Measures of disease frequency

Madhukar Pai, MD, PhD

McGill University, Montreal

Email: madhukar.pai@mcgill.ca

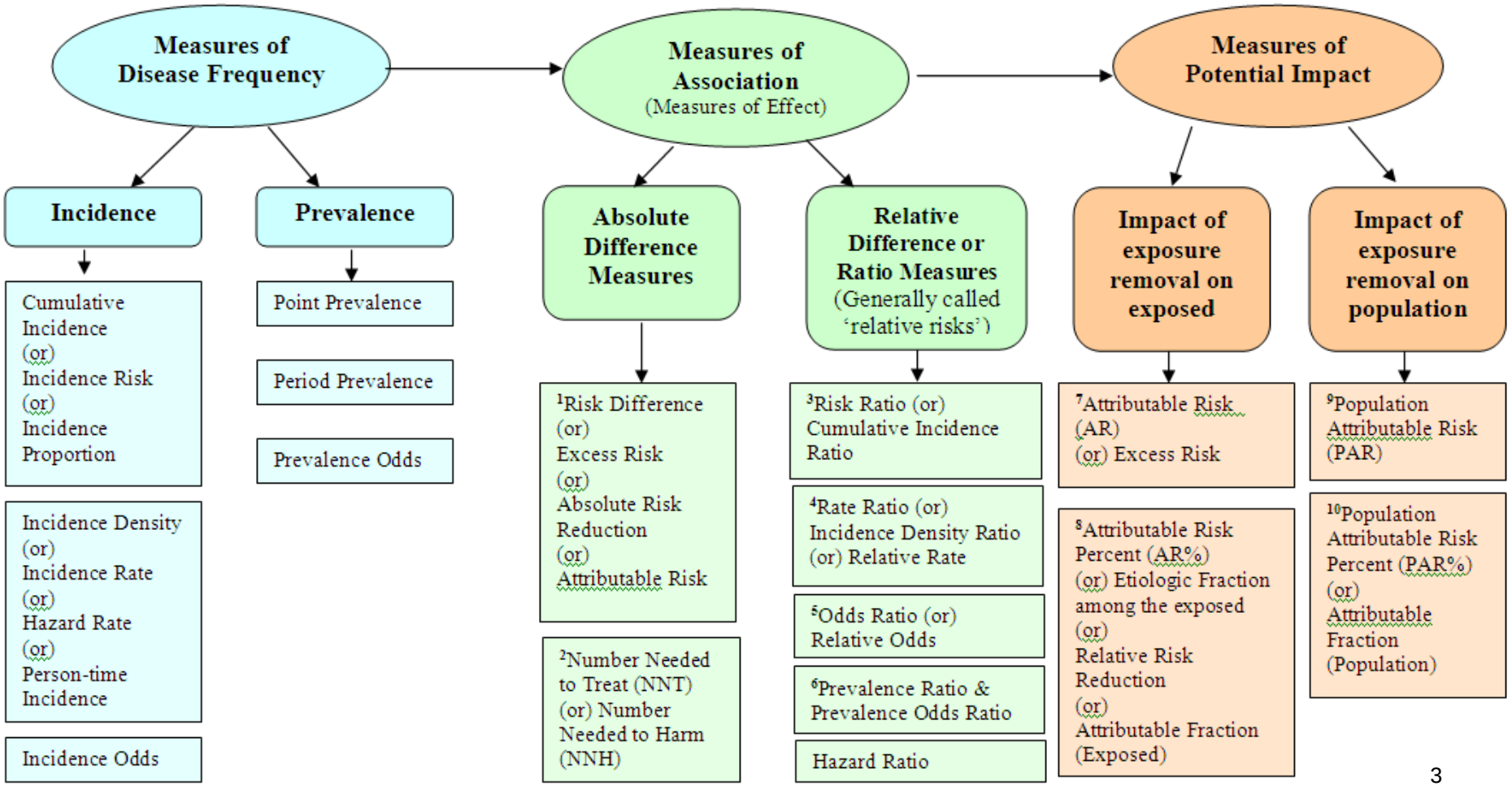
Overview

- Big picture
- Measures of Disease frequency
- Measures of Association (i.e. Effect)
- Measures of Potential Impact

AN OVERVIEW OF MEASUREMENTS IN EPIDEMIOLOGY [VER 3, 2007]

EXPOSURES ? → OUTCOME

Epidemiology is about identifying associations between exposures and outcomes. To identify any association, exposures and outcomes must first be measured in a quantitative manner. Then rates of occurrence of events are computed. These measures are called "*measures of disease frequency.*" Once measured, the association between exposures and outcomes are then evaluated by calculating "*measures of association or effect.*" Finally, the impact of removal of an exposure on the outcome is evaluated by computing "*measures of potential impact.*" In general, measures of disease frequency are needed to generate measures of association, and both these are needed to get measures of impact. There is some overlap between these measures, and terminology is poorly standardized.



Measures of Disease Frequency

- The importance of understanding the “numerator” and the “denominator” [proportions, rates, ratios]
  - Defining the numerator [“case”]
  - Defining the denominator [“population at risk”]
- Quantifying occurrence is usually done using:
  - Incidence
  - Prevalence
- Incidence:
  - Cumulative incidence [incidence risk, incidence proportion]
  - Incidence density [incidence rate; sometimes hazard rate]
- Prevalence:
  - Point prevalence
  - Period prevalence

Rates, Ratios, Proportions

- Three general classes of mathematical parameters.
- Often used to relate the number of cases of a disease [numerator] or health outcome to the size of the source population [denominator] in which they occurred.
- Numerator (“case”) has to be defined
- Denominator (“population size”) has to be defined
  - Epidemiologists have been referred to as “people in search of the denominator”!

Ratio

- Obtained by dividing one quantity by another. These quantities may be related or may be totally independent.
- Usually expressed as: $\frac{x}{y} \times 10^n$

Example: Number of stillbirths per thousand live births.

$$\frac{\#\,\text{stillbirths}}{\#\,\text{live births}} \times 1000$$

- "Ratio" is a general term that includes Rates and Proportions.
- Dictionary: "The value obtained by dividing one quantity by another." [Porta 2008]

Proportion

- A ratio in which the numerator (x) is included in the denominator (y)
- Expressed as: $\frac{x}{y} \times 10^n$ where, 10^n is often 100.

Example: The number of fetal deaths out of the total number of births.

$$\frac{\#\ \text{of fetal deaths}}{\text{live births} + \text{fetal deaths}} \times 100$$

- Answer often read as a percent.
- Dictionary: "A type of ratio in which the numerator in included in the denominator." [Porta 2008]

Rate

- A measure of how quickly something of interest happens.
- Expressed as: $\frac{x}{y} \times 10^n$

Example: The number of new cases of Parkinson's disease which develops per 1,000 person-years of follow-up.

$$\frac{\#\text{ of new cases of Parkinson's disease}}{\text{Total time disease-free subjects observed}} \times 1000$$

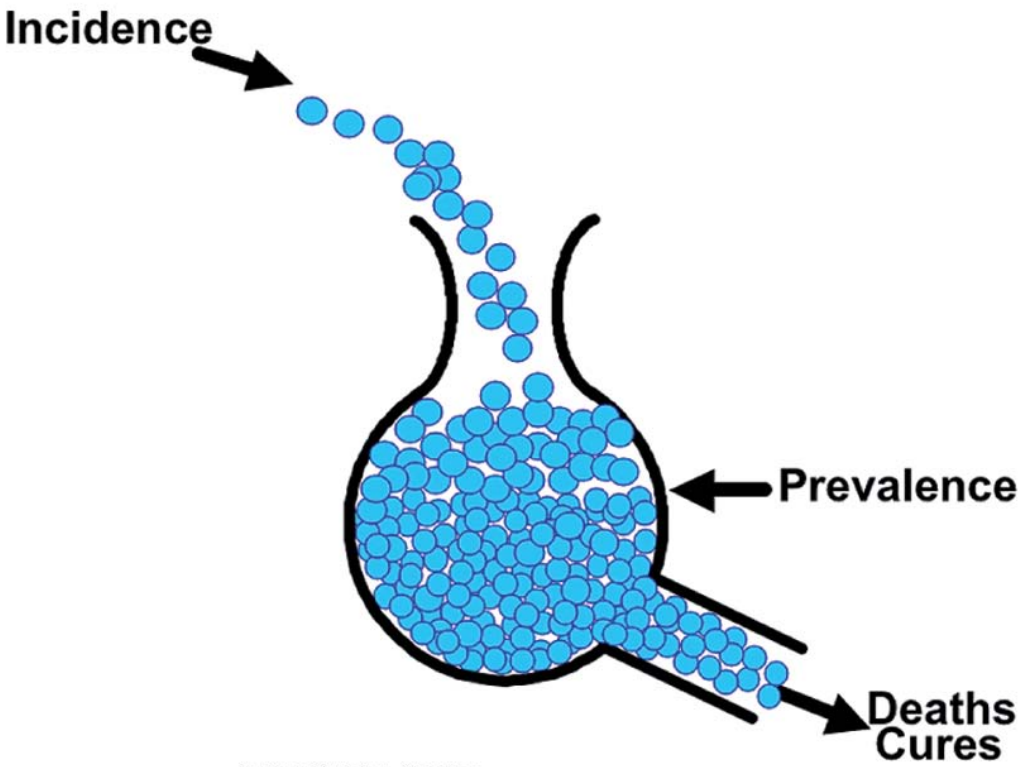
- Time, place and population must be specified for each type of rate.
- In a rate, numerator is not a subset of the denominator
- Rate is not a proportion

Measures of Disease Frequency

- Incidence (I): Measures new cases of a disease that develop over a period of time.
- Dictionary: "The number of new health-related events in a defined population within a specified period of time. May be measured as a frequency count, a rate or a proportion." [Porta 2008]
  - Very helpful for etiological/causal inference
  - Difficult to estimate
  - Implies follow-up over time (i.e. cohort design)

- Prevalence (P): Measures existing cases of a disease at a particular point in time or over a period of time.
- Dictionary: "A measure of disease occurrence: the total number of individuals who have an attribute or disease at a particular time (or period) divided by the population at risk of having the disease at that time or midway through the period. It is a proportion, not a rate." [Porta 2008]
  - Very helpful for quantifying disease burden (e.g. public health)
  - Relatively easy to estimate
  - Implies a cross-sectional design

Prevalence vs. Incidence



Gordis: Epidemiology, 4th Edition.

