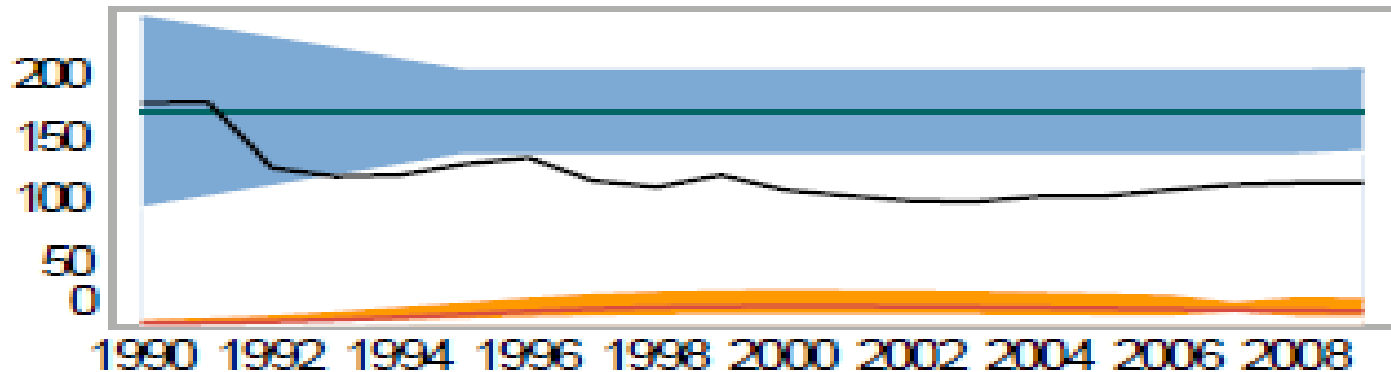
TB incidence per 100.000 1972-2006



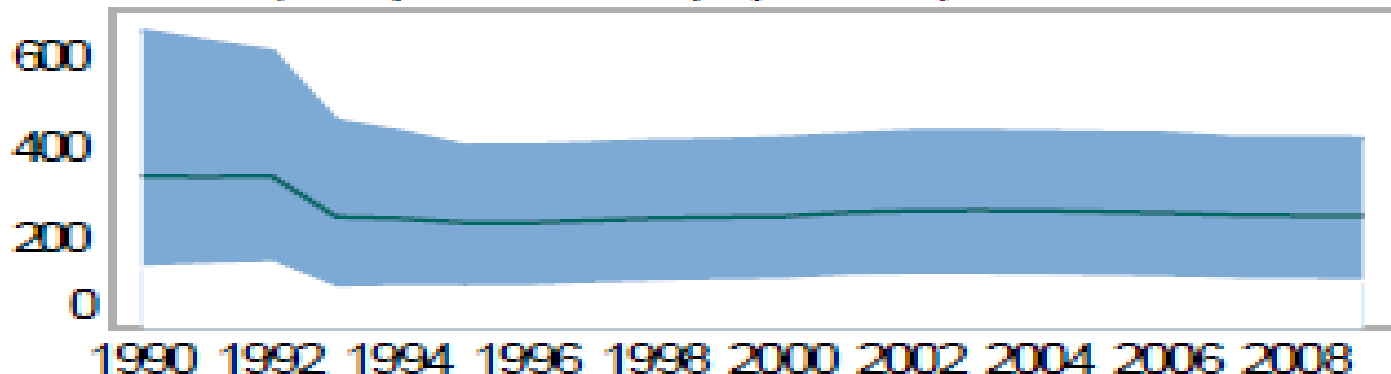
**TB incidence
declined 66%
between 1972
and 2007**

TB Incidence and Prevalence, India, 1990-2008

Incidence (HIV+TB in orange), notifications (black)
(rates per 100 000 population)



Prevalence (rate per 100 000 population)



Incidence

Number of **new cases** of a disease in a defined **population** during a specified **period**

Number of **people at risk** of developing the disease during that **period of time**

Example: A country's Tb incidence is usually expressed as cases/100,000 total population/year

Incidence

- Incidence is a rate and infers risk
- People at risk, not necessarily the entire population, make up the denominator
- New (incident) cases
- Time is important
- Usually relies on passive reporting, BUT uses an established surveillance system and standardized definitions

At Risk

- What does it mean to be at risk?
 - Person must be susceptible.
- Who is not at risk?
 - Person who has (or has had) the disease
 - Person who cannot develop the disease
 - Genetic makeup
 - Immunity
- Example: Incidence of tuberculous infection (new positive tuberculin skin test) in hospital staff. Who is the at-risk population?

Variations on Incidence

Incidence Density

Number of events during a specified period

Person – time denominator (adding all time periods)

- The denominator is calculated by adding the time periods each individual is followed from beginning until the event or are not followed any longer
- time unit is arbitrarily set by the researcher
 - 100 person-years: 100 persons followed for 1 year,
50 persons followed for 2 years,
200 persons followed for 6 months

Variations on Incidence

Cumulative Incidence

Number of events followed for a specified period

Fixed population followed during the same period of time that events occurred

- Estimates the probability that an individual will develop a disease in a specified period of time
- A measure of average risk

Relationship of Incidence and Prevalence

Prevalence = Incidence of disease X duration

- Incidence of drug susceptible TB = 250/100,000/yr
- Duration = 1 year
- Prevalence = 250/100,000
- Incidence MDR TB = 5/100,000/yr
- Duration = 3 years
- Prevalence = 15/100,000

Hailie Geberselassie



Relative Risk

- Risk - the proportion of **participants experiencing the event** of interest. If in a sample of 100, 10 patients relapsed and 90 patients did not, the risk of relapse is $10/100$ (0.10)
- Relative risk (RR) - ratio of risks in 2 groups

How successful is treatment?

