University of Zambia
School of Health Sciences
Department of Biomedical Sciences

Research Methodology, Biostatistics and Epidemiology

BMS 4430/RAD 4620

Introduction to Research, Biostatistics and Epidemiology

S. M. Munsaka, BSc., MSc., PhD

20th February 2023

Research

- Research in common parlance refers to a search for knowledge.
- Once can also define research as a scientific and systematic search for pertinent information on a specific topic.
- In fact, research is an art of scientific investigation.
- The Advanced Learner's Dictionary of Current English lays down the meaning of research as "a careful investigation or inquiry specially through search for new facts in any branch of knowledge."

Aims of research

- To gain familiarity with a phenomenon or to achieve new insights into it (studies with this object in view are termed as *exploratory or formulative research studies);*
- To portray accurately the characteristics of a particular individual, situation or a group (studies with this object in view are known as *descriptive research studies);*
- To determine the frequency with which something occurs or with which it is associated with something else (studies with this object in view are known as *diagnostic research* studies);
- To test a hypothesis of a causal relationship between variables (such studies are known as *hypothesis-testing research studies).*

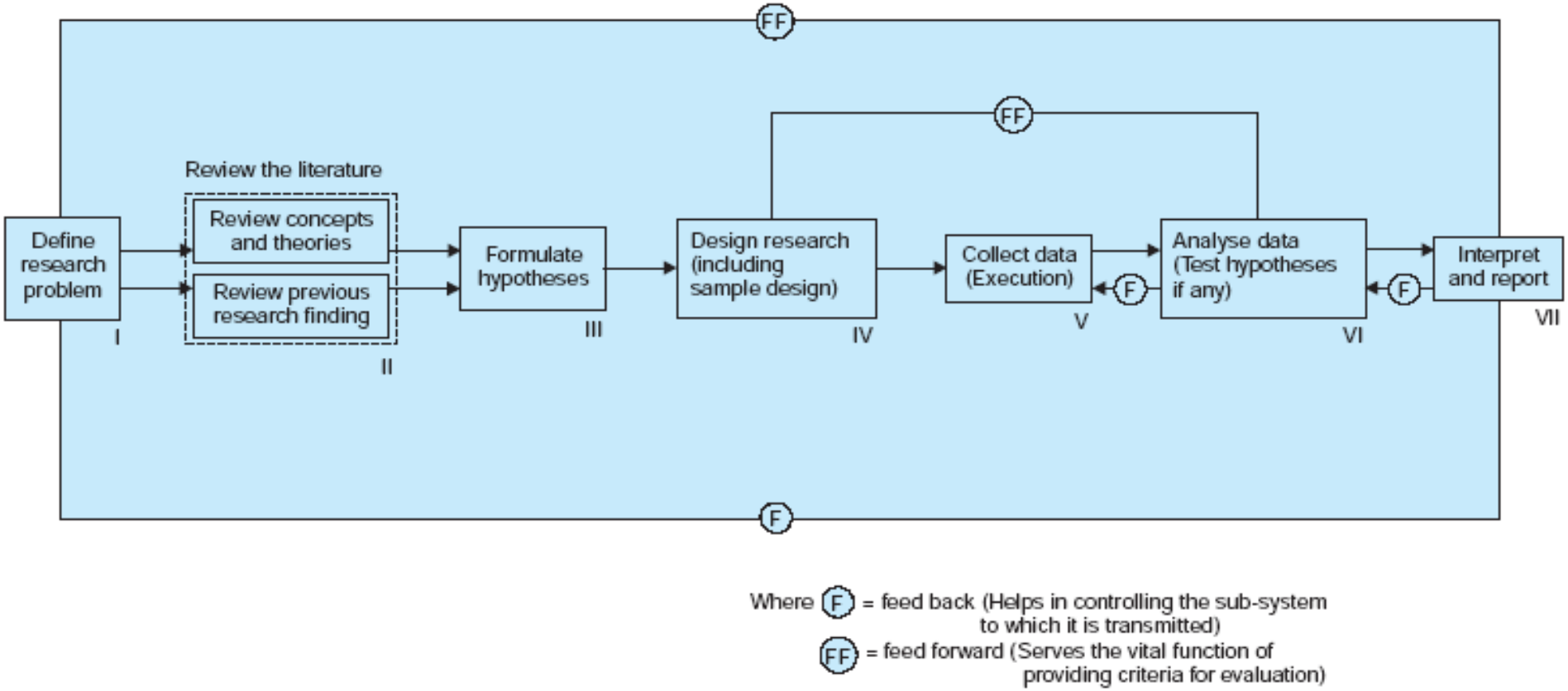
Types of research

- *Descriptive vs. Analytical:*
  - *Descriptive research includes surveys and fact-finding enquiries* of different kinds. Analytical studies identify and quantify associations, test hypotheses, identify causes and determine whether an association exists between variables, such as between an exposure and a disease.