Relationship between incidence and prevalence: IV.

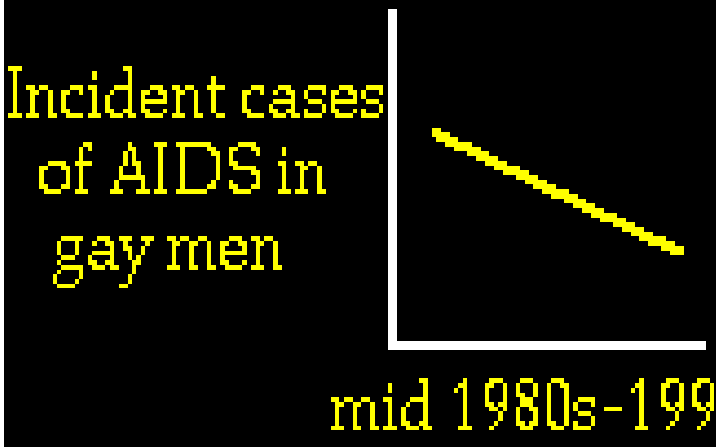
- Prevalence can be viewed as describing a pool of disease in a population.
- Incidence describes the input flow of new cases into the pool.
- Deaths and cures reflects the output flow from the pool.

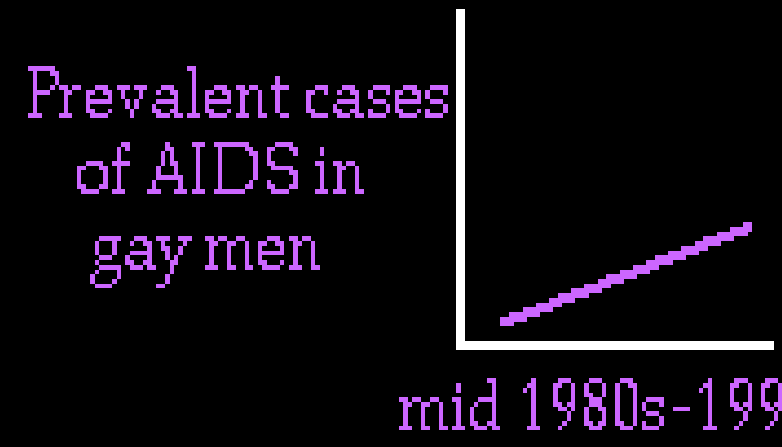
Measures of Disease Frequency

Incidence (I): New

Prevalence (P): Existing

AIDS:





- Anti-retroviral treatment
- Reduce high risk behavior

- Treatments prolong life

Risk

Probability that an individual with certain characteristics such as:

Age

Race

Sex

Smoking status

will experience a health status change over a specified follow-up period (i.e. risk period)

Dictionary: “Probability that an event will occur within a stated period of time.” [Porta 2008]

Assumes:

Does not have disease at start of follow-up.

Does not die from other cause during follow-up (no competing risks).

Risk is often used for prediction at the individual level

Risk

$0 \leq RISK \leq 1$

$0\% \leq \text{percentage} \leq 100\%$

Specify risk period

Example: The 10-year risk that a 45-year-old male will develop prostate cancer is 5%.

Risk can be estimated from:

1. Cumulative Incidence (directly) [note: cumulative incidence is also called “incidence risk”]

2. Incidence density (indirectly via life tables, etc)

Cumulative Incidence

$$CI = \frac{I}{N}$$

I = # of new cases during follow-up

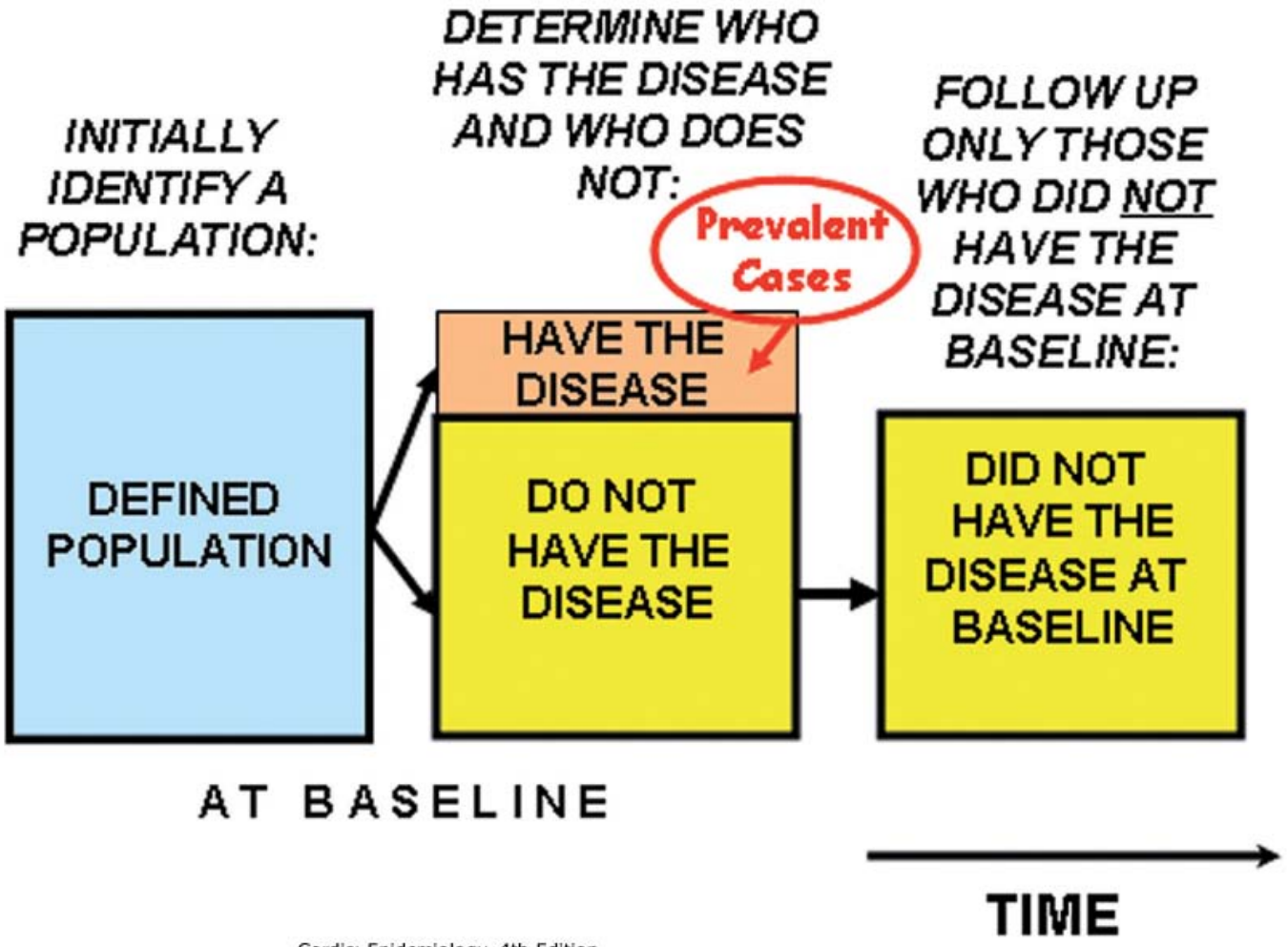
N = # of disease-free subjects at start of follow-up

Measures the frequency of addition of new cases of disease and is always calculated for a given period of time (e.g. annual incidence)

Must always state the time period (since time is not automatically captured in CI)

Dictionary: “Incidence expressed as a proportion of the population at risk. A measure of risk. The proportion of a closed population at risk for a disease that develops the disease during a specified interval.” [Porta 2008]

Prevalent cases must be excluded before follow-up



Example

N=1000 men age 45
I=50 developed prostate cancer
Follow-up = 10 years

No one lost to follow-up
No one withdrew from study

$$\hat{CI} = \frac{I}{N} = \frac{50}{1000} = 0.05 = 5\%$$

10-year risk:

$$\hat{CI} = \frac{I}{N} = 5\%$$

Cumulative Incidence

- Most common way to estimate risk
- Always a proportion (bounded between 0 and 1)
- Assumes a fixed or closed cohort (no exits allowed)
- For brief specified periods of time, e.g. an outbreak, commonly called an Attack “Rate”

- In reality, attrition is a huge problem (losses to follow-up, deaths, competing risks)
- Formula does not reflect continually changing population size for dynamic cohorts (open populations).
- Does not allow subjects to be followed for different time periods.

- In real life, one has to deal with losses, competing risks, attrition, dynamic cohorts, and differential follow-up time!!
- So, rate becomes more relevant

RATE

A measure of how quickly something happens.

Instantaneous Rate

Velocity: 65 mph

Average Rate

Speed: 55 mph average

In epidemiology, we generally measure “average rates” [instantaneous rates are difficult to estimate, but hazard function comes close]

Rate

- Describes how rapidly health events are occurring in a population of interest.

- In epidemiologic studies, we typically measure the average rate at which a disease is occurring over a period of time.
- Rate is always non-negative, but has no upper bound ($0 \leq \text{Rate} \leq \text{infinity}$)
- Rate is not a proportion bounded between 0 and 1

- Example:
  50 new cases per 10,000 person-years

  Interpretation:
  An average of 50 cases occurs for every 10,000 years of disease free follow-up time observed on a cohort of subjects.

