Analyzing epidemiological studies
Epidemiology III.

Epidemiological studies

- Most published studies are analytic or experimental

- These have the aim of discerning a cause-effect relationship between certain factors

- Different types of studies are able to provide different levels of evidence for a causal relationship

- In reality one-to-one cause-effect relationships are rare, we usually encounter „causal webs"

- Ascertainment of cause-effect relationships is one of the central and most difficult tasks of all scientific activities

Hill's causal criteria

➤ *strength of association* (the stronger the more probable)

➤ *consistency* (over space, time, method, research group…)

➤ *dose - response relationship* (larger dose - larger effect)

➤ *chronological relationship* (cause before effect)

➤ *specificity* (one-to-one relationship)

➤ *biological plausibility* (is the relationship plausible at all?)

➤ *coherence* (does it fit with specific established „natural laws")

➤ *analogy* (with similar systems of causation)
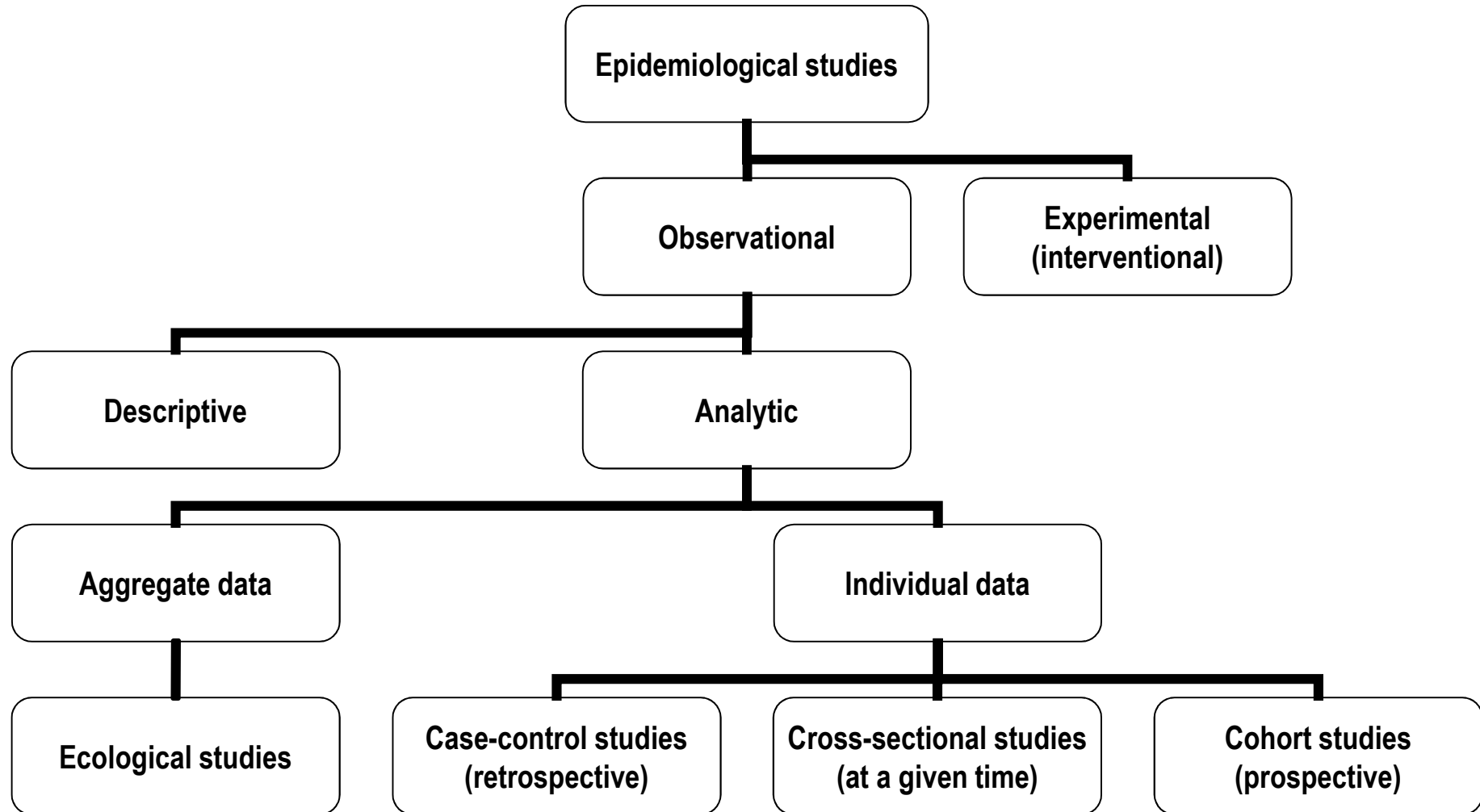
➤ *experimental evidence*