- *Applied vs. Fundamental:*
  - *Research can either be applied (or action) research or* fundamental (to basic or pure) research.
  - *Applied research aims at finding a solution for an* immediate problem facing a society or an industrial/business organisation, whereas *fundamental research is mainly concerned with generalisations and with the formulation of a theory.*

- *Quantitative vs. Qualitative:*
  - *Quantitative research is based on the measurement of quantity* or amount. It is applicable to phenomena that can be expressed in terms of quantity.
  - Qualitative research, on the other hand, is concerned with qualitative phenomenon, i.e., phenomena relating to or involving quality or kind

Research Process



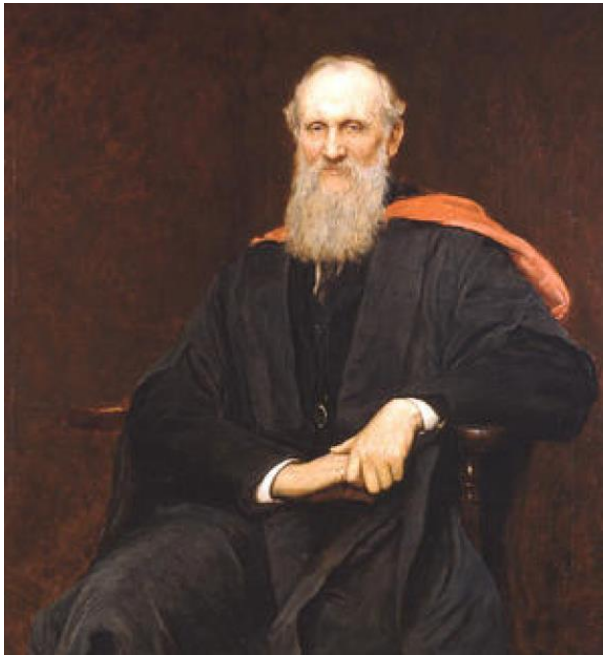
Introduction to statistics

- Statistic or datum: a measured or counted fact or piece of information presented as a figure
  - E.g. height, weight, age, etc
- Statistics or data (plurals) are collected from
  - Experiments/measurements/observations
  - Records
  - Surveys
- Applications: all walks of life including medicine and public health (Biostatistics)
- Statistics is the science of figures; field concerned with the collection of data, classification, summarizing, interpretation, drawing inferences, testing hypothesis, making recommendations etc

Introduction to Biostatistics

- Biostatistics: statistics applied to biological sciences (medicine and public health)

"I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely in your thoughts advanced to the state of Science, whatever the matter may be." Lord Kelvin, 1883