  This type of rate is not easy to use for individual risk prediction [they are useful at the population level]

Dictionary: “The rate at which new events occur in a population.” [Porta 2008]

Incidence density (incidence rate)

$$IR = \frac{I}{PT}$$

I = # of new cases during follow-up

PT = total time that disease–free individuals in the cohort are observed over the study period (total person-time experience of the cohort).

Synonyms: hazard rate*, incidence density rate, person-time incidence.

Dictionary: “The average person-time incidence rate” [Porta, 2008]

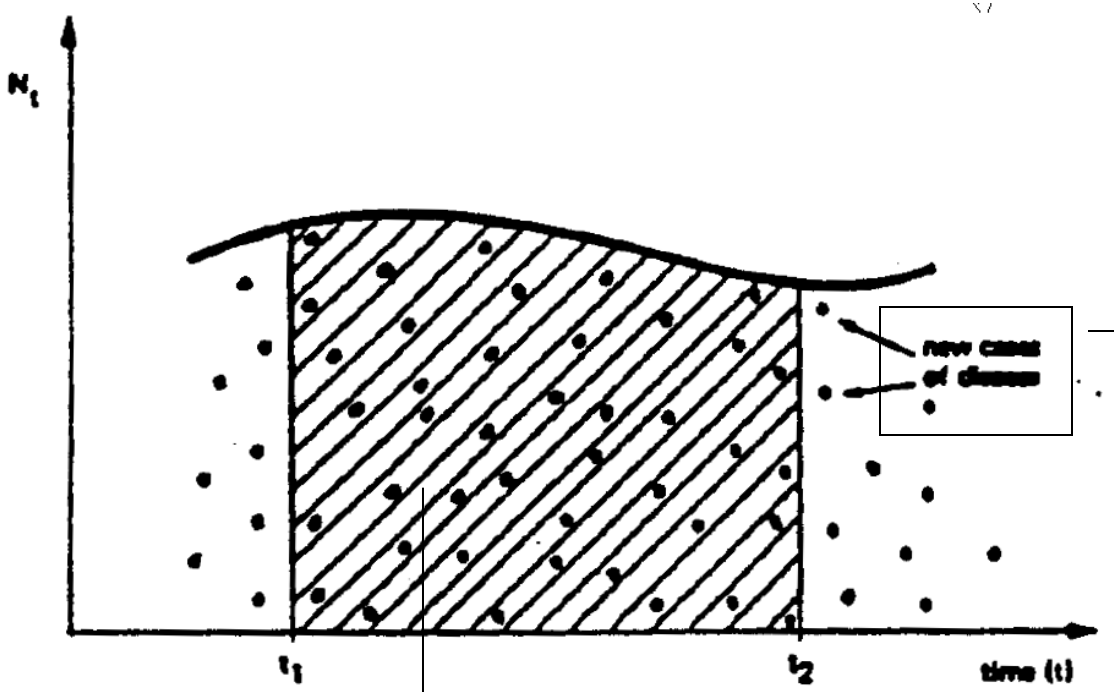
Measures the rapidity with which new cases are occurring in a population

Most sophisticated form of measuring incidence [most difficult as well]

- Accounts for losses, competing risks, dynamic turn-over, differential follow-up time, changes in exposures over time

- *hazard function (in survival analysis) is the event rate at time *t* conditional on survival until time *t* [hazard rate is something like an instantaneous rate]

Incidence “density”: new cases that occur in the total population-time



Numerator

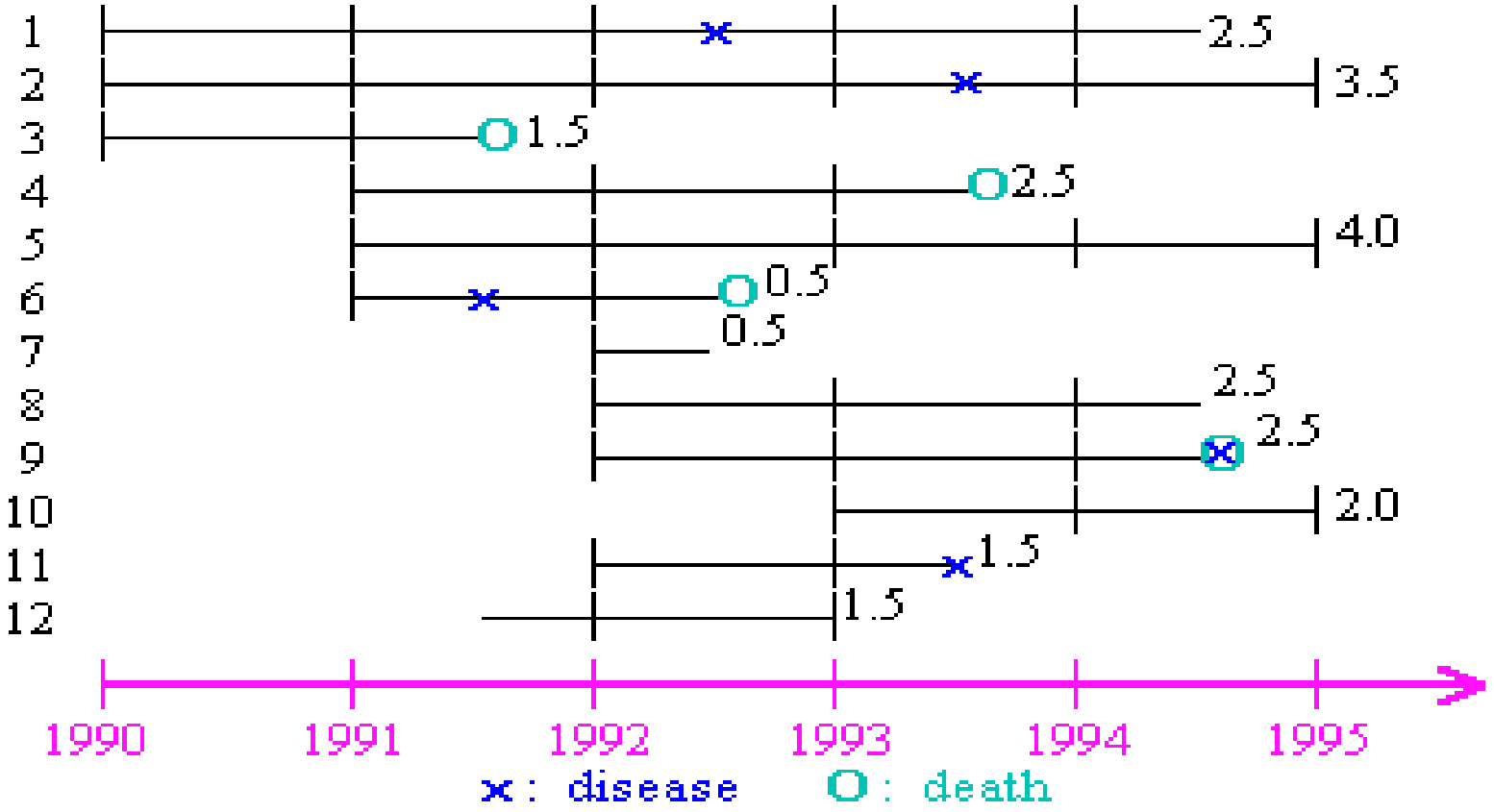
FIGURE 2 *Graphical illustration of the occurrence of new (incident) cases over time in a candidate population (of size N_t at time t)*

Denominator = Area under the curve = aggregate of person-moments [total population-time]

Morgenstern IJE 1980

Example

Hypothetical cohort of 12 initially disease-free subjects followed over a 5-year period from 1990 to 1995.



Example, cont.

$$\hat{IR} = \frac{I}{PT} = \frac{5}{25PY} = 0.20$$

$$= 20 \text{ new cases per } 100 \text{ person - years}$$

$$\text{or } 200 \text{ new cases per } 1000 \text{ person - years}$$

Study questions:

1) Is the value of 0.20 a proportion?
2) Does the value of 0.20 represent an individual's risk of developing disease?

Confusing Risk with Rate

- The term “Rate” is often been used incorrectly to describe a measure of risk (cumulative incidence).

  e.g.,

  - Attack Rate
  - Death Rate
  - Case-Fatality Rate

- When reading Epidemiologic literature, one should be careful to determine the actual measure being reported.

