Study Designs in Epidemiology

Headlines

- Epidemiological research
- Classification of designs
- Qualitative methods
- Quantitative methods
- Choice of design

Epidemiological Research

- **Lab research**: applies knowledge of basic sciences towards development of procedures and strategies to prevent, control and understand mechanisms of health-related phenomena

- **Epidemic investigations**: study of outbreaks, in local populations, to identify agent(s), transmission mode(s), and possible control measure(s)

- **Population-based (field) research**: study of distribution, determinants, control measures of health-related phenomena in chosen populations, followed by application of suitable biostatistical techniques which may allow generalization of results

Data Collection Methods

- **Primary:** where the investigator is the first to collect the data. Sources include: medical examinations, interviews, observations, etc. Merits: less measurement error, suits objectives of the study better. Disadvantage: costly, may not be feasible.

- **Secondary**: where the data is collected by OTHERS, for other purposes that those of the current study. Sources include: individual records (medical / employment); group records (census data, vital statistics)

Study design: Definition

A study design is a specific plan or protocol for conducting the study, which allows the investigator to translate the conceptual hypothesis into an operational one.

Comparison (I)

Qualitative

- Understanding
- Interview/observation
- Discovering frameworks
- Textual (words)
- Theory generating
- Quality of informant more important than sample size
- Subjective
- Embedded knowledge
- Models of analysis: fidelity to text or words of interviewees

Quantitative

- Prediction
- Survey/questionnaires
- Existing frameworks
- Numerical
- Theory testing (experimental)
- Sample size core issue in reliability of data
- Objective
- Public
- Model of analysis:parametric, non-parametric

Comparison (II)

Qualitative
- **Methods**
  - Focus Groups
  - Interviews
  - Surveys
  - Self-reports
  - Observations
  - Document analysis
  - **Sampling**: Purposive
- **Quality Assurance:**
  - **Trustworthiness**: Credibility, Confirmability, Dependability, Transferability
  - **Authenticity:** Fairness, Ontological, Educative, Tactical, Catalytic

Quantitative
- **Methods**
  - Observational
  - Experimental
  - Mixed
  - **Sampling**: Random (simple, stratified, cluster, etc) or purposive
- **Quality Assurance**:
  - **Reliability:** Internal and External
  - **Validity**: Construct, Content, Face

Qualitative Research Types

Postpositivist	Interpretivist	Critical	Alternative/ Arts-Based
does not claim to provide universal answers but seeks to ask questions instead	*multiple interpretations of the same phenomena must be allowed for, and that no truth is attainable*		
Grounded Theory	Ethnography *description and interpretation of a cultural or social group or system*	Critical Theory	Personal Experience
	Phenomenology: *the science or study of phenomena, things as they are perceived*	Feminist	Narrative Inquiry
	Case Study		Performance
	Life Story/Oral History		Portraiture
	Biography		Collage

Qualitative Research Techniques

- **Participant observation (field notes)**
- **Interviews / Focus group discussions with key infomants**
- **Video / Text and Image analysis (documents, media data)**
- **Surveys**
- **User testing**

Involves Skills of

- **Observing**

- **Conversing**

- **Participating**

- **Interpreting**

Qualitative Techniques (I)

- **Participant observation**
  - Gains insight into understanding cultural patterns to determine what's necessary and needed in tool development (complementary to interviews)
- **Interviews/Focus groups with stakeholders**
  - Explores how tools are used and could be used in a novice programming course
  - Gains insight into the meaning of tools for students for learning to program

Qualitative Techniques: (II)

- **Data analysis**
  - Themes arising from data would provide insight into current "learning to program" issues and see what is important to students / teachers / administrators
- **Survey**
  - Useful for verifying results on a larger scale
- **User Testing**
  - Useful for triangulating results

Rigor in Qualitative Research

- **Dependability**
- **Credibility**
- **Transferability**
- **Confirmability**

Quantitative Designs

Quantitative designs

- **Observational**: studies that do not involve any intervention or experiment.

