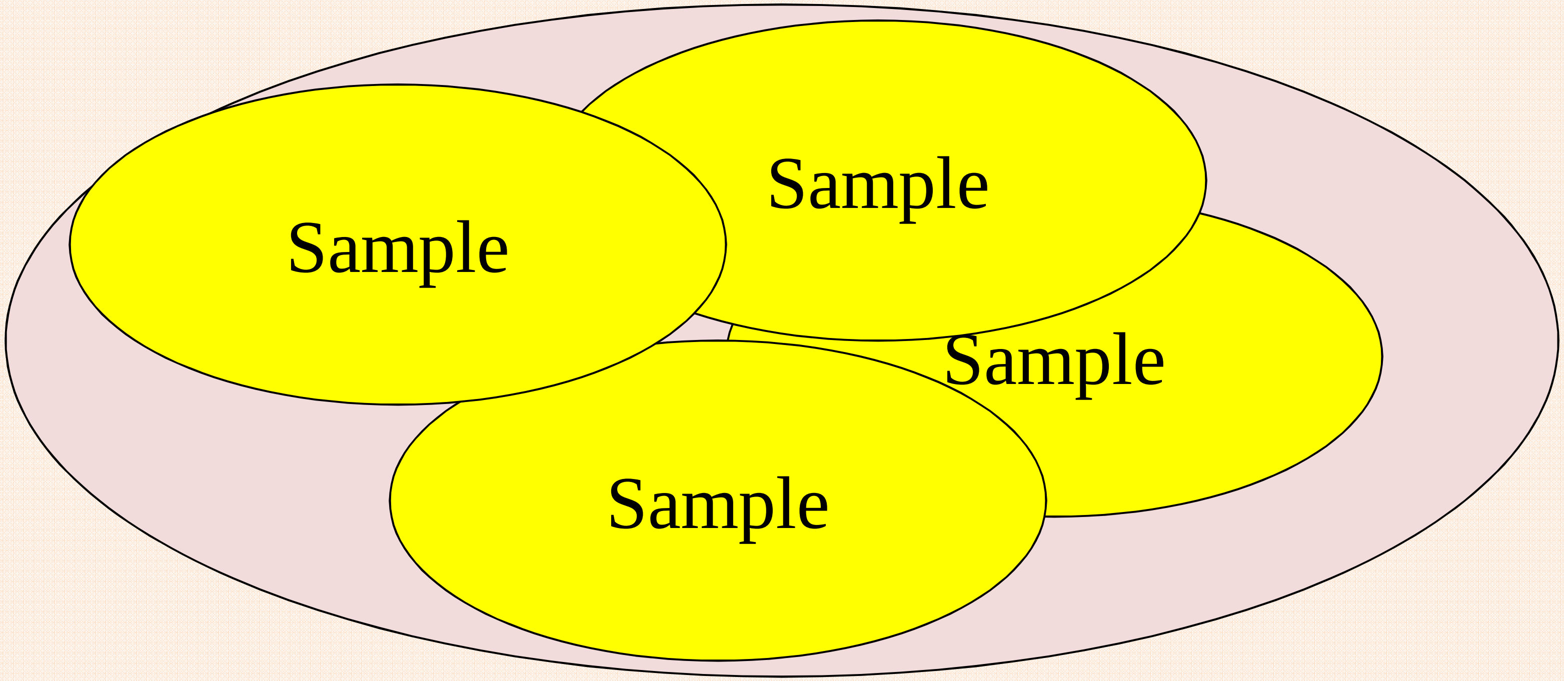
Sampling, Sampling distribution & Estimates

MANIPAL UNIVERSITY

Sampling

Population



Probability/ Random sampling

Method of selecting a sample from a population by giving all individuals some probability of selection

Methods of random sampling

- Simple random sampling
- Stratified random sampling
- Systematic sampling
- Cluster sampling
- Multistage sampling

Simple random sampling

Method of selecting a sample from a population by giving all possible samples equal probability of selection

Lottery method/ Chit method/ Random number table

Advantages

All members get equal chance.