Risk vs Rate

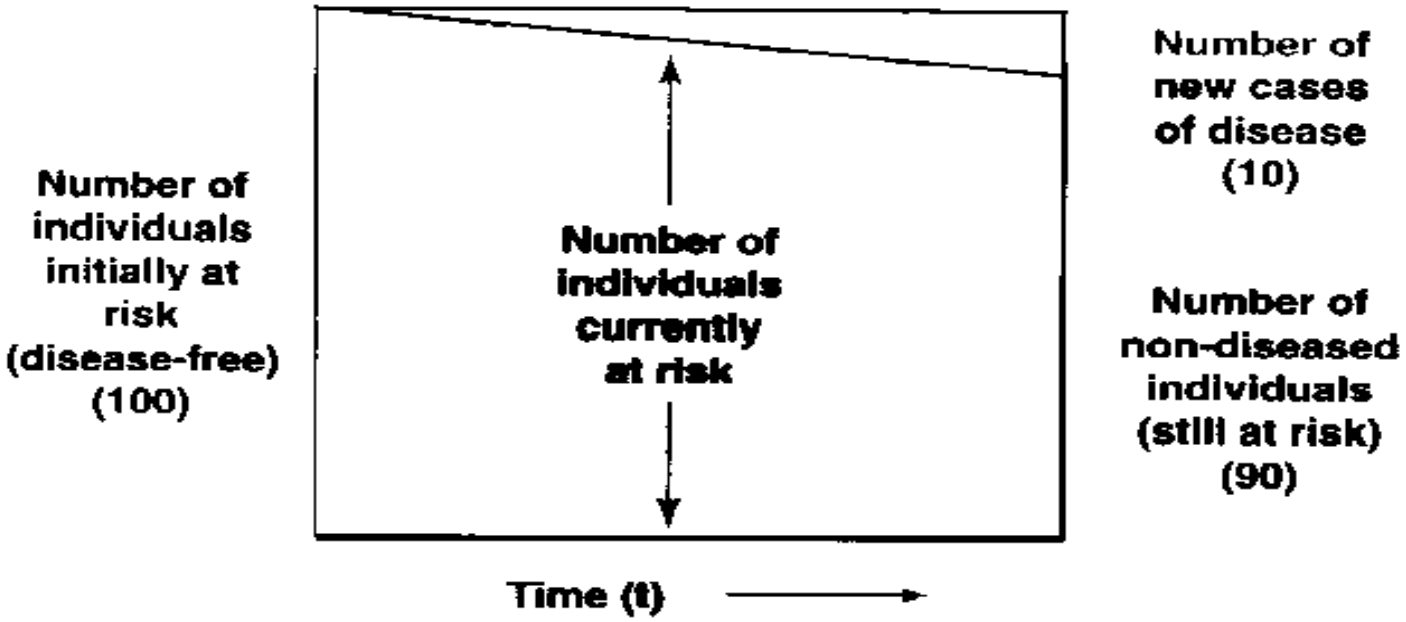
RISK

- E.g. Cumulative incidence
- Proportion (always between 0 and 1)
- Probability that an individual will develop a disease during a specific period
- Use for individual prognosis
- More assumptions
- Cannot handle variable follow-up times, attrition, competing risks
- Easy to compute in a fixed cohort with few losses; but gets difficult with open populations with longer follow up and losses

RATE

- E.g. Incidence density
- Non-negative and no upper bound
- Describes how rapidly new events occur in a specific population
- Use for etiological comparisons
- Fewer assumptions
- Can handle variable follow-up times, attrition, competing risks
- Can be computed even with open populations with losses and longer follow up

Risk vs Odds



Thus, it is possible to calculate the risk and the odds of developing the disease during the study period as:

Risk = 10/100 = 0.10 = 10%

Odds of disease = 10/90 = 0.11 = 11%

Dictionary: "Odds is the ratio of the probability of occurrence of an event to that of non-occurrence." [Porta, 2008]

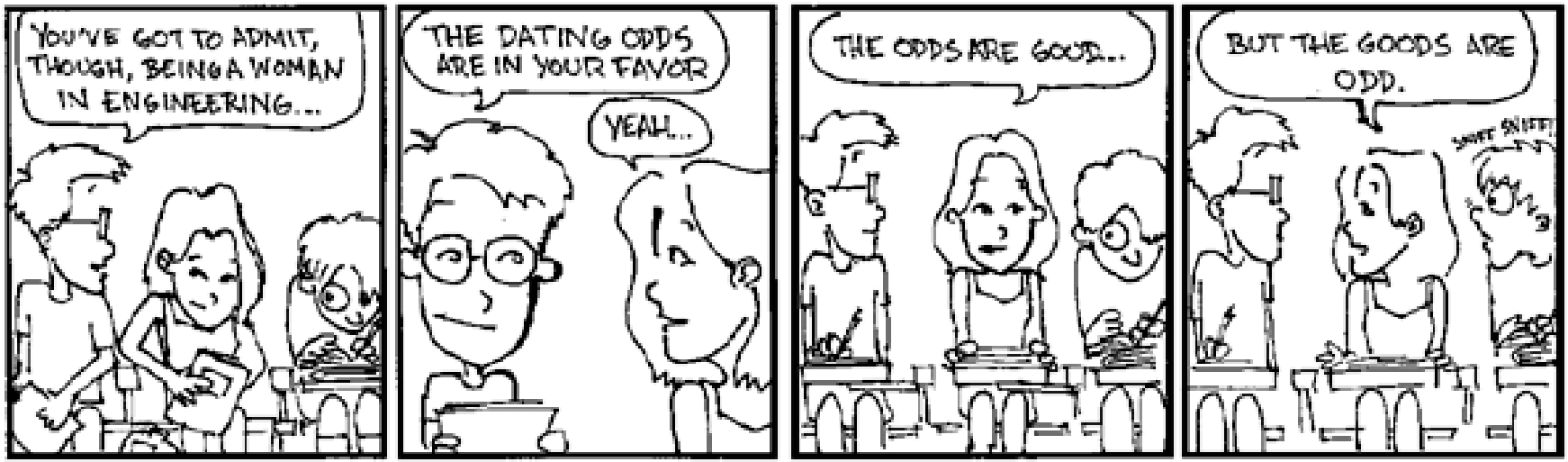
Source: Silva 1999

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

- To go from Probability to Odds:
  - Odds = P / (1 – P)
  - E.g. If P = 0.20, Odds = 0.20 / 0.80 = 0.25
- To go from Odds to Probability:
  - Probability = Odds / (1 + Odds)
  - E.g. If Odds = 0.25, P = 0.25 / 1.25 = 0.20


YOU'VE GOT TO ADMIT, THOUGH, BEING A WOMAN IN ENGINEERING...
THE DATING ODDS ARE IN YOUR FAVOR
YEAH...
THE ODDS ARE GOOD...
BUT THE GOODS ARE ODD.
SNIFF SNIFF!
JORGE CHAM ©THE STANFORD DAILY

Prevalence

- Measures existing cases of a health condition
  - Inherently biased towards inclusion of “survivors”
- Primary outcome of a cross-sectional study (e.g. sample surveys)
- Two types of Prevalence
  - Point prevalence
  - Period prevalence

Point Prevalence

$$P = \frac{C}{N}$$

C = # of observed cases at time t

N = Population size at time t

Measures the frequency of disease at a given point in time

Dictionary: “A measure of disease occurrence: the total number of individuals who have an attribute or disease at a particular time (or period) divided by the population at risk of having the disease at that time or midway through the period. It is a proportion, not a rate.” [Porta 2008]

Point Prevalence Example

Suppose there are 150 individuals in a population and, on a certain day, 15 are ill with the flu. What is the estimated prevalence for this population?

$$P = \frac{15}{150} = 10\%$$

Prevalence

Useful for:

- Assessing the health status of a population.
- Planning health services.
- Often the only measure possible with chronic diseases where incident cases cannot be easily detected (e.g. prevalence of hypertension)

Not very useful for:

- Identifying risk factors (etiology): confusion between risk factors for survival vs. risk factors for developing disease
- Makes no sense for conditions that are acute and short duration (e.g. diarrhea)

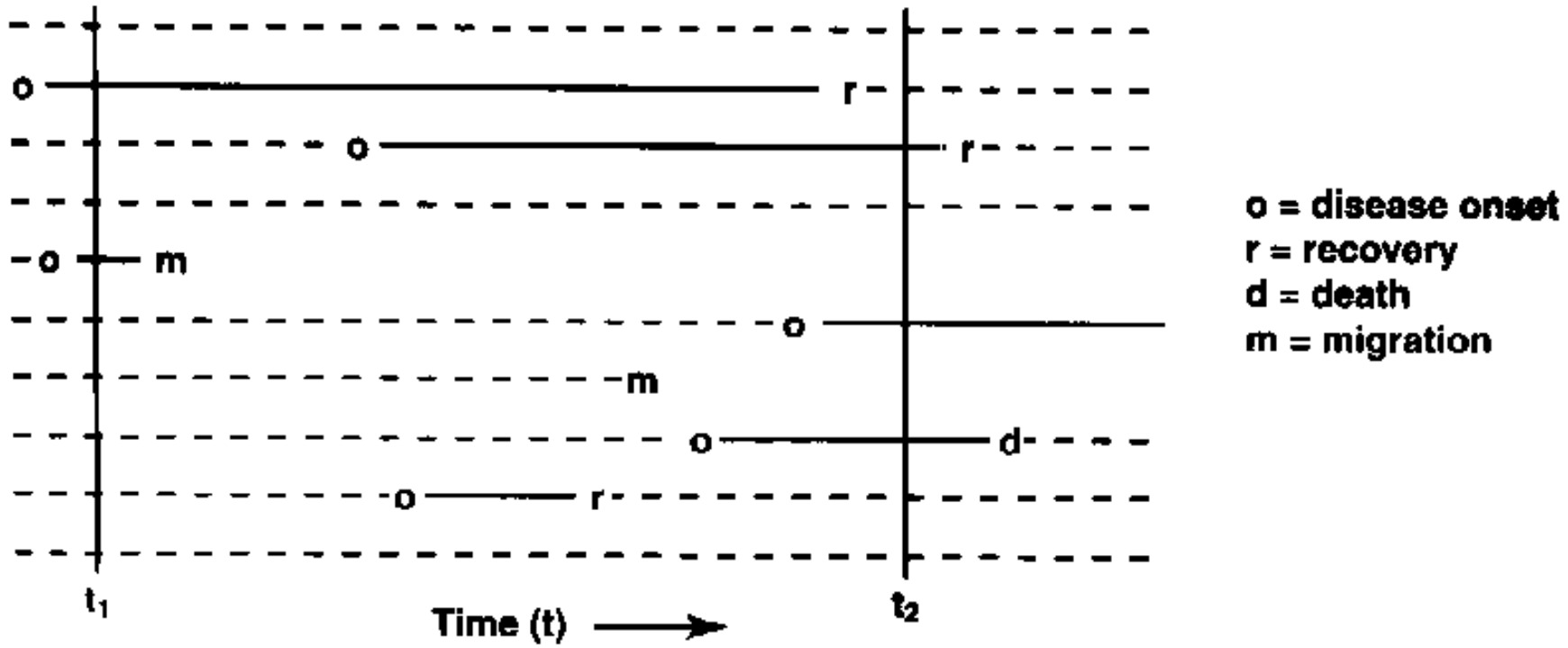
Period Prevalence

$$PP = \frac{C + I}{N}$$

- C = the # of prevalent cases at the beginning of the time period.
- I = the # of incident cases that develop during the period.
- N = size of the population for this same time period.