Basic biostatistical/epidemiological concepts

- null hypothesis

- alternative hypothesis

- representativity

- randomization

- statistical significance

- confidence interval

- random error

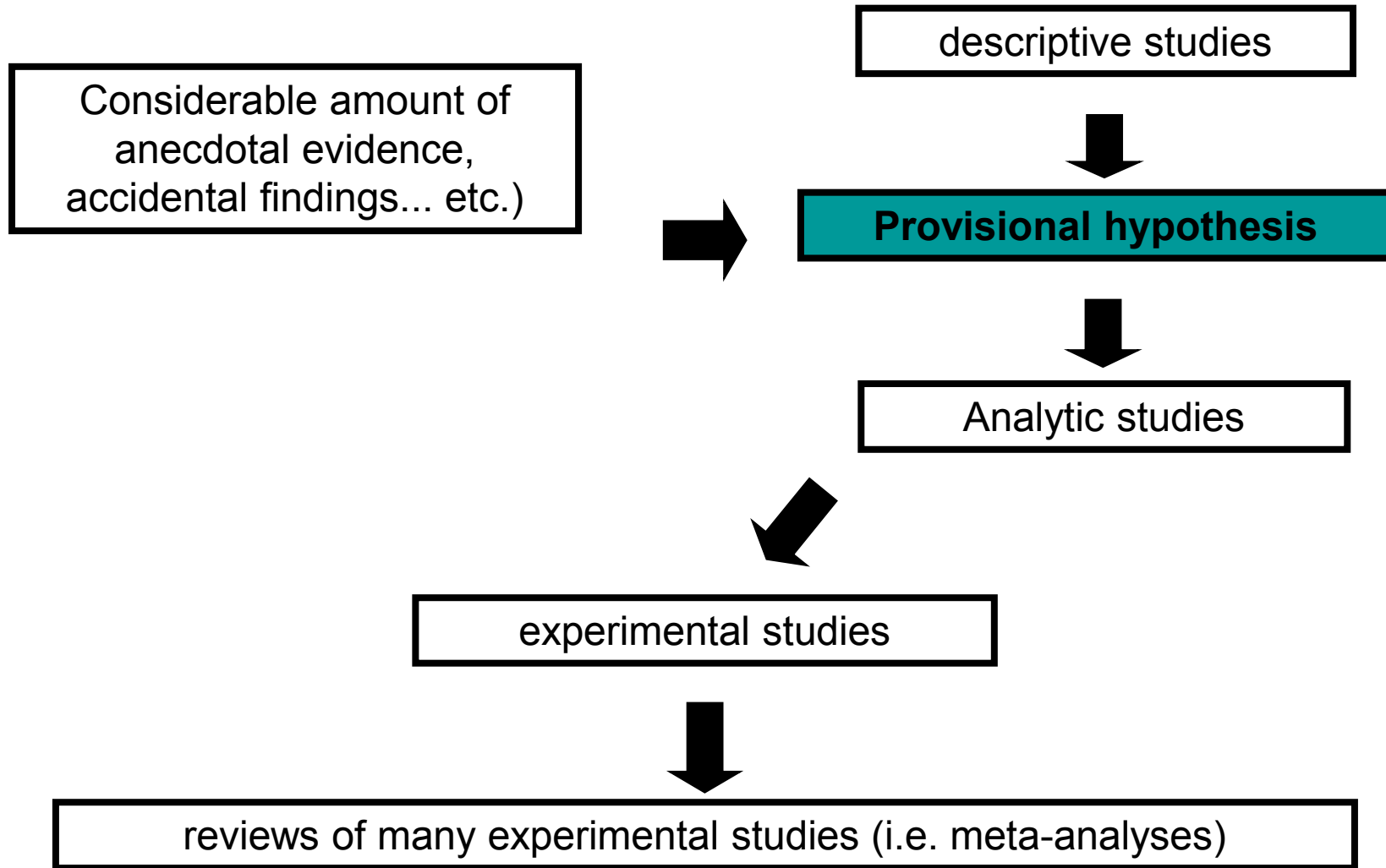- systematic error

- confounding factor

Epidemiological studies

- Main functions are to:

  ✓ collect,

  ✓ analyze and

  ✓ utilize health-related information in order to improve population health.

- Planning epidemiological studies involves:

  ✓ professional (medical, epidemiological, ethical)

  ✓ administrative and

  ✓ economic considerations.

Types of epidemiological studies

```
                    ┌──────────────────────────┐
                    │  Epidemiological studies │
                    └──────────────────────────┘
                          │
            ┌─────────────┴──────────────┐
   ┌─────────────────┐          ┌──────────────────┐
   │  Observational  │          │   Experimental   │
   │                 │          │ (interventional) │
   └─────────────────┘          └──────────────────┘
        │
  ┌─────┴──────────────┐
┌──────────────┐   ┌──────────────┐
│  Descriptive │   │   Analytic   │
└──────────────┘   └──────────────┘
                        │
              ┌─────────┴──────────┐
     ┌────────────────┐      ┌──────────────────┐
     │ Aggregate data │      │  Individual data │
     └────────────────┘      └──────────────────┘
            │                         │
            │              ┌──────────┼──────────────┐
 ┌────────────────────┐ ┌────────────────────┐ ┌────────────────────┐ ┌────────────────────┐
 │ Ecological studies │ │ Case-control       │ │ Cross-sectional    │ │ Cohort studies     │
 │                    │ │ studies            │ │ studies            │ │ (prospective)      │
 │                    │ │ (retrospective)    │ │ (at a given time)  │ │                    │
 └────────────────────┘ └────────────────────┘ └────────────────────┘ └────────────────────┘
```

The typical course of epidemiological investigation

descriptive studies

Considerable amount of anecdotal evidence, accidental findings... etc.)

Provisional hypothesis

Analytic studies

experimental studies

| reviews of many experimental studies (i.e. meta-analyses) |
|---|

Strength of evidence of studies

Systematic review or meta-analysis of RCTs

Double-blind RCTs

Single-blind RCTs

Randomized, controlled trials (RCTs)

Non-randomized / uncontrolled experimental studies

„Regular" cohort studies

Historical cohort studies

Case-control studies

Ecological studies

Cross-sectional studies

Expert opinions, anecdotal reports

Ecological studies

- These studies analyze exposition based on aggregated with the method of correlation analysis (e.g. number of cigarettes sold and mortality rate)

- Conceptually, the ecological component in this kind of study is an issue of data analysis and not study design.