- **Experimental**: studies that entail manipulation of the study factor (exposure) and randomization of subjects to treatment (exposure) groups

Observational Designs

Observation Methods

- **Selected Units**: individuals, groups
- **Study Populations**: cross-sectional, longitudinal
- **Data collection timing**: prospectively, retrospectively, combination
- **Data collection types**: primary, secondary

Study populations

- **Cross-sectional:** where only **ONE** set of observations is collected for every unit in the study, at a certain point in time, disregarding the length of time of the study as a whole

- **Longitudinal:** where **TWO or MORE** sets of observations are collected for every unit in the study, i.e. follow-up is involved in order to allow monitoring of a certain population (cohort) over a specified period of time. Such populations are AT RISK (disease-free) at the start of the study.

Observational Designs (Classification I)

- **Exploratory**: used when the state of knowledge about the phenomenon is poor: small scale; of limited duration.

- **Descriptive**: used to formulate a certain hypothesis: small / large scale. Examples: case-studies; cross-sectional studies

- **Analytical**: used to test hypotheses: small / large scale. Examples: case-control, cross-sectional, cohort.

Observational Designs (Classification II)

- **Preliminary** (case-reports, case-series)
- **Basic** (cross-sectional, case-control, cohort [prospective, retrospective])
- **Hybrid** (two or more of the above, nested case-control within cohort, etc)
- **Incomplete** (ecological, PMR, etc)
- **Others** (repeated, case cross-over, migrant, twin, etc)

Case-series:
Clinical case series

- **Clinical case-series:** usually a coherent and consecutive set of cases of a disease (or similar problem) which derive from either the practice of one or more health care professionals or a defined health care setting, e.g. a hospital or family practice.

- A case-series is, effectively, a **register** of cases.

- **Analyse cases together to learn about the disease**.

- Clinical case-series are of value in epidemiology for:
  - Studying symptoms and signs
  - Creating case definitions
  - Clinical education, audit and research

Case series:
Population based

- When a **clinical case-series** is complete for a defined geographical area for which the population is known, it is, effectively, a **population based case-series** consisting of a population register of cases.

- Epidemiologically the most important case-series are registers of serious diseases or deaths (usually NCDs), and of health service utilisation, e.g. hospital admissions.

- Usually compiled for administrative and legal reasons.

Case series:
Natural history and spectrum

- By delving into the past circumstances of these patients, including examination of past medical records, and by continuing to observe them to death (and necropsy as appropriate), health professionals can build up a picture of the **natural history of a disease**.

- Population case-series is a systematic extension of this series but which includes additional cases, e.g. those dying without being seen by the clinicians.

- Add breadth to the understanding of the spectrum and natural history of disease.

Case series: Population

- **Full epidemiological use of case-series data needs information on the population to permit calculation of rates**

- **Key to understanding the distribution of disease in populations and to the study of variations over time, between places and by population characteristics**

- **Case-series can provide the key to sound case control and cohort studies and trials**

- **Design of a case-series is conceptually simple**

- **Defines a disease or health problem to be studied and sets up a system for capturing data on the health status and related factors in consecutive cases**

Case series:
Requirements for interpretation

To make sense of case-series data the key requirements are:

- The diagnosis (case definition) or, for mortality, the cause of death
- The date when the disease or death occurred **(time)**
- The place where the person lived, worked etc **(place)**
- The characteristics of the person **(person)**
- The opportunity to collect additional data from medical records (possibly by electronic data linkage) or the person directly
- The size and characteristics of the **population at risk**

Case series: Additional data

- **Case-series data can be linked to other health data either in the past or the future, e.g. mortality data can be linked to hospital admissions including at birth and childhood, cancer registrations and other records to obtain information on exposures and disease.**

- **Cases may also be contacted for additional information.**

- **This type of action may turn a case-series design into a cohort design.**

Case series: Strengths

Population case-series permit two arguably unique forms of epidemiological analysis and insight.

- Paint a truly national and even international **population perspective on disease**.

- The disease patterns can be **related to aspects of society or the environment** that affect the population but have no sensible measure at the individual level e.g. ozone concentration at ground level and the thickness of the ozone layer in the earth's atmosphere.