Example: one year prevalence: proportion of individuals with the disease at any time during a calendar year. It includes cases arising before and during the year. Denominator is total population during the time period.

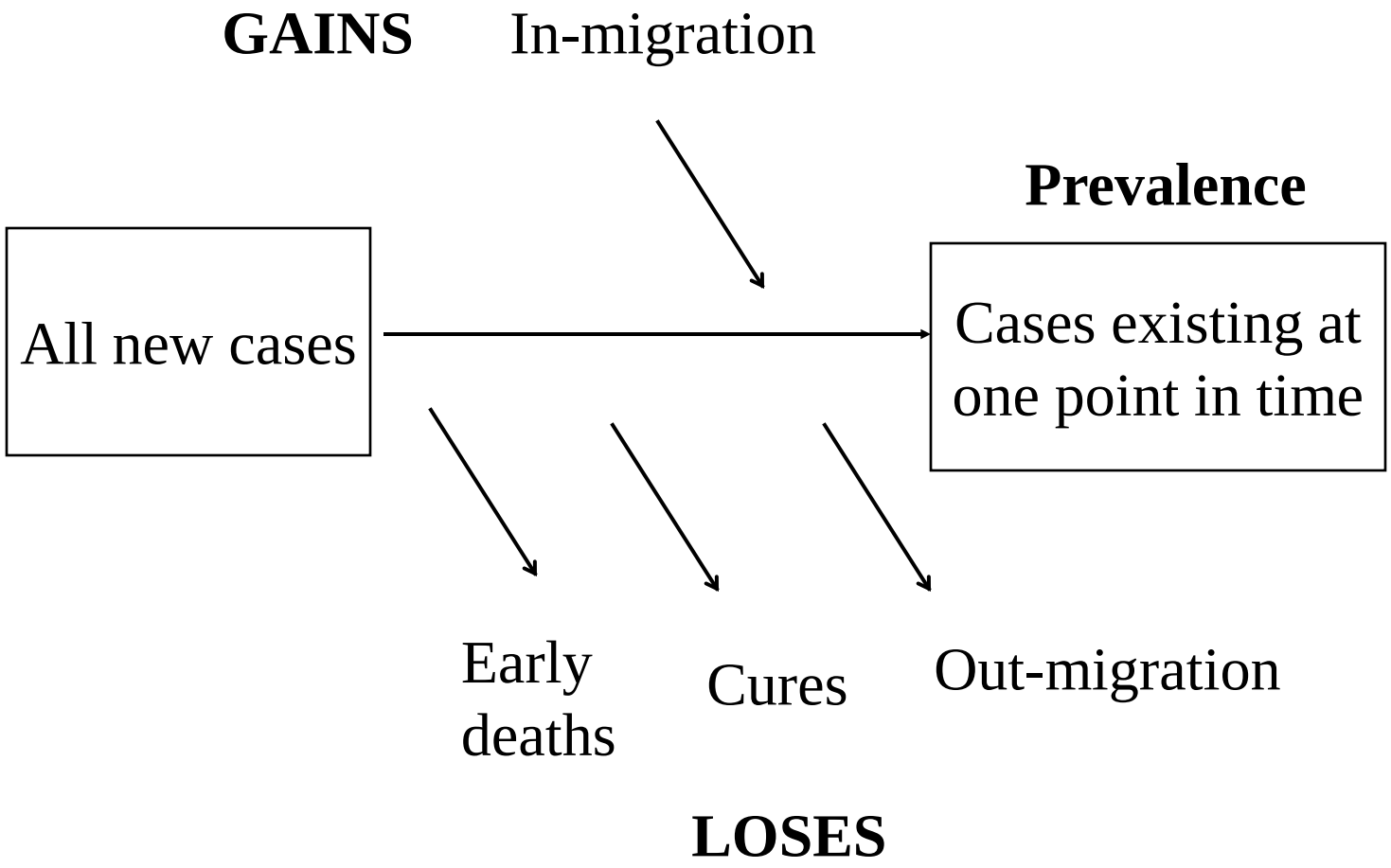
Prevalence: example



Point prevalence at time t_1 = 2/10 = 20%
Point prevalence at time t_2 = 3/8 = 38%
Period prevalence between t_1 and t_2: 6/10 = 60%

Source: Silva 1999

What impacts prevalence estimation?



What factors can increase prevalence?

Prevalence

Longer duration

Prolongation of life without cure

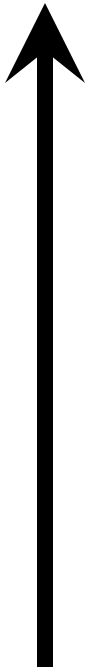
Increased incidence

In-migration of cases

Out-migration of healthy people

In-migration of susceptible people

Better diagnosis/reporting



Source: Beaglehole, 1993

What factors can decrease prevalence?

Prevalence

Shorter duration

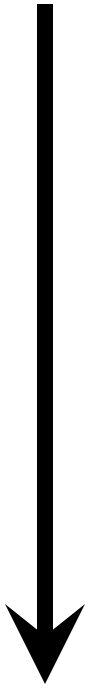
High case fatality

Decreased incidence

In-migration of healthy people

Out-migration of cases

Improved cure rates



Source: Beaglehole, 1993

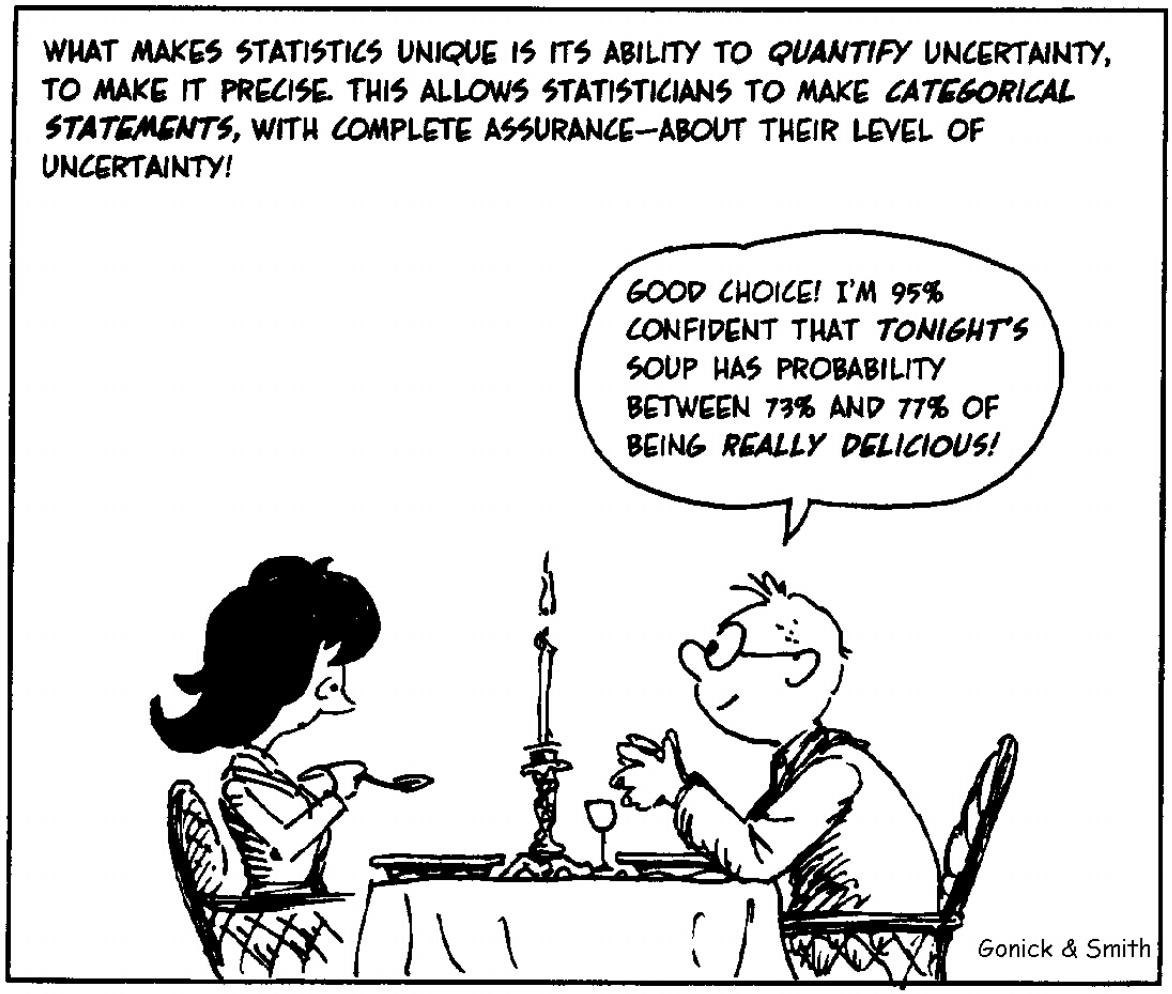
Relation between measures of disease frequency

- Relationship between prevalence and incidence:
  - Prevalence = Incidence Rate X Average Duration*

*Assumptions: steady state population and rare disease

- Relationship between incidence risk and rate:
  - Risk = Incidence Rate X Duration of the period of risk

Key issue to understand: all measures are “estimates” [subject to error]