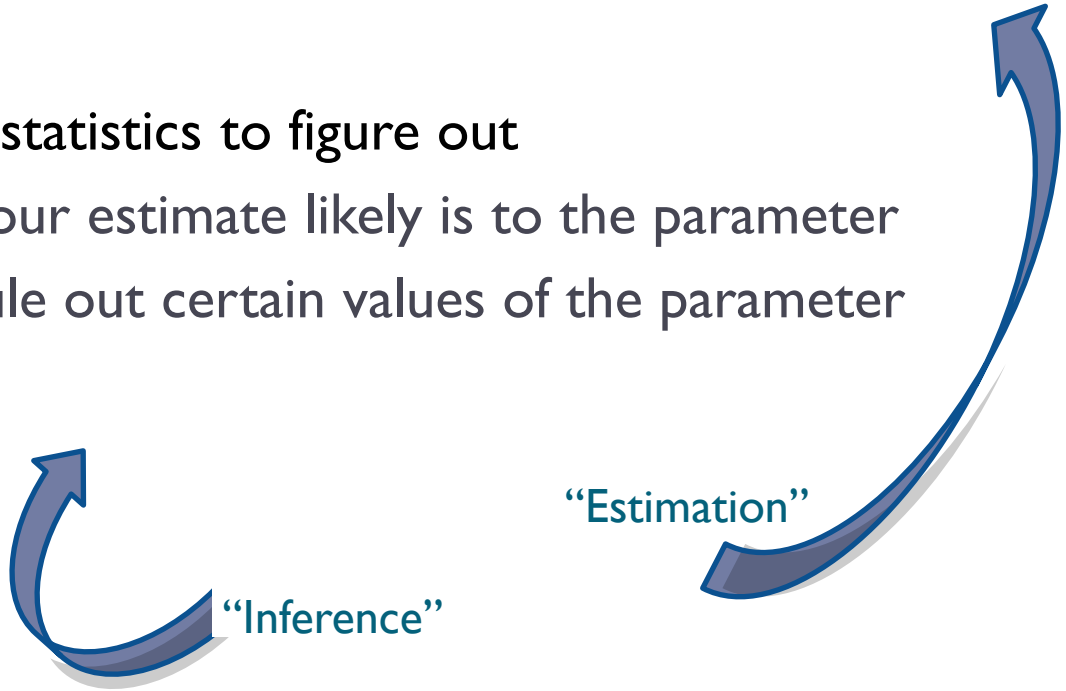


Lord Kelvin, 1883

What is Statistics

- Statistics assumes there is an *unknown, true* value out there
- Instead we collect a *sample* and use our sample to *estimate* the population parameter

- Then we use statistics to figure out
  - how close our estimate likely is to the parameter
  - if we can rule out certain values of the parameter



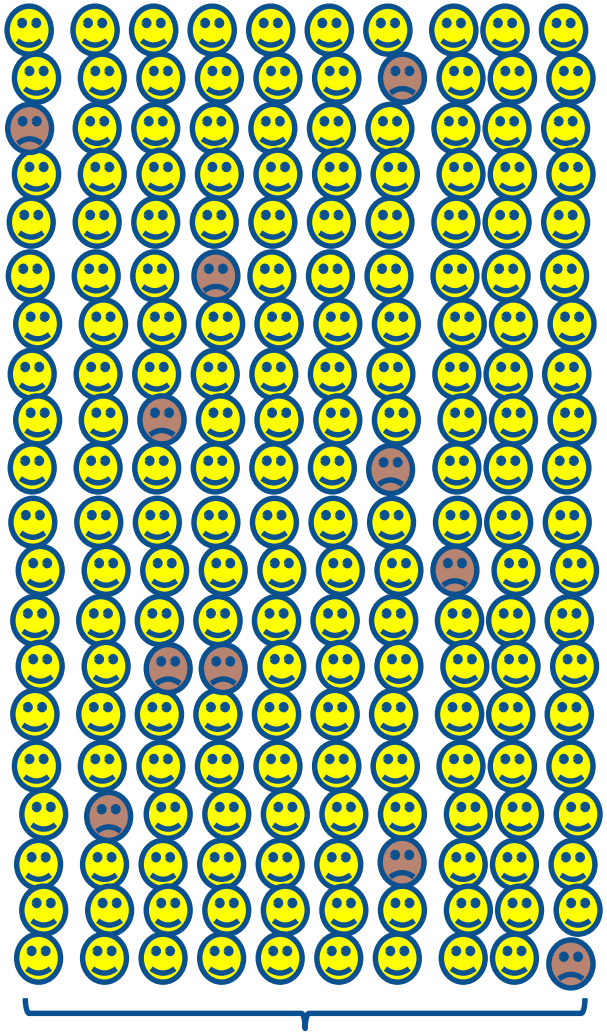
Example: Estimation

Suppose we want to know the prevalence of a certain disease in a specific population

Truth:

- Population is 200 people
- 10 people in population have disease
- Prevalence of Disease is 10/200 = 5%

If we could test all 200 people, we would know the true prevalence (*the "parameter"*).