Cross-sectional Studies
(Community health studies, surveys)

- **Characteristics:** detects point prevalence; relative conditions; allows for stratification

- **Merits:** feasible; quick; economic; allows study of several diseases / exposures; useful for estimation of the population burden, health planning and priority setting of health problems

- **Limitations:** temporal ambiguity (cannot determine whether the exposure preceded outcome); possible measurement error; not suitable for rare conditions; liable to survivor bias

- **Effect measure:** Odds Ratio

Case - Control Studies

- **Characteristics**: two source populations; assumption that non-cases are representative of the source population of cases.

- **Merits**: least expensive; least time-consuming; suitable for study of rare diseases (especially NCDs)

- **Limitations**: not suitable for rare exposures; liable to selection bias and recall bias; not suitable for calculation of frequency measures.

- **Effect measure:** Odds Ratio

Cohort Studies

- **Characteristics**: follow-up period (prospective; retrospective)
- **Merits:** no temporal ambiguity; several outcomes could be studied at the same time; suitable for incidence estimation
- **Limitations (of prospective type):** expensive; time-consuming; inefficient for rare diseases; may not be feasible
- **Effect measure:** Risk Ratio (Relative Risk)
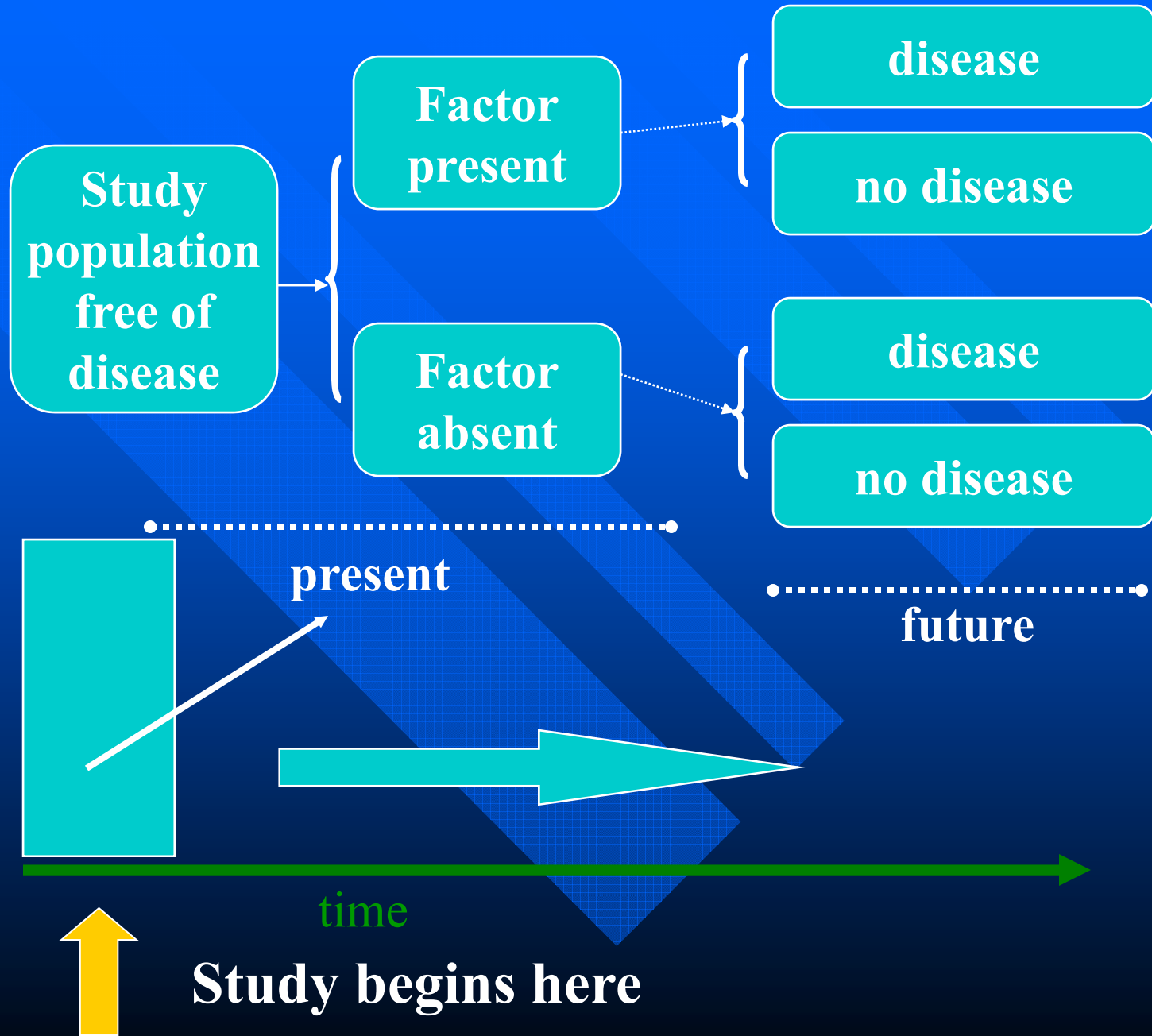
Ecological studies (I)