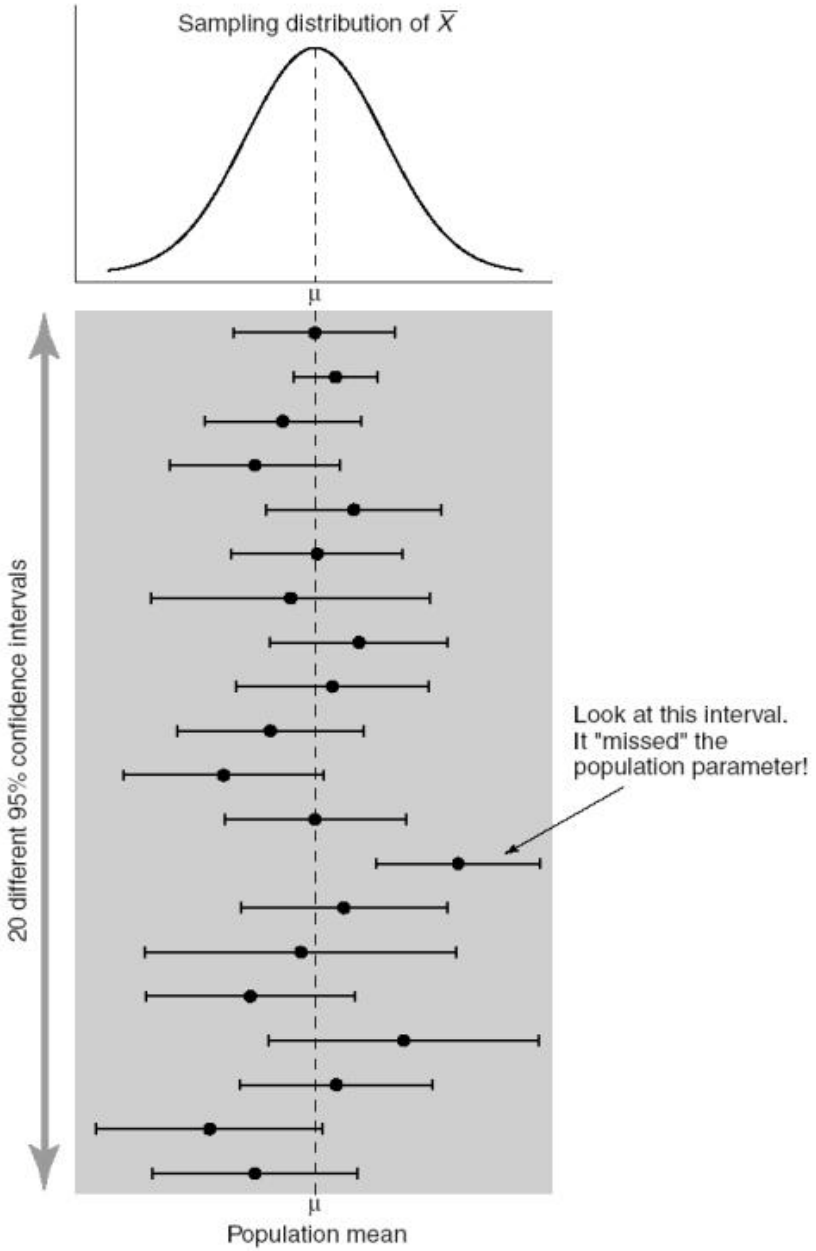


FIGURE 16.2 A sampling distribution of the mean (based on all possible samples of size 100) and an illustration of the 95 percent confidence intervals for twenty possible samples. The width of the intervals will be slightly different because they are estimated from different random samples. In the long run, 95 percent of confidence intervals will capture the population mean.

Therefore,
all estimates must
be reported
with a confidence
intervals

What are 95% confidence intervals?

- The interval computed from the sample data which, were the study repeated multiple times, would contain the true effect 95% of the time

- Incorrect Interpretation: "There is a 95% probability that the true effect is located within the confidence interval."
  - This is wrong because the true effect (i.e. the population parameter) is a constant, not a random variable. Its either in the confidence interval or it's not. There is no probability involved (in other words, truth does not vary, only the confidence interval varies around the truth).

Useful reading: Primer on 95% CI by American College of Physicians

Crude vs. adjusted rates

- Crude rates are useful, but not always comparable across populations
- Example: crude death rate in Sweden is higher than in Panama [Rothman text]
- Why?
- Confounding by age
- Age standardization is nothing but adjustment for confounding by age

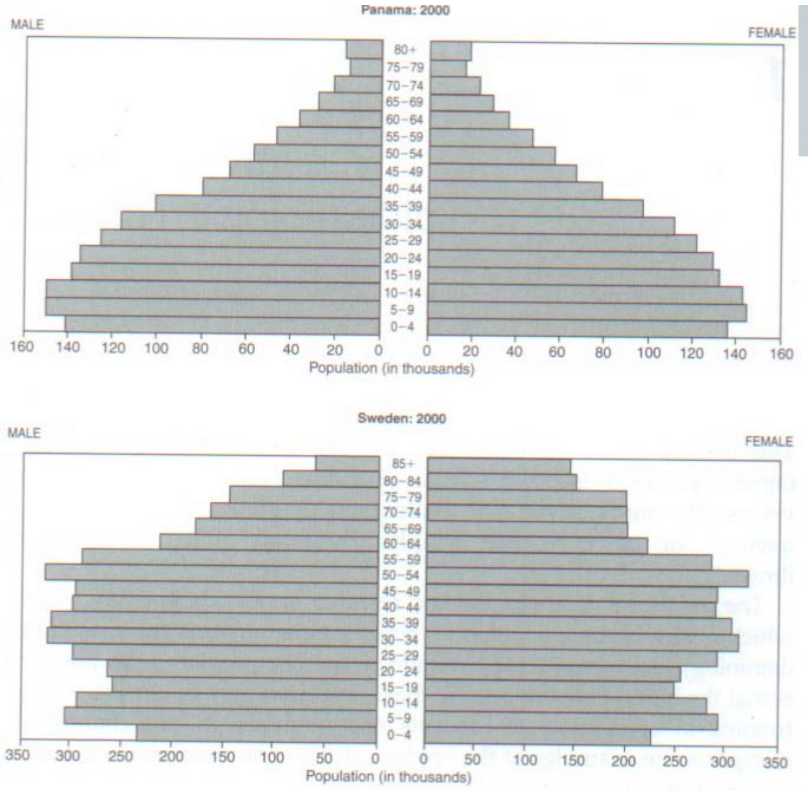
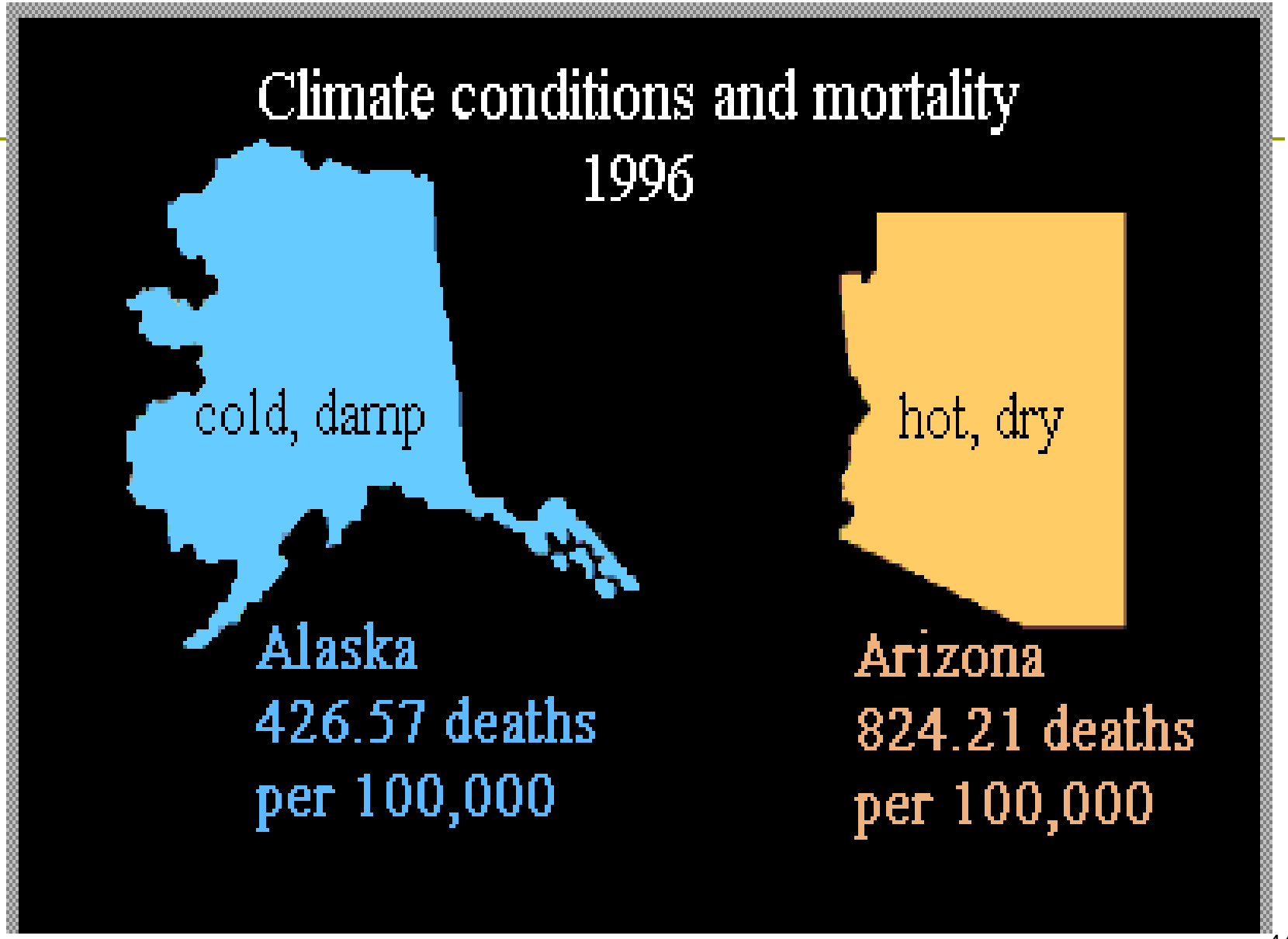


Figure 1–1. Age distribution of the populations of Panama and Sweden (population pyramids). Source: U.S. Census Bureau, International Data Base.