- Relationship between exposure and outcome at the individual level can not be described (ecological fallacy)

CHD mortality and cigarette-consumption in American states
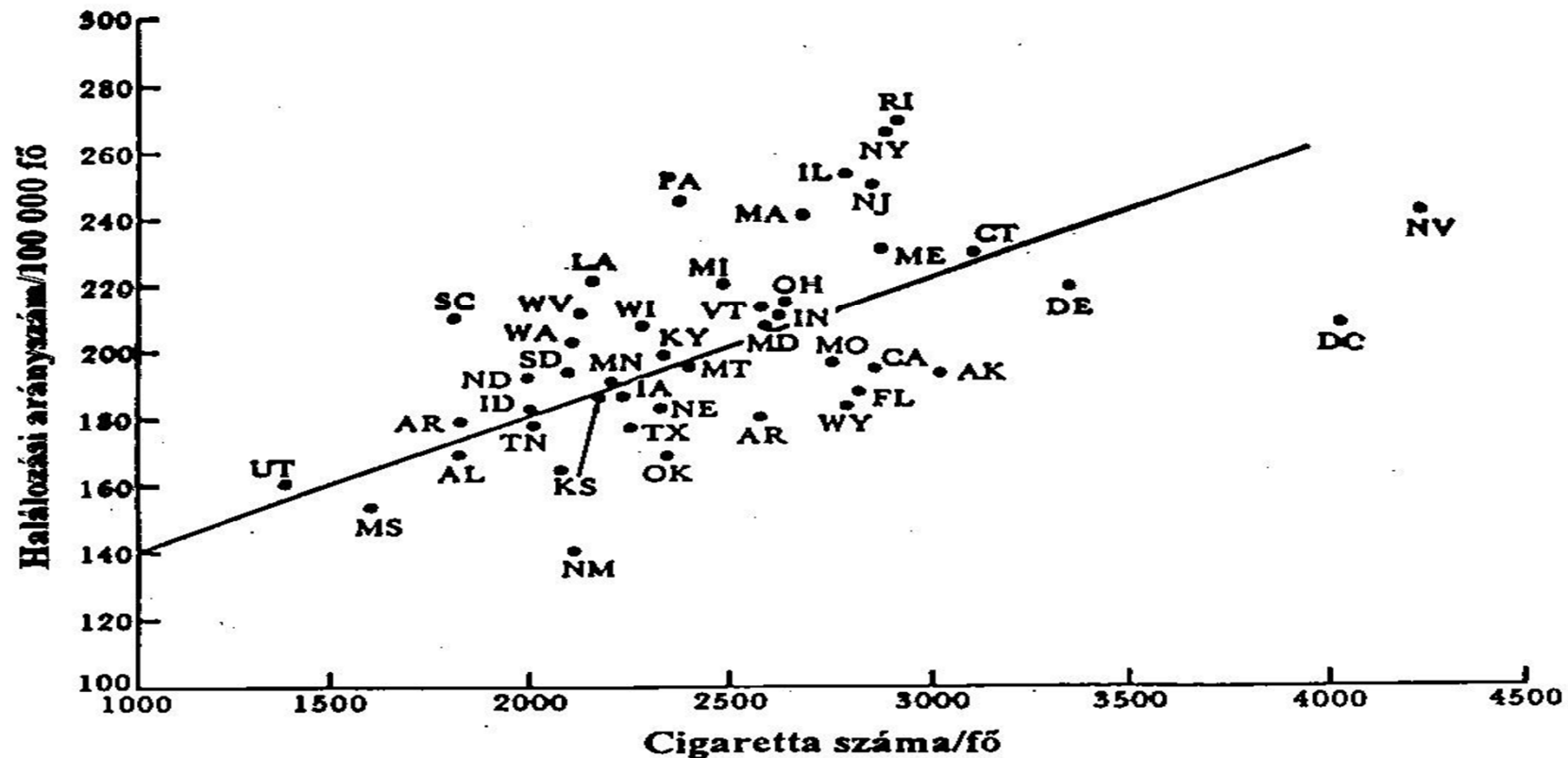
Forrás: Friedman G. D.: Cigarette smoking and geographic variation in coronary heart disease mortality in the United States. *J. Chronic Dis.* 20:769, 1967.

Ecological fallacy: example

- Imagine a study of the rate of coronary heart disease in the capital cities of the world relating the rate to average income.

- Within the cities studied, coronary heart disease is higher in the richer cities than in the poorer ones.

- We might predict from such a finding that being rich increases your risk of heart disease.

- In the industrialised world the opposite is the case - within cities such as London, Washington and Stockholm, poor people have higher CHD rates than rich ones.

- The ecological fallacy is usually interpreted as a major weakness of ecological analyses.

- Ecological analyses, however, informs us about forces which act on whole populations.

Descriptive studies: the National Health Interview Survey 2003 (OLEF 2003)

Distribution of total amount of gratuities paid in 2003 between various health care sectors