- **These are studies where exposure data relating to a place (say hardness of water, which could be collected on individuals) are correlated with health data collected on individuals but summarised by place (say CHD rates).**

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

- **What is missing: relationship between exposure and outcome at the individual level (incomplete design)**

Ecological studies (II)

- **Cross-sectional, case-control and cohort studies and trials (and not just population case-series) could also be analysed in relation to such "ecological" variables and such units of analysis.**

- **Most ecological analyses are based on population case-series.**

- **Ecological analyses are subject to the ecological fallacy.**

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.

Experimental Designs

Experimental Study Design

A study in which a population is selected for a planned trial of a regimen, whose effects are measured by comparing the outcome of the regimen in the **experimental** group versus the outcome of another regimen in the **control** group. Such designs are differentiated from observational designs by the fact that there is **manipulation of the study** **factor** (exposure), and **randomization** (random allocation) of subjects to treatment (exposure) groups.

Why Performed ?

1. Provide stronger **evidence** of the effect (outcome) compared to observational designs, with maximum confidence and assurance

2. Yield more **valid results**, as variation is minimized and bias controlled

3. Determine whether experimental treatments are safe and effective under **"controlled environments"** (as opposed to "natural settings" in observational designs), especially

   when the margin of expected benefit is doubtful / narrow (10 - 30%)
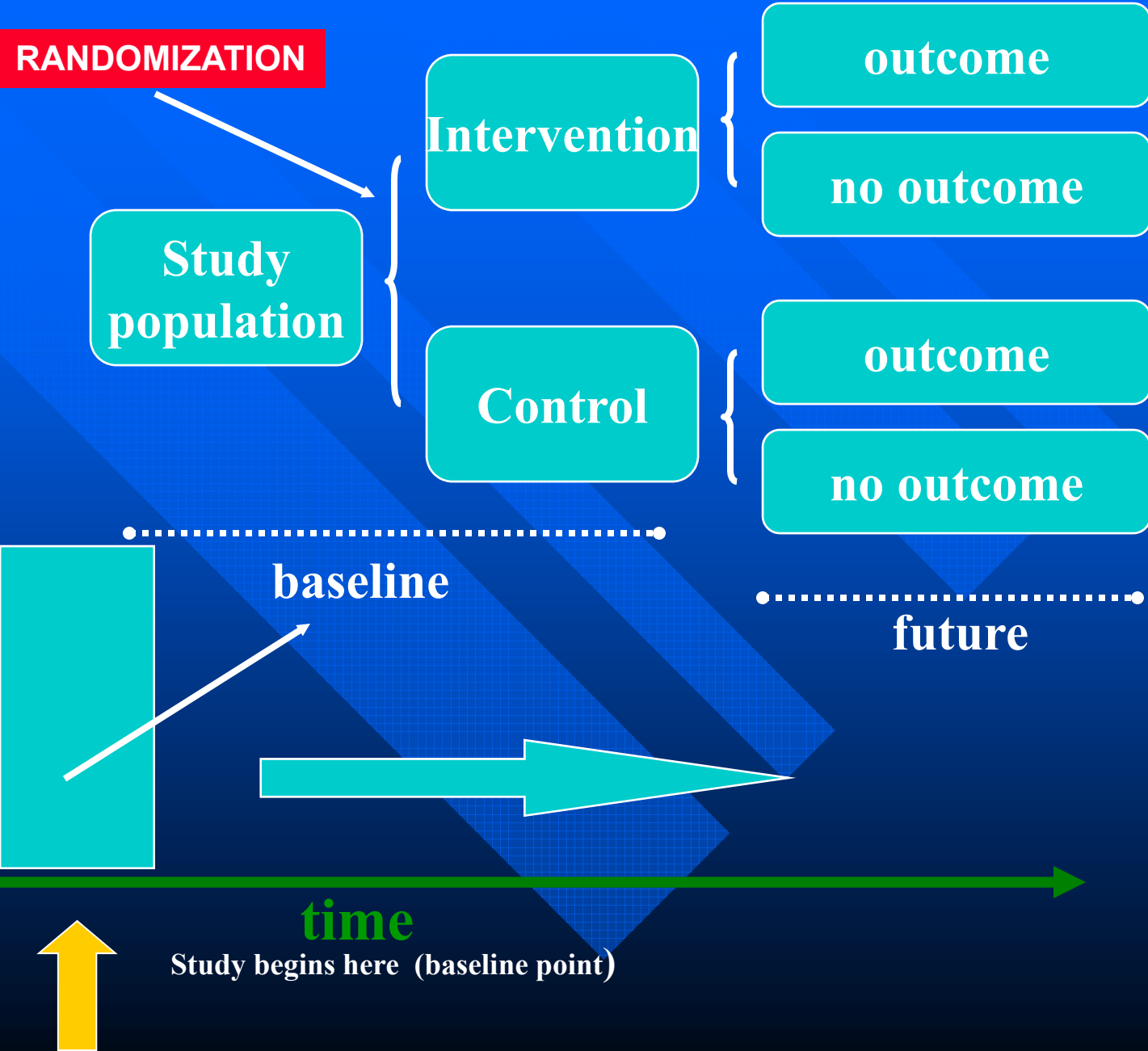
Types of trials
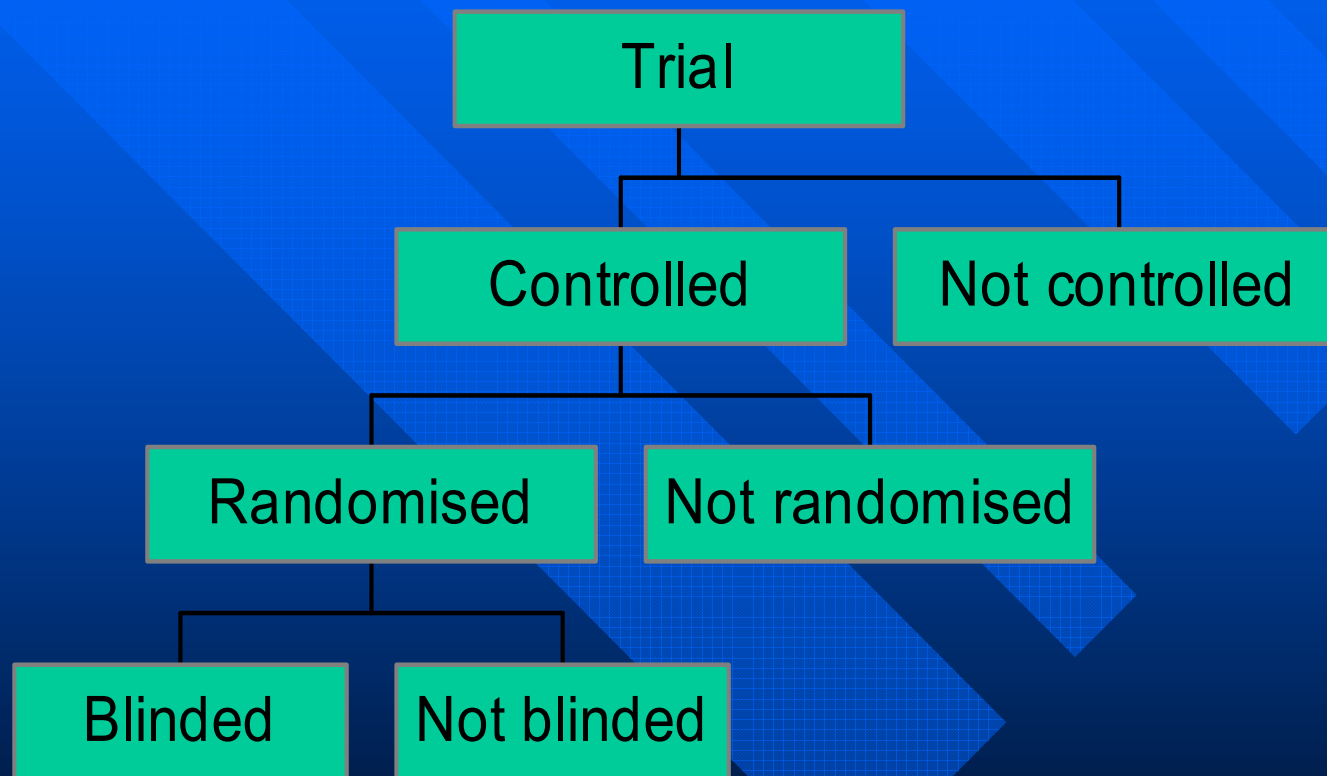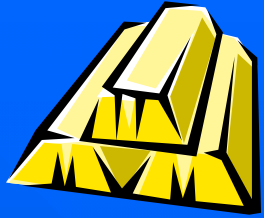
RCT Advantages (I)