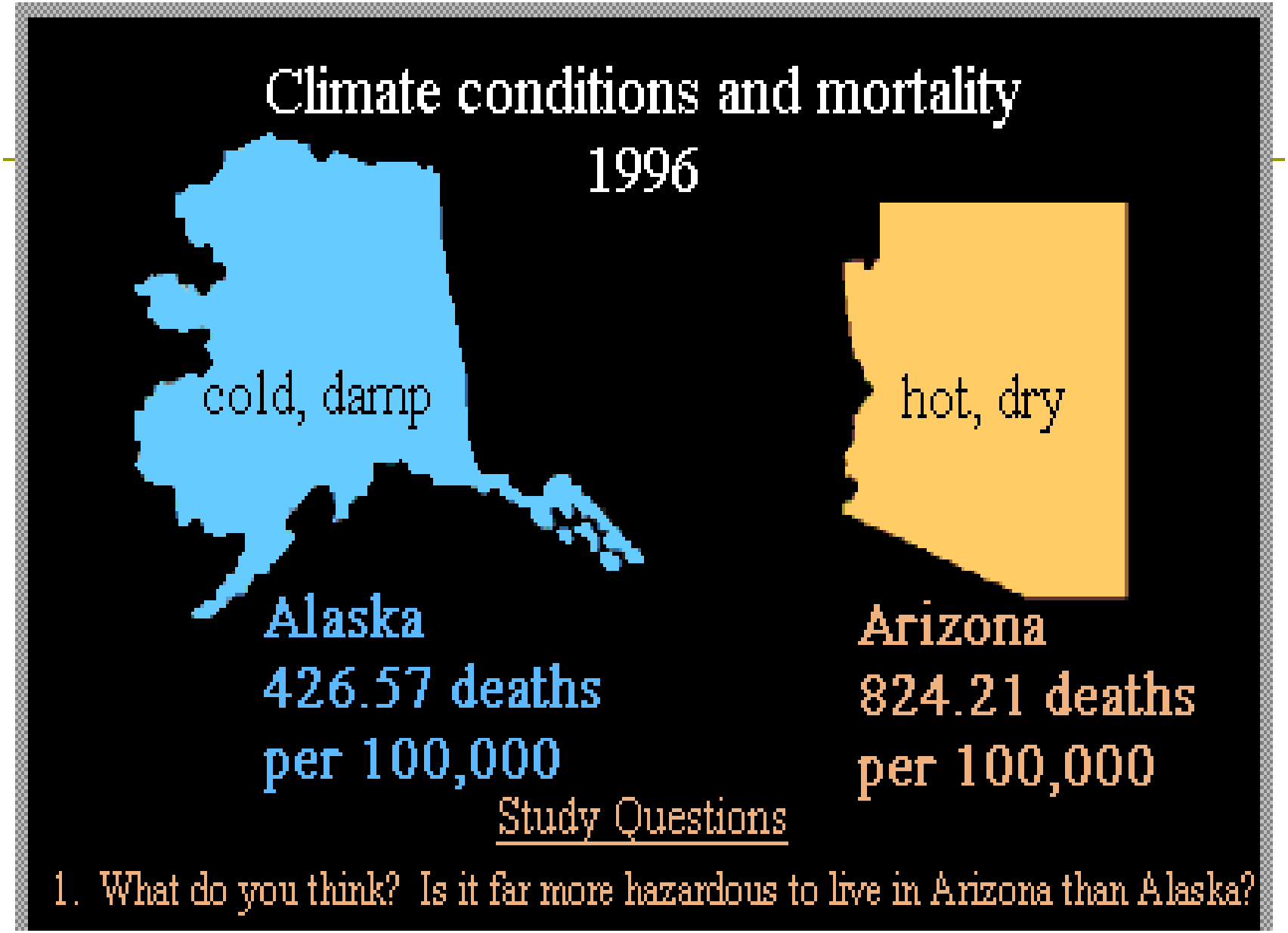
Rothman KJ, 2002

Example from Kleinbaum:



Kleinbaum et al. ActivEpi

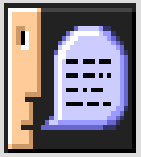
Example from Kleinbaum:



Example from Kleinbaum:

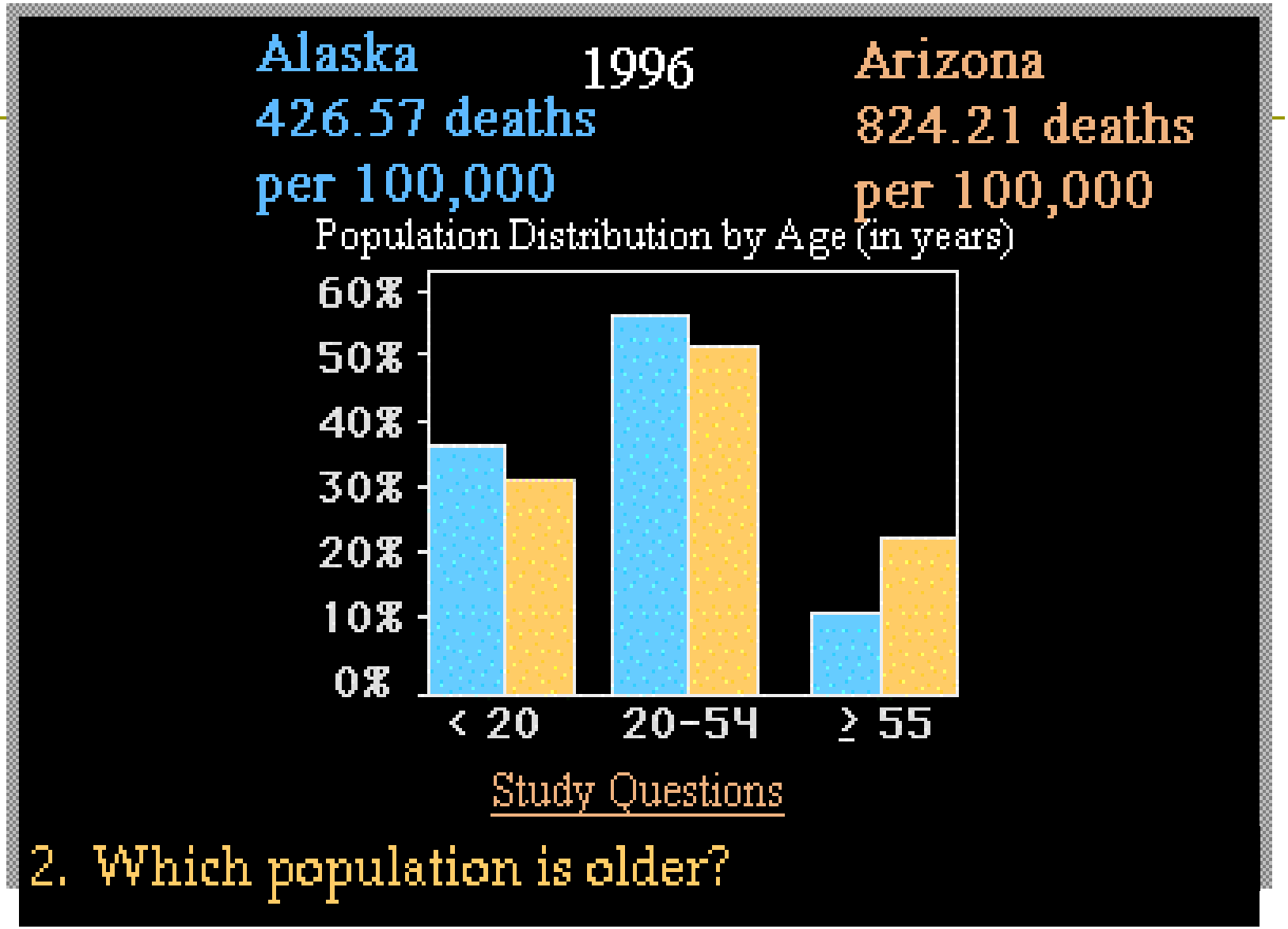
Study Questions

1. What do you think? Is it far more hazardous to live in Arizona than Alaska?



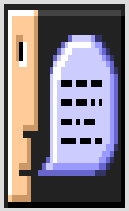
These two rates are crude rates because they represent the overall mortality experience in 1996 for the entire population of each state. Crude rates do not account for any differences in these populations on factors such as age, race or sex that might have some influence on mortality. Without consideration of such factors, it would be premature to make such a conclusion.

Example from Kleinbaum:



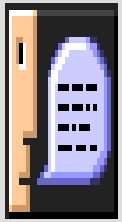
Example from Kleinbaum:

2. Which population is older?



Arizona. The dry, warm climate of Arizona attracts many older persons than does Alaska.

3. Why should we expect relatively more deaths in Arizona than in Alaska?



There are relatively more older persons living in Arizona, and older persons are at high risk of dying.