GP services (%) — 25%

hospital inpatient care (%) — 42%

specialist ambulant care (%) — 33%
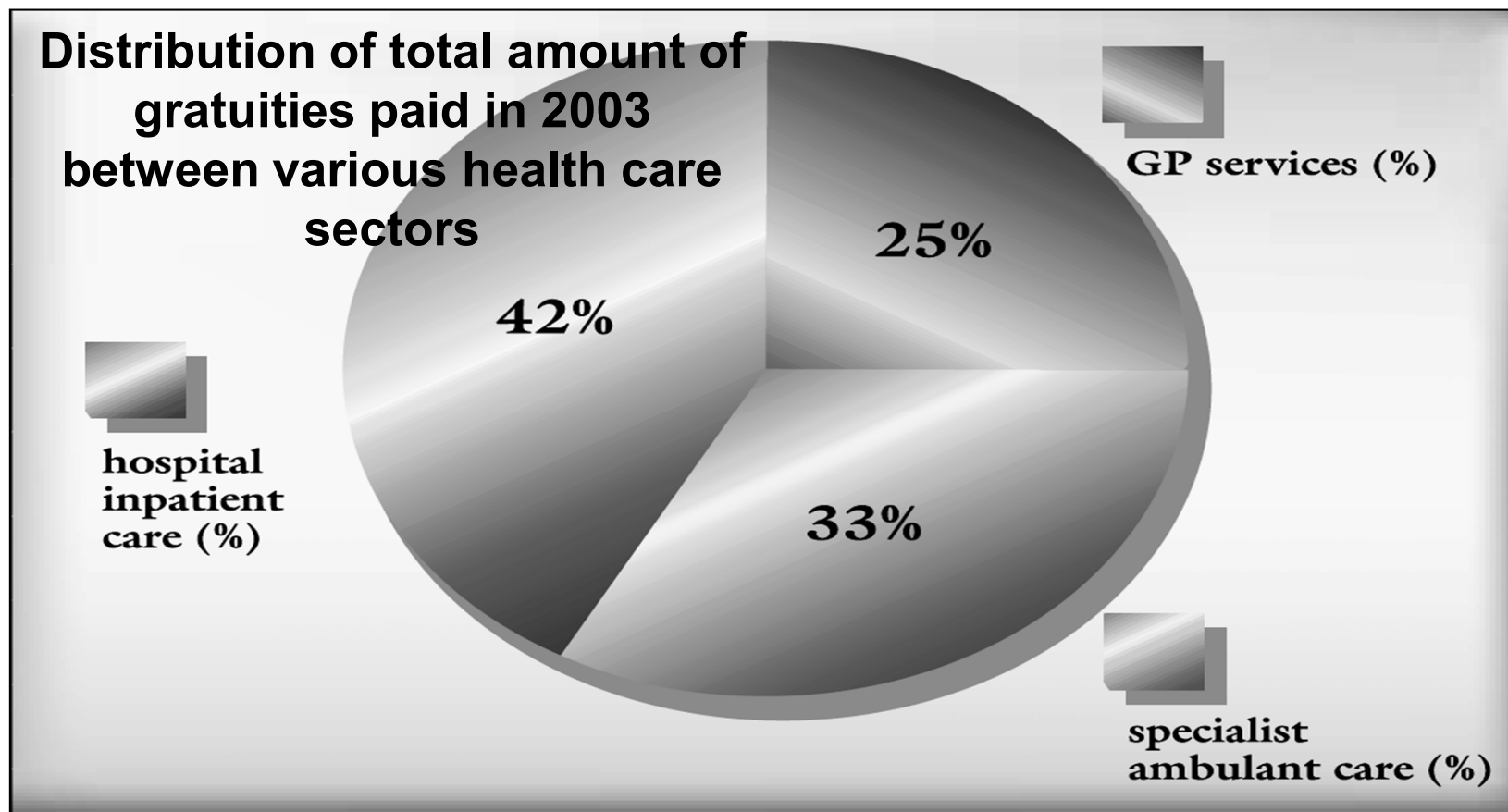
Source: National Health Interview Survey (OLEF) 2003. Executive Update. National Center for Epidemiology, 2003. http://www.oek.hu.

Descriptive studies

- Populational data
- Individual data (case, case-series)
  - Personal factors ⟶ **Who?**
    (age, gender, marital status)

  - Place ⟶ **Where?**
    (geographical and social environment)

  - Time ⟶ **When?**
    (changes in a long or short period of time, seasonal changes)

Cross-sectional studies

These studies observe the
- exposition factor and the
- disease

at a same, given time on an individual level

These studies inform us about the frquency of the disease and the exposition factor at a given time so it estimates prevalence.
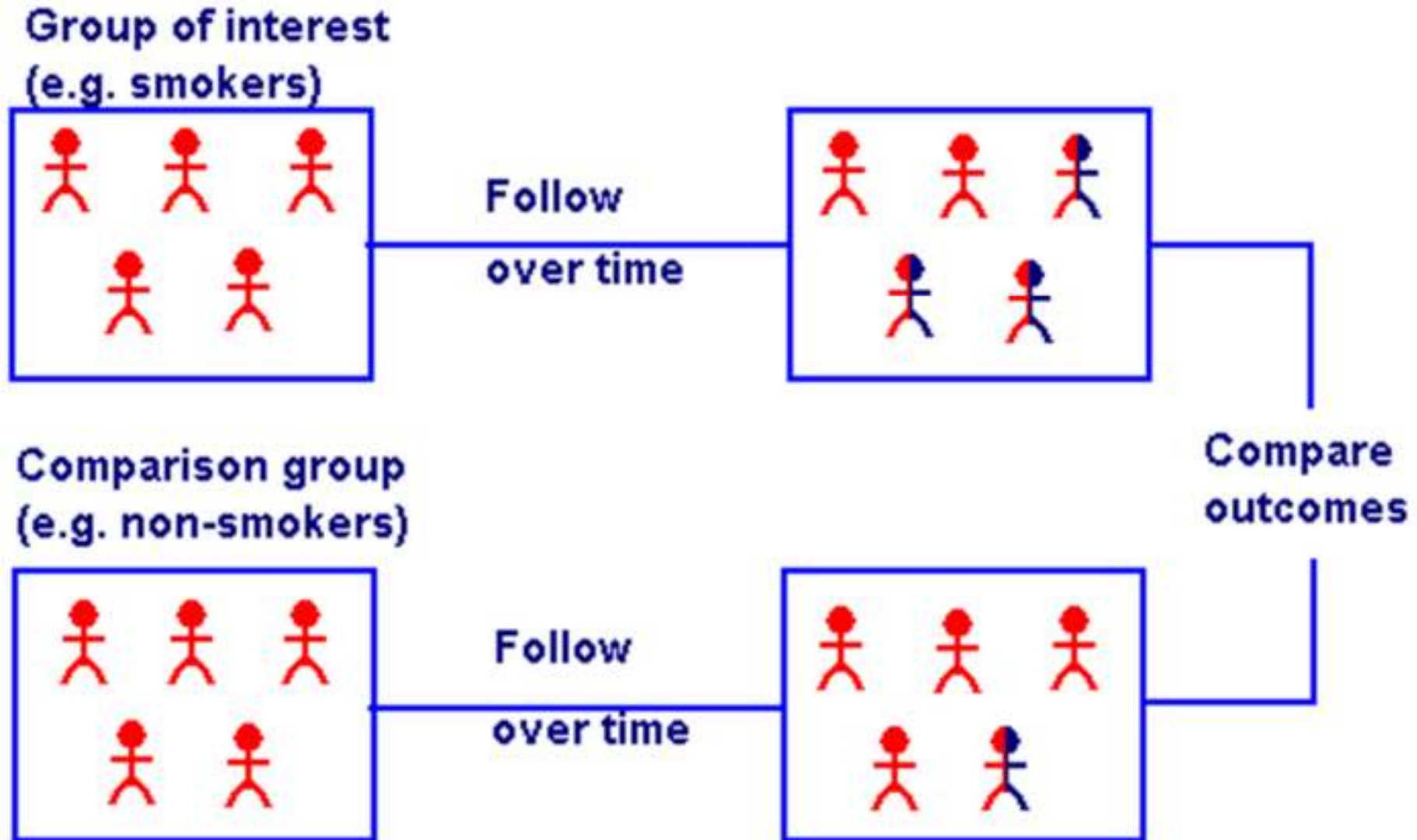
Cross-sectional study