Population

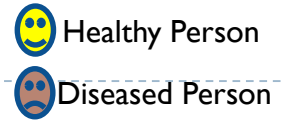
Healthy Person

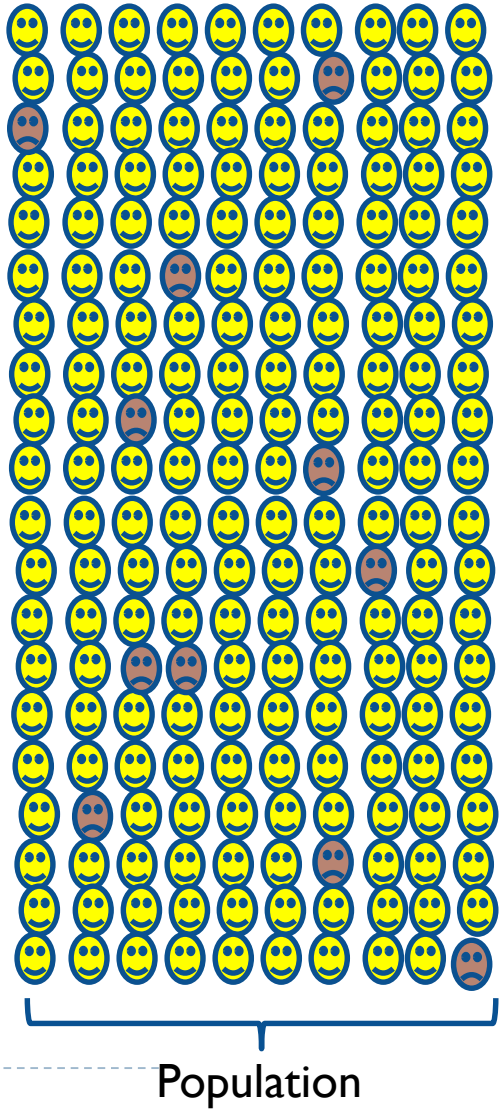
Diseased Person

Example: Estimation

We cannot measure all 200, so we take a *sample*

In this case, 2/20 in our sample have disease, so our *estimate* for the prevalence is 10%





Introduction
B: Types of Studies

- Many studies are one of two types:
  - **Observational:** You observe data but don't actively intervene
    - E.g., compare mortality in breast cancer patients choosing radical mastectomy versus lumpectomy
  - **Experimental:**
    - **Randomized control trials** (RCT): Randomly assign patients to treatments
    - E.g., randomly assign half the patients to radical mastectomy and half to lumpectomy

Introduction
B: Types of Studies

- Big difference between observational studies and randomized control trials:
  - In observational studies, patients choosing versus not choosing treatment likely to differ in ways other than just treatment
    - May be more health-conscious (better diet, more exercise, etc., visit doctor more often, etc.)
  - In randomized control trials, randomization makes treatment groups comparable with respect to other factors like diet, exercise, etc.

Introduction
B: Types of Studies

- If see difference in mortality between groups:
  - In observational study, can't conclude treatment caused difference in mortality

    Maybe health-conscious behavior caused difference

  - In clinical trials, can conclude that treatment caused difference in mortality

    Randomization makes groups comparable in ways other than treatment received (e.g., health-conscious behavior)

Introduction
B: Types of Observational Studies

- Two common observational studies:
  - Cohort study: Study one group (cohort) with risk factor (e.g., smoking) & another without
    - Which group has more disease over next 5 years?
    - Prospective study: looking from now to future
  - Case/control study: Compare group with disease (cases) to group of controls
    - Which group engaged in risk factor more?
    - Retrospective: looking backward in time
  - Cross-Sectional Study
    - Assessment at one time point
    - Assessment of exposure and outcome at the same time