Estimates are easy to calculate

Disadvantages

Difficulty in constructing sampling frame

Minority subgroups might not get represented

Stratified random sampling

The population is first divided into strata or subpopulations according to gender, disease severity, geographical location

A sub sample is drawn from each strata.

Why stratification?

Administrative convenience

To get representation for all subgroups

To get more precise estimate

Cluster sampling

The population is first divided into a number of clusters like slums, localities, family, physician, clinic, school etc.

A simple random sample of clusters is selected and all or a sample of elementary units/ belonging to the selected clusters are studied.

Administrative convenience

Cost effective

Less precise estimate

Why we prefer probability sampling?

Generalizability of findings

Derivation of estimates and its precision

Derivation of sampling distribution

Applicability of theory of hypothesis testing

Estimation of bias in the estimate

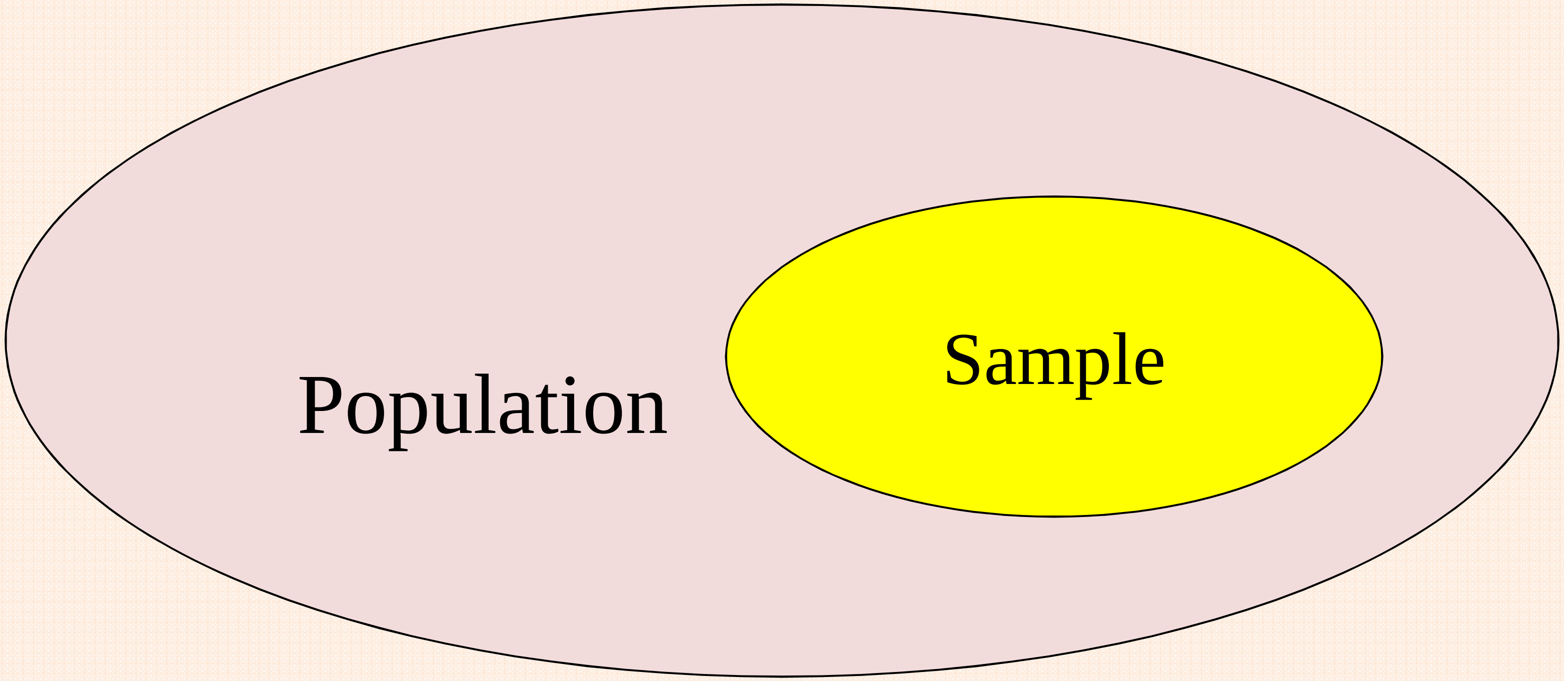
Non probability sampling methods

Snowball sampling

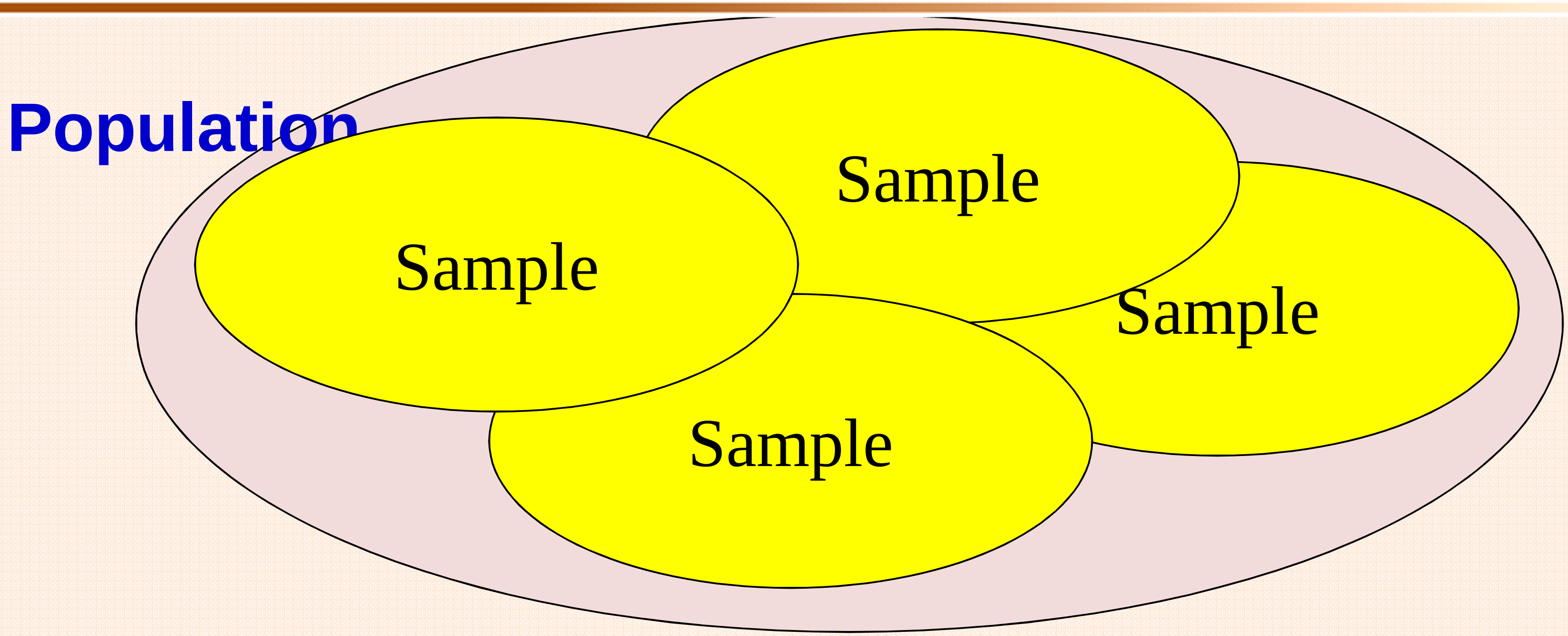
Purposive sampling

Quota sampling

The Sampling Distribution



Sampling distribution



If repeated samples of same size are taken from a population and the estimate is calculated from each sample, the estimate will vary, that is, they can have a distribution over a central value which is **the sampling distribution of that estimate**