Disease

	Yes	No
Exposed	a	b
Non-exposed	c	d

$$P = \frac{a}{a + b} \quad \text{vs.} \quad P = \frac{c}{c + d}$$

$$P = \frac{a}{a + c} \quad \text{vs.} \quad P = \frac{b}{b + d}$$

Cohort studies

Group of interest (e.g. smokers)

Follow over time

Comparison group (e.g. non-smokers)

Follow over time

Compare outcomes

Cohort study

	Disease	
	Yes	No
Exposed	a	b
Non-exposed	c	d

Study design →

Cannabis consumption and psychosis in a Dutch cohort study I.

- Van Os et. al. conducted the large-scale study between 1997 and 1999

- They selected 4045 non-psychotic individuals from the general population, and determined whether each individual used cannabis or not

- They then examined the incidence of psychosis in these subjects at 1 and 3 years

Source: Van Os et. al. Am J Epidemiol Vol. 156, No. 4, 2002.

Cannabis consumption and psychosis in a Dutch cohort study II.

Cannabis use	Psychosis	Healthy	Total
Yes	8 a	304 b	312 a+b
No	c 30	d 3622	c+d 3652
Total	38 a+c	3926 b+d	3964

Source: Van Os et. al. Am J Epidemiol Vol. 156, No. 4, 2002.

Measures of Risk I.

<u>Absolute risk</u>: incidence in a given population

Risk of psychosis in exposed & unexposed?

<u>Relative risk</u> (RR):

$$\frac{\text{absolute risk in exposed } (R_{EXP})}{\text{absolute risk in unexposed } (R_{UNEXP})}$$

Relative risk of psychosis in exposed compared to the risk in the unexposed?

<u>Attributable risk</u> (AR): $R_{EXP} - R_{UNEXP}$

Risk of psychosis attributable to exposure?

Attributable fractions

- **Attributable fraction in exposed**

  describes the percentage of the incidence of the exposed group that occurs because of exposition

- **Attributable fraction in the whole population**

  the proportion of the total risk of a disease in a population that can be attributed to exposure

Defining risk (formulas)