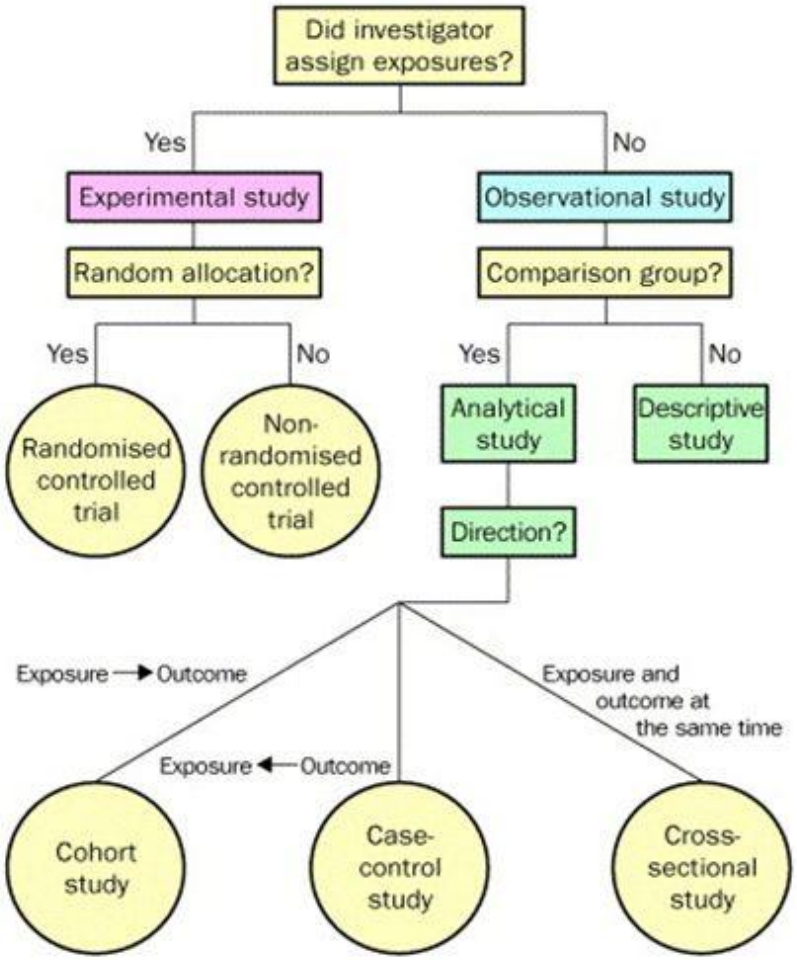
Types of Epidemiological Studies

Observational Studies

- **Case-Control Study**
  - Inexpensive to conduct
  - Small number of subjects
    - Controls represent general population
  - Able to study rare diseases
    - Prevalence < 5%
  - Results obtained quickly
  - Recall bias
  - Matching of controls may mask or give a spurious association
  - Limited to 1 outcome variable
  - Can not calculate incidence rates, prevalence, excess risk

- **Cohort Studies**
  - Expensive to conduct
  - Large number of subjects
  - Disease detected in yearly check-ups
  - Study time is long
  - Establishes sequence of events
    - Subjects are classified in relation to exposure before becoming ill
  - Study several outcome variables
  - Calculate incidence rates, prevalence and excess risk

Observational studies



Many types of studies - right research design depends on the question we ask

Observational studies

- Large proportion of research
- Can be valuable (e.g. AE) but also many disadvantages (confounding, bias)

Without comparison group – **descriptive:**

(do not try to qualify the relationships but give us a perspective of what is happening in the population, prevalence or experience of a group)

- Case reports, case series, qualitative studies, some cross-sectional studies (surveys)

With comparison group - **analytical**

(attempts to qualify relationship between two factors – effect of an intervention / exposure on an outcome)

- Cohort studies, case-control studies, some cross-sectional studies

Introduction
B: Types of Studies

- Usual sequence:
  - Epidemiologists analyzing observational data identify risk factor
    - Cholesterol and heart disease
    - Smoking and lung cancer, etc.
  - Other observational studies and animal studies confirm relationship
  - Intervention developed to alter risk factor
  - RCT definitively answers whether intervention alters mortality/heart attacks, etc.
- Observational studies & clinical trials often complement each other
  - Consistent results bolster evidence

Why Measurement?

- Measurements supply the numbers used in statistical analysis.

- No matter how profound the theoretical formulations, how sophisticated the design, and how elegant the analytic techniques, researchers cannot compensate for **poor measures**.

- We will discuss the *role* of measurement, and the *types* of measurement.

Types of Measurements (Variables)

- **Categorical (Nominal)**
  - Mutually exclusive and exhaustive
  - No particular order
  - E.g. profession, Disease condition, Sex
    - Binary (dichotomous): two outcomes
      - E.g. Sex; male and female, HIV status; Positive and negative