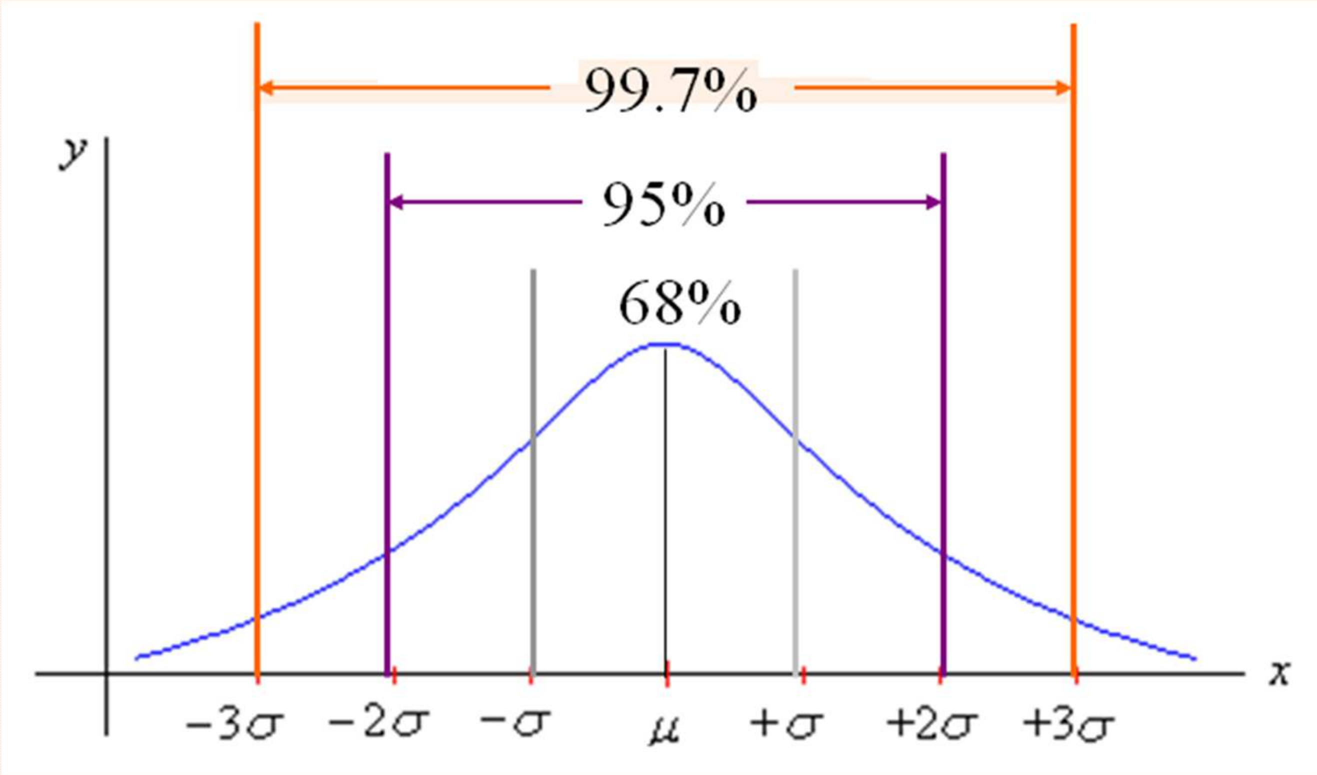
Sampling Error

Sampling error is the difference between estimate and actual parameter

Standard deviation of the sampling distribution, known as the standard error, is taken as a good estimate of sampling error

Normal Curve

Approximate percentage of area within given standard deviations



Confidence interval of population mean

An interval estimate of the unknown population mean, based on a random sample from the population.