$$I_{exp} = \frac{a}{a+b} \qquad I_{non\text{-}exp} = \frac{c}{c+d}$$

$$\text{Relative risk (RR)} = \frac{I_{exp}}{I_{non\text{-}exp}}$$

$$\text{Attributable risk (AR)} = I_{exp} - I_{non\text{-}exp}$$
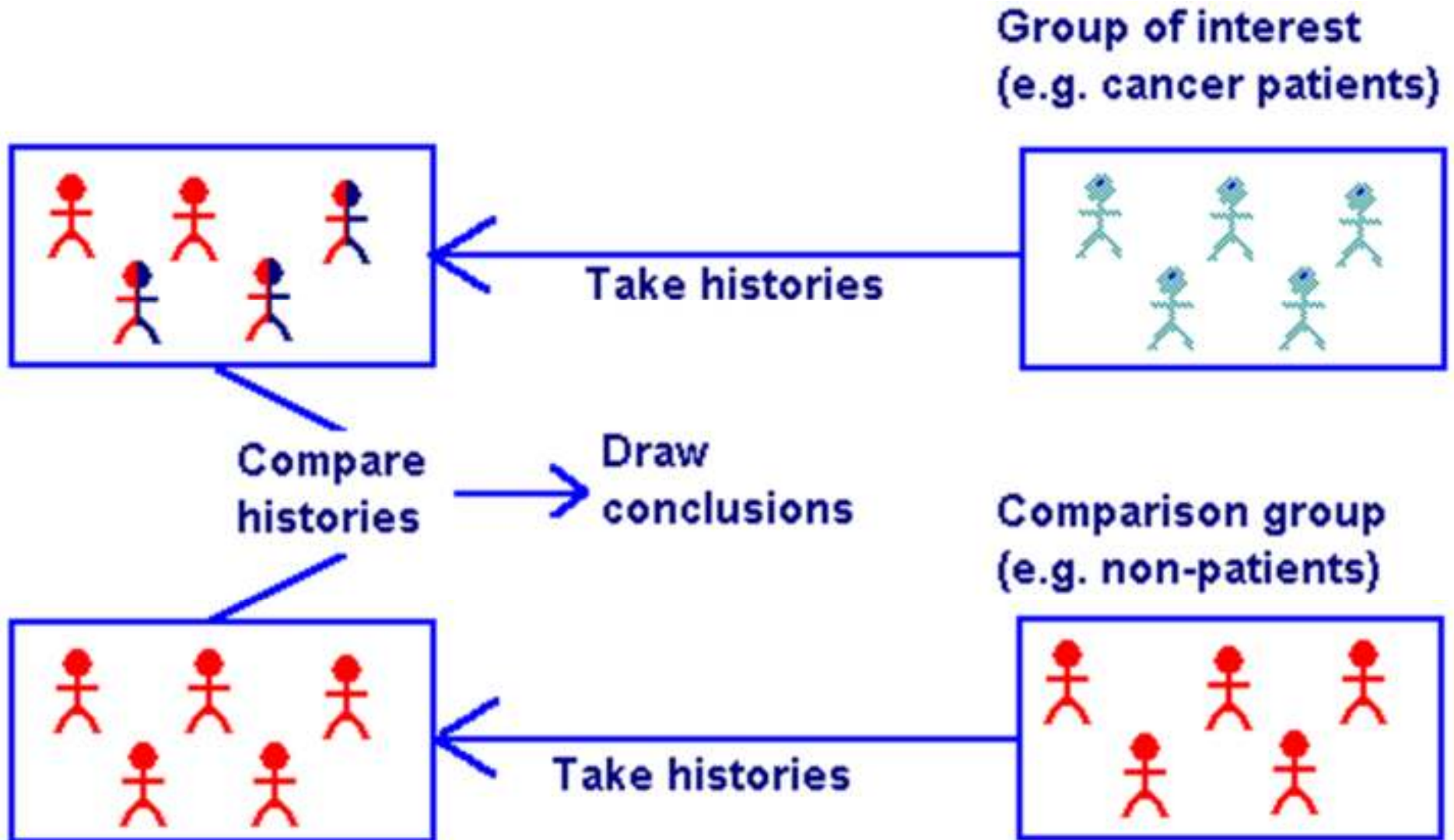
$$\text{Attributable fraction in exposed(\%)} = \frac{I_{exp} - I_{non\text{-}exp}}{I_{exp}} \times 100$$

$$\text{Attributable fraction in population (\%)} = \frac{I_{pop} - I_{non\text{-}exp}}{I_{pop}} \times 100$$

Some characteristics of cohort studies

- Exposure is measured prior to the onset of disease.

- The connection between an exposure and multiple outcome measures can be assessed simultaneously.

- Study design is prospective, but may be historic.

- Incidence can be measured directly.

- Relative and attributable risk can be calculated from incidence figures.

- Usually quite expensive and time-consuming.

- Studies typically require large efforts in organization and management, compliance of subjects is variable, many subjects may discontinue their participation.

Case-control studies

Group of interest (e.g. cancer patients)

Take histories

Compare histories

Draw conclusions

Comparison group (e.g. non-patients)

Take histories

Case-control studies

Disease

	Case group (sick people)	Control group
Exposition in anamnesis	a	b
No exposition in anamnesis	c	d

Study design

Retrolental Fibroplasia

- First cases described in 1942, in the USA (Terry).

- Almost exclusively in premature infants.

- Characterized by non-inflammatory fibrosis and vascular proliferation in the ocular bulbs with various outcomes ranging from complete recovery to permanent blindness.

- Usually develops within the first month of life, and never after the third month.

- More common in the Caucasian population than in Afro-Americans.

- No seasonal, geographical or genetic incidence pattern.

- First studies focused on the connection between RFP and birth weight

Case-control studies

Retrolental fibroplasia

Incidence of RLF by birth weight (Fletcher, 1955)[*]

Birth weight	Premature infants		
	RLF	Controls	Total
< 1701 g	128	261	389
1700 - 2500 g	8	75	83
Total	136	336	472

Source: Lowe CR. Epidemiology: A guide to teaching methods. WHO, 1973.

Measures of Risk II.

Odds:
$$\frac{\text{prevalence of exposure among cases (or!) controls}}{1-(\text{prevalence of exposure among cases (or!) controls})}$$

What are the odds of having been born with very low   birth weight (<1701 g) among RFP cases and controls?

Odds ratio (OR):
$$\frac{\text{odds of exposure among cases}}{\text{odds of exposure among controls}}$$

How many times higher are the odds of having been born   with very low birth weight among RFP cases than among   controls?

Some characteristics of case-control studies

- Incidence cannot be directly measured.

- Odds of past exposure can be calculated in both case and control groups.

- Study design is retrospective.

- Multiple exposures can be assessed in connection with a single outcome measure.

- The relative risk of becoming a case on exposure as compared to non-exposure can be approximated with the odds ratio of the case and control groups.

- Relatively low-budget and requires relatively little time.

- A longer stretch of time can be assessed with regards to exposure.

- Assessment of exposure may be biased by inaccurate recall of subjects.