Example from Kleinbaum:

Alaska	1996	Arizona
426.57 deaths per 100,000		824.21 deaths per 100,000

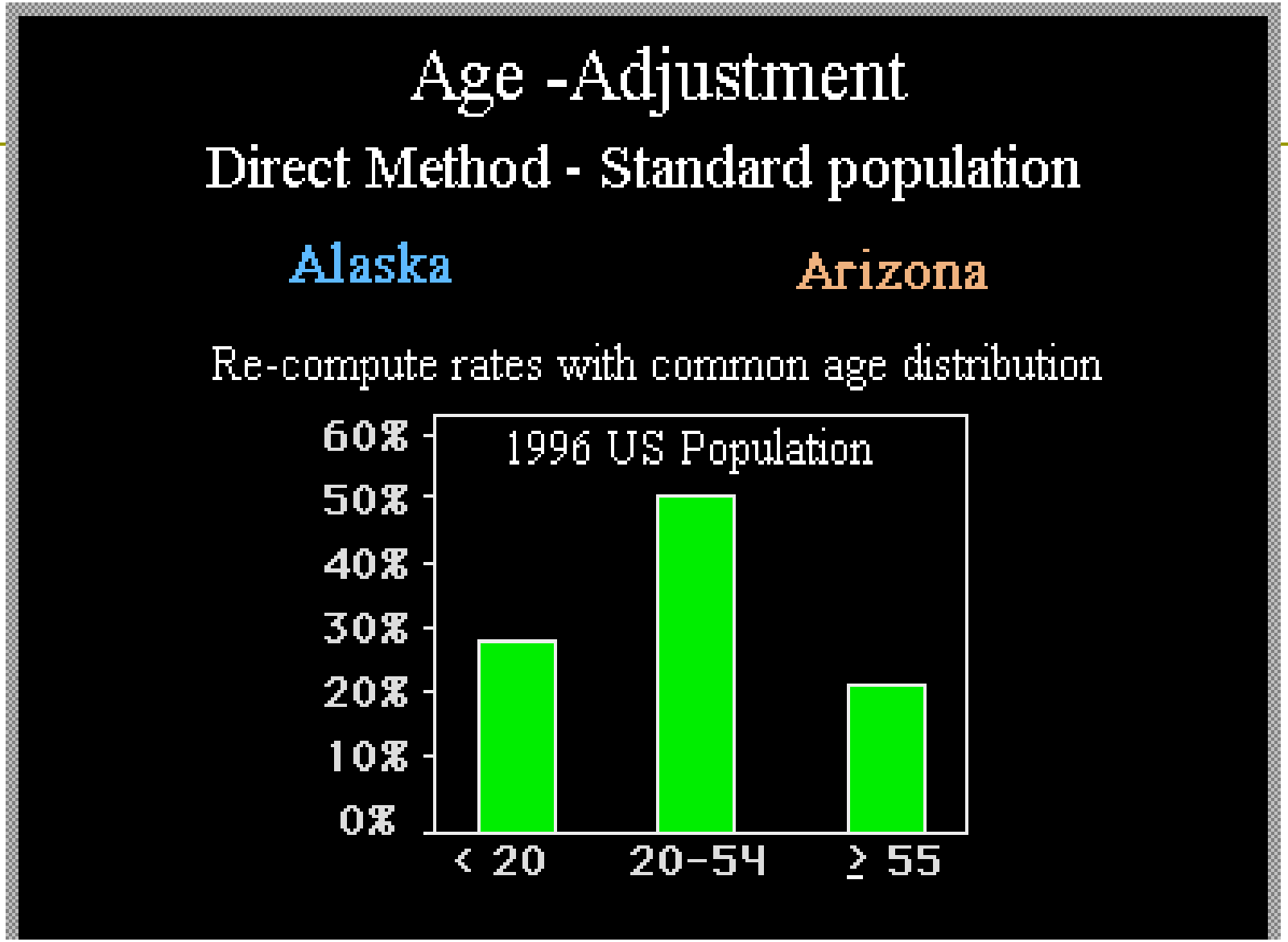
AGE **distorts** the comparison.

Age -Adjustment

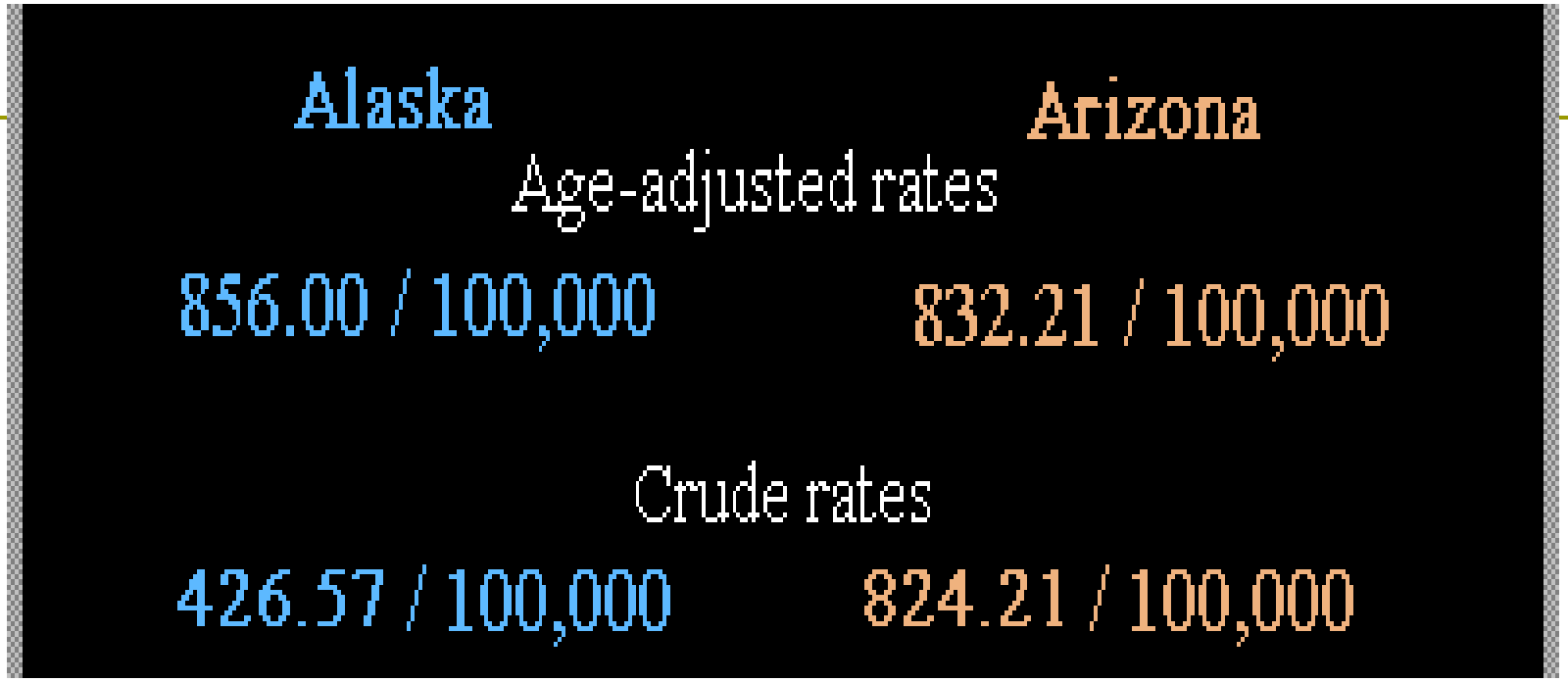
↓

Age-Adjusted Rates

Example from Kleinbaum:



Example from Kleinbaum:



Controlling for any age differences in the two populations, the overall mortality rate is higher in Alaska, the cold damp climate then in Arizona, the warm dry climate.

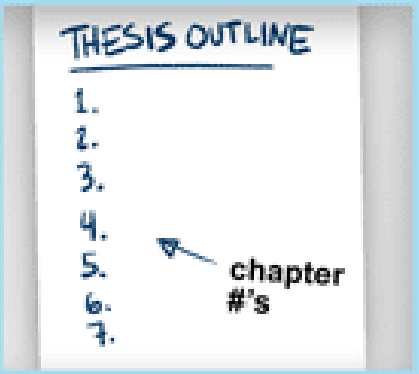
Readings for this week

- Rothman text:
  - Chapter 3: Measuring disease occurrence and causal effects
- Gordis text:
  - Chapter 3: Measuring the occurrence of disease: morbidity
  - Chapter 4: Measuring the occurrence of disease: mortality

WRITING YOUR THESIS OUTLINE

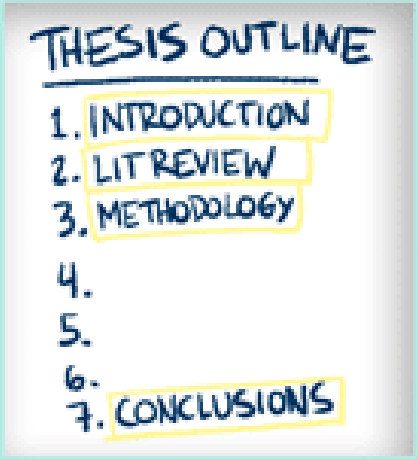
NOTHING SAYS "I'M ALMOST DONE" TO YOUR ADVISOR/ SPOUSE/PARENTS LIKE PRETENDING YOU HAVE A PLAN

STEP 1 Aim for a respectable number of chapters:



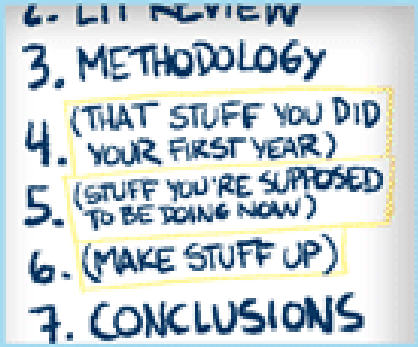
5 = "That's IT??"
6-7 = "Not bad"
8+ = "Are you crazy??"

STEP 2 Fill in the "freebies":



You're half way done!

STEP 3 Make up titles for the "meat" chapters:



(It'll be years before you actually have to work on that later chapter, and by then your thesis topic will have changed anyway)

STEP 4 Voilá! You just bought yourself another two years

www.phdcomics.com