The interval (Mean- 2 x s /√n, Mean +2 x s /√n) is an approximate 95% confidence interval for population mean

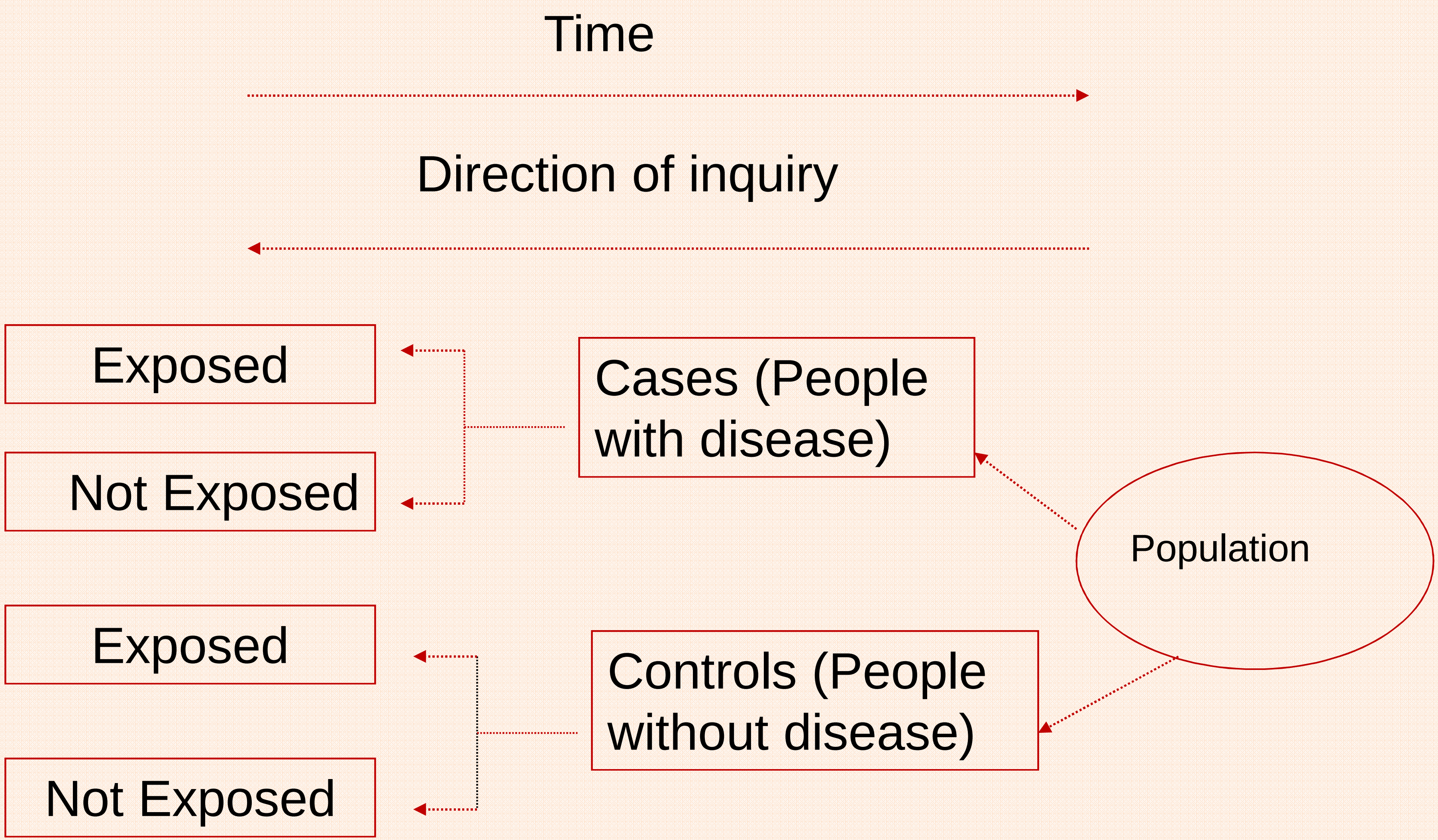
95% CI of population mean = Mean ± 2 SEM