The effects of O$_2$ therapy on the incidence of RLF in premature infants (based on a cohort! study)

The effects of O$_2$ therapy in premature infants (Patz,1952)[*]

| O$_2$ therapy	Premature infants		
	RLF	No RLF	Total
65-70% O$_2$ for 4-7 weeks	17	11	28
< 40% O$_2$ for < 2 weeks	6	31	37

Errors in case-control studies

- Selection bias

- Recall distortion

- Observational bias

- Relative risk and odds ration may be different in case of frequent diseases

Recommended use of epidemiological studies

	Case-control	Cohort	Cross-sectional
	study		
Rare disease	+++++	-	-
Long latency period	+++	-	-
Observation of more expositional factors	++++	+++	++
Rare exposition factor	-	+++++	-
Observing more outcomes of one exposition factor	-	+++++	++
Observing the connection in time	-	+++++	-
Calculating incidence	-	+++++	-

Randomised clinical trials
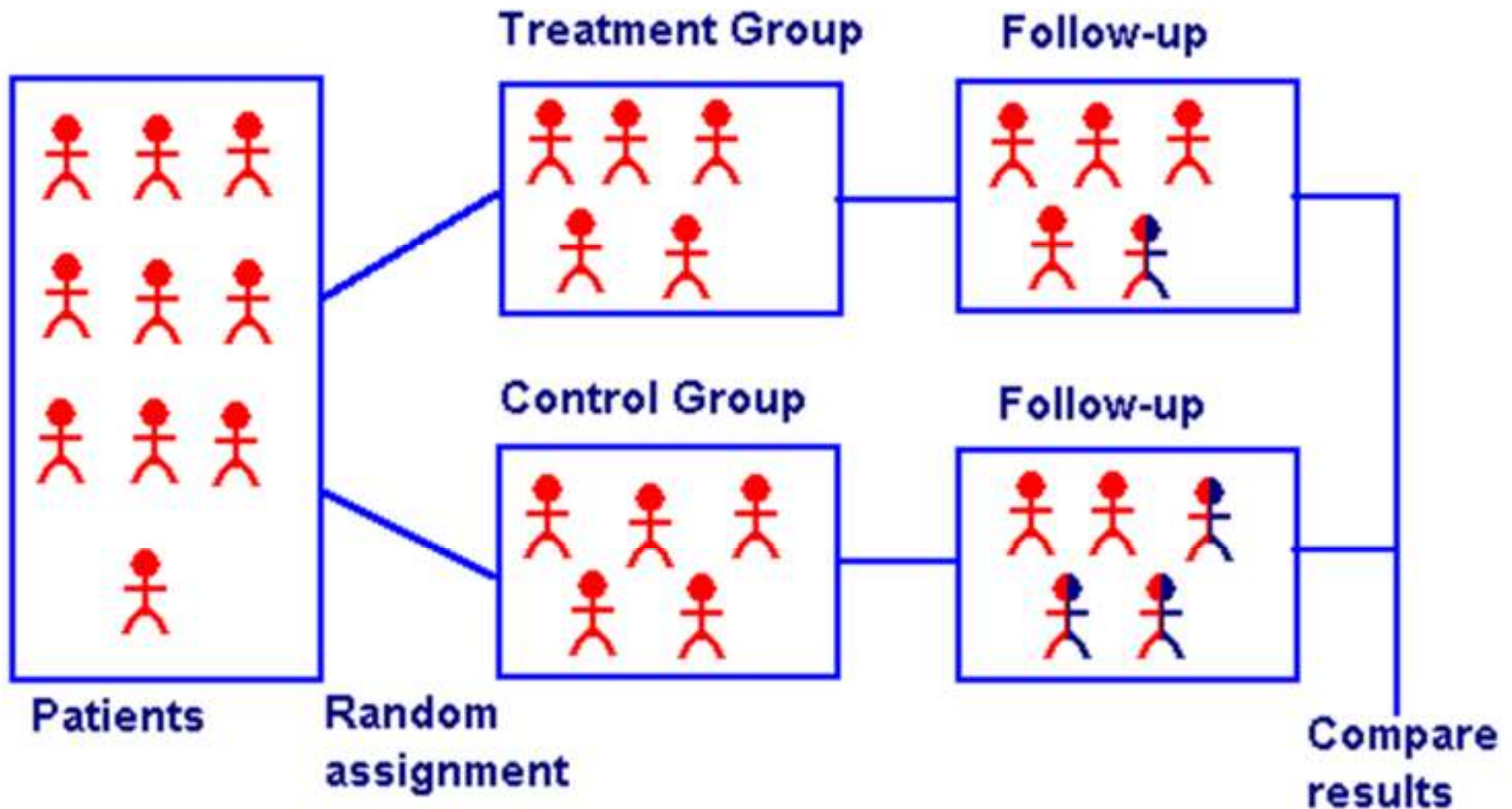
Double blind trials
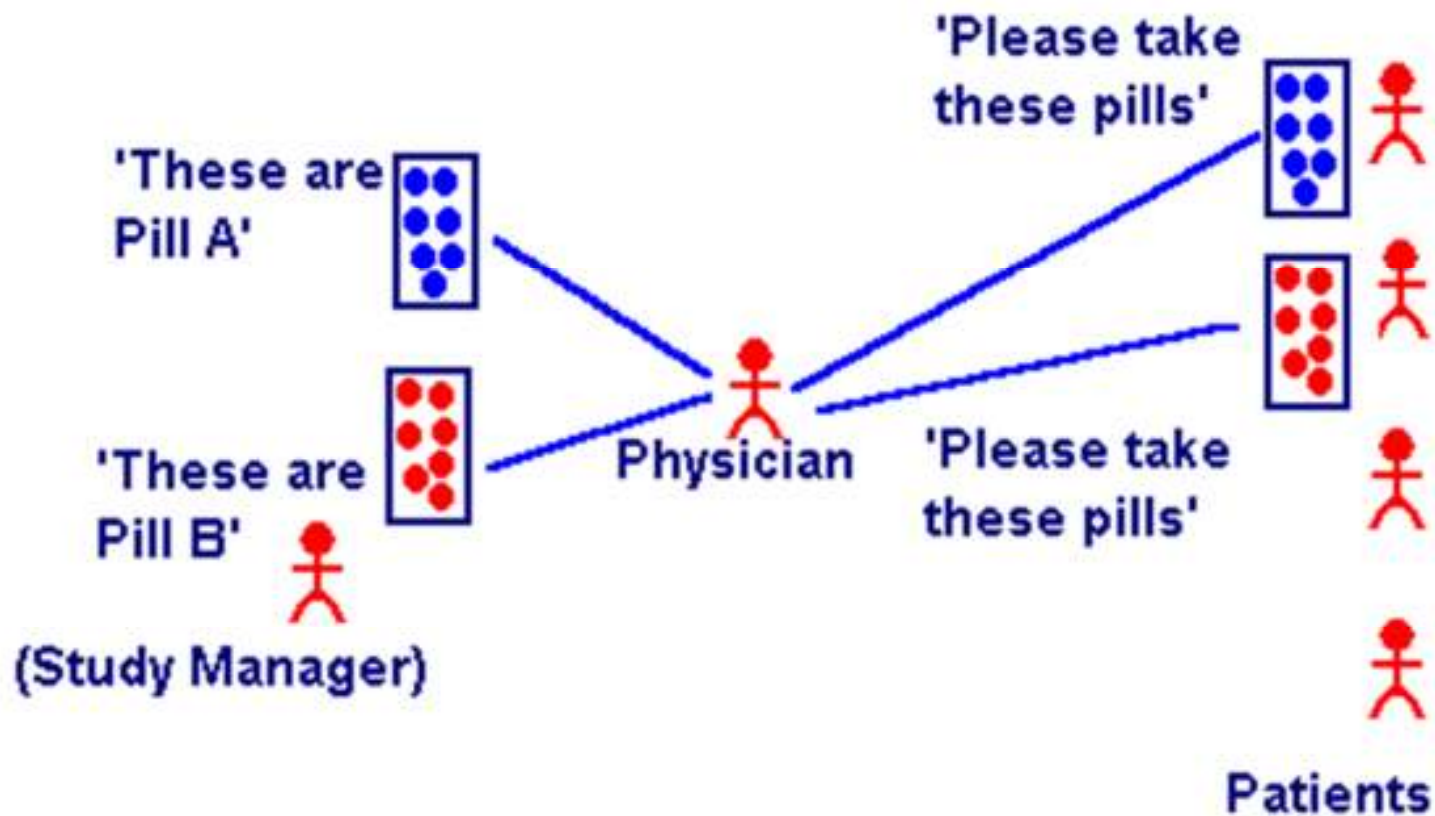
Plan an epidemiological study for the following!