Your Minister of Health asks, “How are we doing with treatment? Are there any problems?”

How do you answer the question?

A. “I haven’t a clue.”

SACKED immediately

B. “I think we have about a 98% cure rate.”

SACKED after Minister presents in international meeting

D. “I need to ask my statistician.”

Statistician promoted to director to replace you

C. “We have done a **retrospective cohort review** and have found that—”

E. “We have done a **prospective cohort study** and we have found that—.”

JOB SECURITY!



Risk of TB relapse

--- more patients with HIV infection relapse within 1 year of completion of treatment compared with patients who do not have HIV infection. We have quantified the difference as follows:—”

Relative Risk for TB Relapse

- HIV + Group: N = 786
 - Relapse 94
- HIV - group: N = 522
 - Relapse -20

Study design- Prospective cohort

The Key to All Knowledge (according to Jerry)

THE 2X2 TABLE!!!

		Outcome (Disease)	
		+	-
Exposure (Risk Factor)	+		
	-		

BEHOLD!

THE TRUTH!

Relative Risk of TB Relapse

Exposure \ Outcome	Relapse	No Relapse
HIV +	A 94	B 692
HIV -	C 20	D 502

A+B
786

C+D
522

A+C 114

B+D 1194

A+B+C+D
1308

Relative Risk for TB Relapse

- HIV + Group: N = 786
 - Risk of Relapse $94/786 = 12\%$
- HIV - group: N = 522
 - Risk of Relapse $20/522 = 4\%$
- Relative risk $94/786 / 20/522 = 3.0$

Interpreting Relative Risks

RR = 1 Risk in exposed is equal to the risk in the unexposed (no association)

RR > 1 Risk in the exposed is greater than the risk in the unexposed (+ association)

RR < 1 Risk in exposed is less than the risk in the unexposed (- association)

Odds Ratio

- Odds - a way of expressing the chance of an event, by dividing the **number of individuals in a sample who experienced the event by the number who did not**. If in sample of 100, 10 patients relapsed and 90 patients did not, the odds of relapse are 10/90 (1:9 or 0.11)

$$n / N-n$$

- Odds ratio (OR) - the ratio of the odds of an event in one group to the odds of an event in another group

$$n / N-n \bigg/ n / N-n$$

How successful is treatment?

Your minister of health asks, “How are we doing with treatment? Are there any problems?”

How do you answer the question?

Your answer, “We have done a **retrospective cohort review** and have found that more patients with HIV infection relapse within 1 year of completion of treatment compared with patients who do not have HIV infection. We have quantified the difference as follows:—”

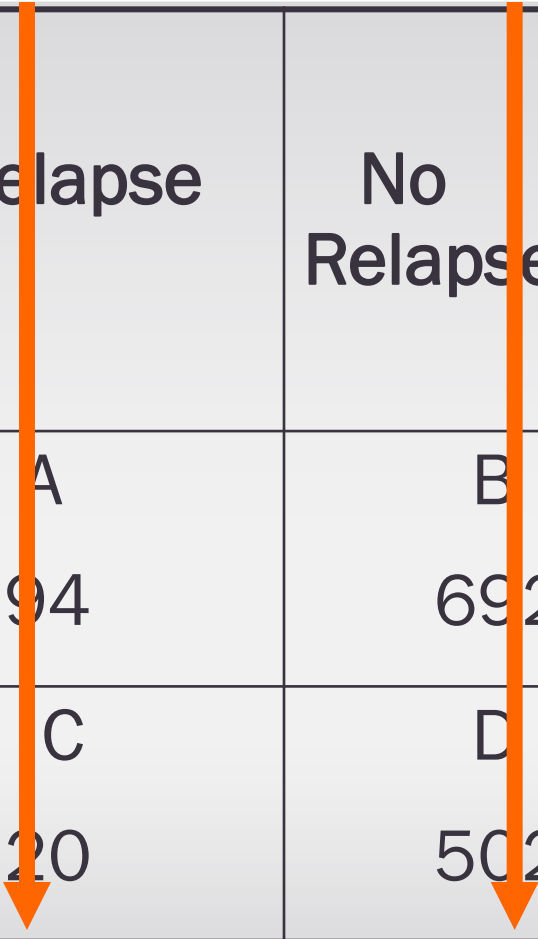
Odds Ratio for TB Relapse

- HIV + Group: $N = 786$
 - Relapse- $n=94$
- HIV - group: $N = 522$
 - Relapse - $n=20$

Study design- Case control (What's wrong here?)

Odds Ratio for TB Relapse

Exposure	Relapse	No Relapse
Outcome		
HIV +	A 94	B 692
HIV -	C 20	D 502



A+B
786

C+D
522

A+C 114

B+D 1194

A+B+C+D
1308

Odds Ratio for TB Relapse

- HIV + Group: N = 786

- Relapse - 94

- No relapse - 692

Odds 14%

- HIV- group: N = 522

- Relapse - 20

- No Relapse - 502

Odds 4%

When the risk is small, OR ~ RR

Interpreting Odds Ratios

OR = 1 Odds of being exposed is equal among individuals **with** and **without** disease (no association)

OR > 1 Odds of being exposed is greater among individuals **with** disease than among individuals **without** disease (+ association)

OR < 1 Odds of being exposed is less among individuals **with** disease than among individual **without** disease (- association)

Limitations of Relative Risk and Odds Ratios

Strong statistical associations between outcome and markers do not necessarily imply that the marker can discriminate between persons likely to have the outcome and those who do not.

The accuracy of a binary marker (HIV+ or HIV-) in predicting outcome is better summarized by the sensitivity and specificity of the marker for the outcome.