Confidence interval for population proportion

- An approximate 95% confidence interval for the parameter *p* *(Prevalence or incidence)* is

$$\hat{p} \pm 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Estimation of Odds Ratio: Case-Control study



•odds of exposure among cases

= 29/205

•odds of exposure among controls = 135/1307

OC use	Cases	Controls	Total
Yes	29	135	164
No	205	1307	1512
Total	234	1442	1676

•odds ratio = $\frac{29/205}{135/1307}$ = 1.37

Where odds ratio = 1, this implies no difference in effect

Interpretation

How many times more (or less) the exposed group likely to get the event compared to unexposed group

- ⇨ Exposure may be a protective factor
- ⇨ Exposure may be a risk factors

Simple interpretation

OR = 1	No association
OR < 1	A negative association
OR > 1	A positive association

Exposure	Cases	Controls	Total
Yes	a	b	a+b
No	c	d	c+d
Total	a+c	b+d	a+b+c+d = N

OR = ad/bc

Computation of 95% confidence interval for OR

95% C.I. of OR = $\left(OR\, e^{-1.96(S.E)},\ OR\, e^{+1.96(S.E)}\right)$

$$S.E = \sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}$$

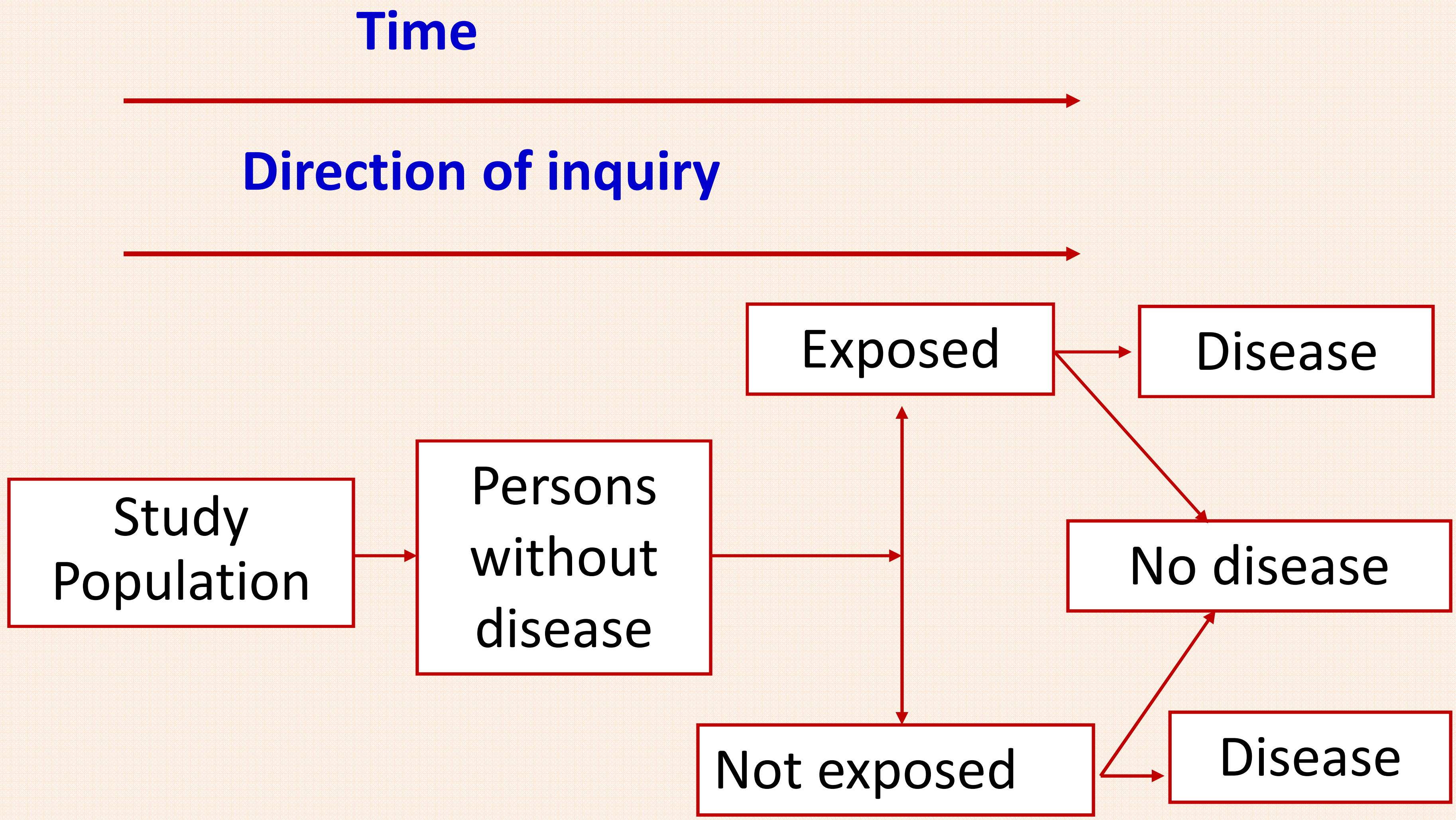
Further interpretation

95% C.I of OR contains 1

95% C.I. of OR below 1

95% C.I. of OR is above 1

Estimation of Relative Risk: Cohort study



Smoking & Coronary heart disease- Cohort study

Smoking	Develop CHD	No CHD	Total
Yes	84	2916	3000
No	87	4913	5000
Total	171	7829	8000