- Alcohol consumption and myocardial infarction

- Workplace stress and depression

- Chewing gum and labial herpes (cold sores)

- Regular exercise and ankle injuries

- Tea consumption and gastric cancer

- Acupuncture therapy and C-type viral hepatitis

A brief guideline for analyzing epidemiological studies I.

1. What is the primary objective of the study?
2. What is the central hypothesis?
   - Are there others?
3. What does the investigators' causal model look like?
4. What type of study design was used?
   - Why do you think it was selected?
   - What are the general strengths and weaknesses of the type of study conducted?
   - What strengths and weaknesses can you identify in this particular study?
5. How was the outcome of interest defined?
6. How was the exposure of interest defined?
   - What criteria were applied?
   - How was the exposure measured?
   - What do we know about the quality of the exposure measurement?

A brief guideline for analyzing epidemiological studies II.

7. What are the potential sources of error?

   - Random error?

   - Systematic error (bias)?

   - Confounding?

8. In what ways could the concerns you identified in #7 influence the findings?

9. Is there an association observed?

   - How would you state/interpret the results?

10. Where do the findings agree or diverge from those described or alluded to in the introduction and/or discussion?

11. How strong is the case for causality?

12. What are the public health implications of the study?

13. Did the study provide novel information in general? What is it's scientific relevance?

Class exercise

You will be presented with two case studies. Try to answer at least some of the following questions in each case:

1. What was the general aim of the study?
2. What was the exposure of interest?
3. What was the primary outcome measure?
4. What was the original hypothesis of the authors?
5. What type of study design did they employ?
6. Why do you think they chose that study type?
7. What advantages and disadvantages does the study type have?
8. What potential confounding factors can you identify?

Breast feeding and risk of breast cancer in young women

Women diagnosed with breast cancer between 1982 and 1985 living in 11 health regions in Britain were selected for the study if diagnosis was made before their 36th birthday. One control per case, matched for age, was selected at random from the list of the case's general practitioner. Age matching was achieved by selecting control pairs who's date of birth was at most six months apart from the respective case. A further criteria for selecting controls was that the control had to be registered at the GP-s office before the diagnosis of her corresponding case pair was established. If a patient could not be interviewed, neither was their control pair. If a selected control person could not be interviewed a second was selected (and so forth until a control interview was finally obtained). Both cases and controls were limited to Caucasian females who had not had prior malignancies and were not suffering from severe mental retardation or any psychiatric conditions.

Study participants were contacted in their homes by trained survey personnel between January 1984 and February 1988. Corresponding cases and controls were interviewed by the same person. Of the 1049 women who met case-selection criteria, 755 (75%) could be interviewed. Of the 755 initially selected control pairs, 675 (89%) were willing to participate in the study. A second control pair had to be selected in 68 cases and a third in 12 cases.

Teenagers, sex, and risk taking

This study provides a baseline examination of teenage sexual relationships in 1991 in a sample drawn from 9 state and independent schools in southwest England; a follow up was conducted in 1992. 1025 students in year 11 and aged 15-16 responded in 1991; the follow up in 1992 showed 315 returning the questionnaire. Information was requested on knowledge of sexually transmitted diseases (STDs), age at first intercourse, sexual intercourse without use of contraception, the frequency of "one night stands" or relationships measured in days, and knowledge of friends who had STDs. Relative risk ratios were calculated for those engaging in sexual activity before and after the age of 16 years. The results showed that in the follow up more girls and persons with more academic interests, as represented by the school subjects taken for the General Certificate of Secondary Education, responded. There were no differences by social class. The average age of the sample was 16.9 years.