- the "gold standard" of research designs. They thus provide the most convincing evidence of relationship between exposure and effect. Example:

  » *trials of hormone replacement therapy in menopausal women found no protection for heart disease, contradicting findings of prior observational studies*

RCT Advantages (II)

- **Best evidence study design**
- **No inclusion bias (using blinding)**
- **Controlling for possible confounders**
- **Comparable Groups (using randomization)**

RCT Disadvantages

- **Large trials (may affect statistical power)**
- **Long term follow-up (possible losses)**
- **Compliance**
- **Expensive**
- **Public health perspective ?**
- **Possible ethical questions**

Choice of Design (I)

Depends on:
- Research Questions
- Research Goals
- Researcher Beliefs and Values
- Researcher Skills
- Time and Funds

Choice of design (II)

It is also related to:
- **Status of existent knowledge**
- **Occurrence of disease**
- **Duration of latent period**
- **Nature and availability of information**
- **Available resources**

Comparing study designs

- **Theme**
- **Ease**
- **Timing**
- **Maintenance and continuity**
- **Costs**
- **Ethics**
- **Data utilisation**
- **Main contribution**
- **Observer bias**
- **Selection bias**
- **Analytic output**

Conclusion (I)

- **Qualitative designs are complementary to quantitative designs, are important in study of social determinants of health problems**
- **Quantitative designs have a common goal to understand the frequency and causes of health-related phenomena**
- **Seeking causes starts by describing associations between exposures (causes) and outcomes**

Conclusion (II)

- **Case-series is a coherent set of cases of a disease (or similar problem).**
- **Cases are compared with reference group, we have a case control study**
- **In a population studied at a specific time and place (a cross-section) the primary output is prevalence data, though association between risk factors and disease can be generated.**
- **In cross-sectional studies, we are looking for both exposure and outcome**
- **In case-control studies, we know the outcome, looking for the exposure**
- **In cohort studies, we know the outcome, following up looking for the outcome in question**

Conclusion (III)

- If the population in a cross-sectional survey is followed up to measure health outcomes, this study design is a cohort study.
- If the population of such a study are, at baseline, divided into two groups, and the investigators impose a health intervention upon one of the groups the design is that of a trial.
- Studies based on aggregated data are commonly referred to as ecological studies.
- Mostly, ecological studies are mode of analysis, rather than a design.
- Interpretation and application of data are easier when the relationship between the population observed and the target population is understood
- RCTs represent the "gold standard" of research designs. They thus provide the most convincing evidence of relationship between exposure and effect..