Risk Ratio or Relative Risk

- Risk or incidence of event in smokers

  = **84/3000**

- Risk or incidence of Event in non-smokers

  = **87/5000**

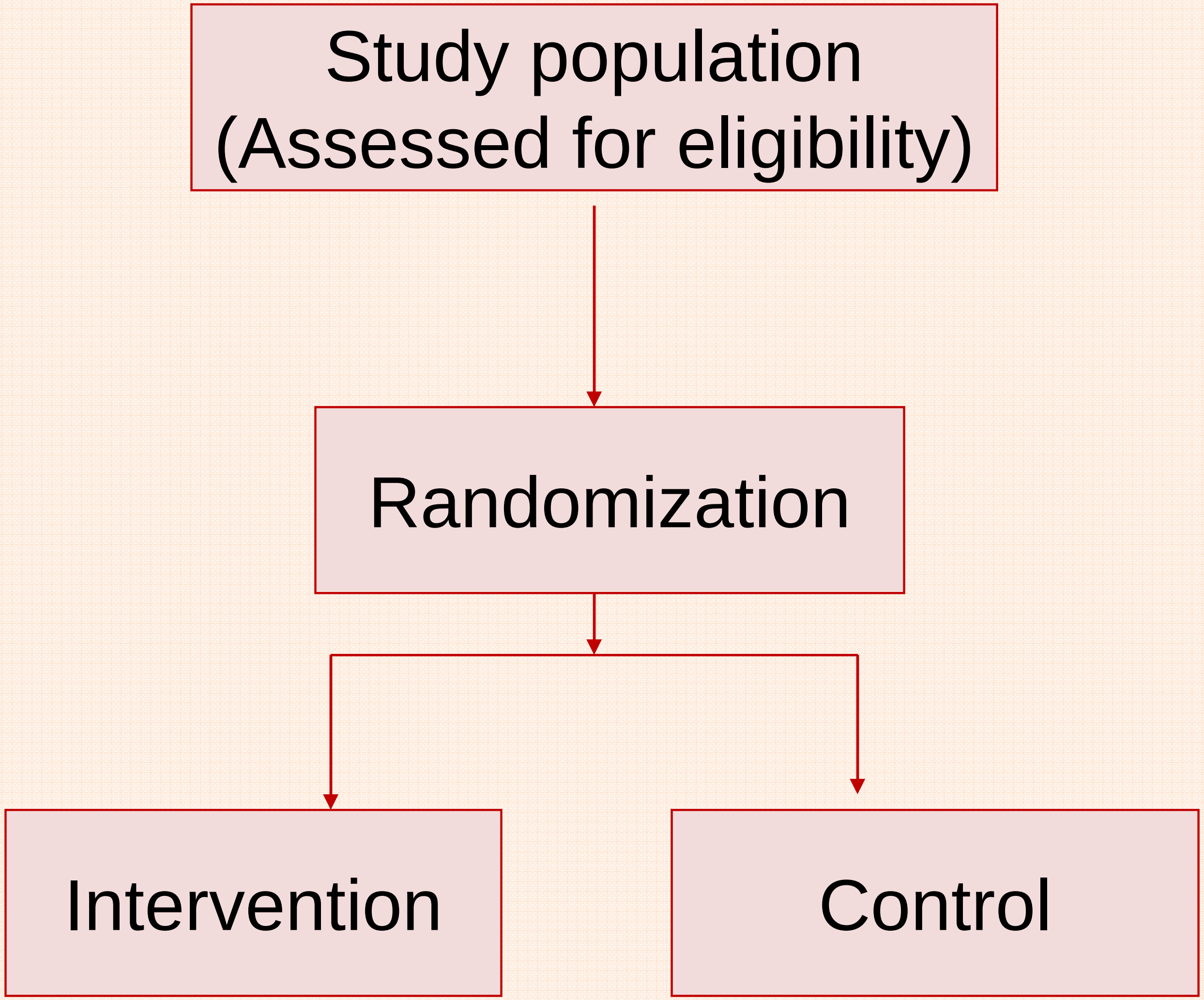
Smoking	Develop CHD	No CHD	Total
Yes	84	2916	3000
No	87	4913	5000
Total	171	7829	8000

- Risk ratio = $\frac{84/3000}{87/5000} = \frac{0.028}{0.017} = 1.64$

  = $\frac{\text{Risk in smokers}}{\text{Risk in non-smokers}}$

Where RR = 1, this implies no difference in effect

Randomized Controlled Trial



Comparing Azithromycin with Amoxycillin

Treatment	Clinical failure	No event	Total
Azithro	4	44	48
Amoxy	7	49	56
Total	11	93	104

Risk Ratio or Relative Risk

- Risk of event on Azithro

  = **4/48**

- Risk of event on Amox

  = **7/56**

Treat	Clinical failure	No event	Total
Azithro	4	44	48
Amoxy	7	49	56
Total	11	93	104

- Risk ratio = $\frac{4/48}{7/56} = \frac{0.083}{0.125} = 0.66$

  = $\frac{\text{Risk on Azithro}}{\text{Risk on Amox}}$

Where RR = 1, this implies no difference in effect

Computation of 95% CI for RR

Confidence limits = $\left(RR\, e^{-1.96(SE)}, RR\, e^{1.96(SE)}\right)$

where $SE = \sqrt{\frac{1}{a} - \frac{1}{a+b} + \frac{1}{c} - \frac{1}{c+d}}$

Interpretation

Relative risk represent how many times more (or less) likely the disease occurs in the exposed group as compared with the unexposed.

RR = 1	No association
RR < 1	Negative association
RR > 1	Positive association

95% CI of RR contains 1

95% CI of RR is below 1

95% CI of RR is above 1