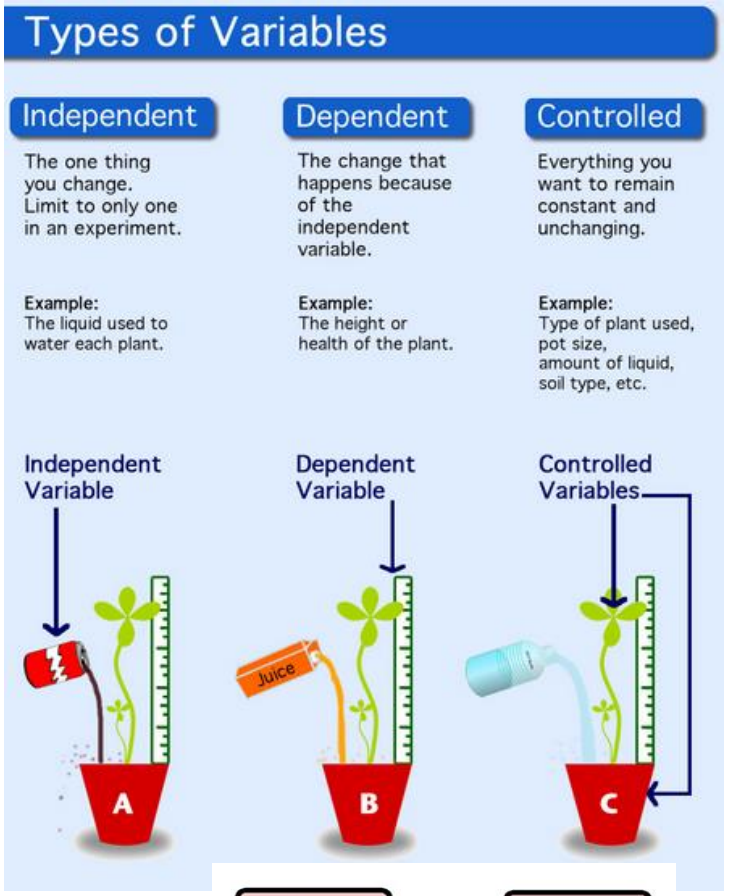
- **Ordinal**
  - Categorical measurements (variables) that can be ordered
    - E.g. severity of symptoms, age ranges

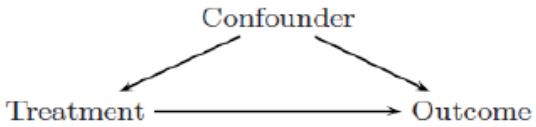
- **Continuous (interval)**
  - Numerical variable with many possible values, no upper limit

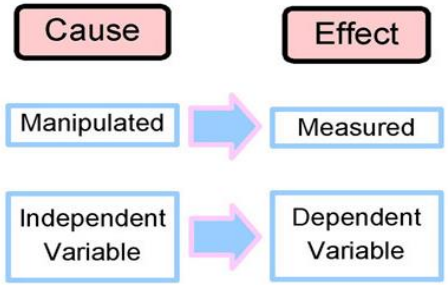
- **Discrete variables** e.g. 1, 2,3,4 etc
  - Has the most statistical information
    - Egg age, blood pressure, blood glucose levels, CD4, viral load etc

Types of Measurements (Variables)

- Dependent variable (Experimental or Response variable)
  - The variable that you measure
  - The **dependent variable** responds to the **independent variable**. It is called **dependent** because it "depends" on the **independent variable**.
  - E.g. outcomes; death, CD4, viral load, blood glucose
- Independent variable (predictor or descriptor variable)
  - Not controlled by experimenter
  - E.g. race, sex, HIV status
- Adjustment variable (cofounder)
  - A variable that can affect both the dependent and independent variable

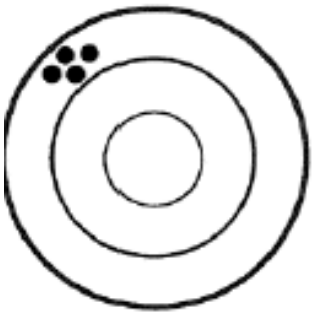




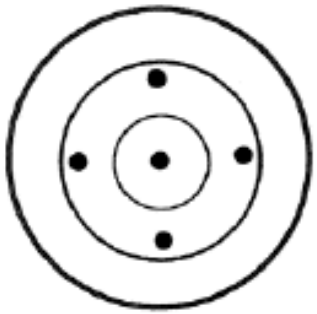


Precision

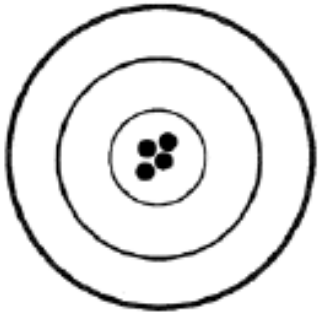
- Definition: the degree to which a variable has nearly the same value when measured several times.
- Example: Child's weight is not always the same if they are moving.



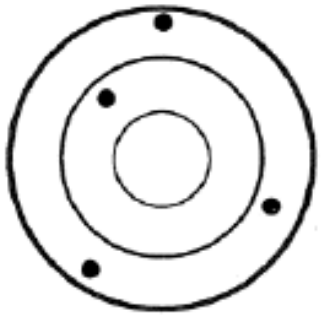
Good precision
Poor accuracy



Poor precision
Good accuracy



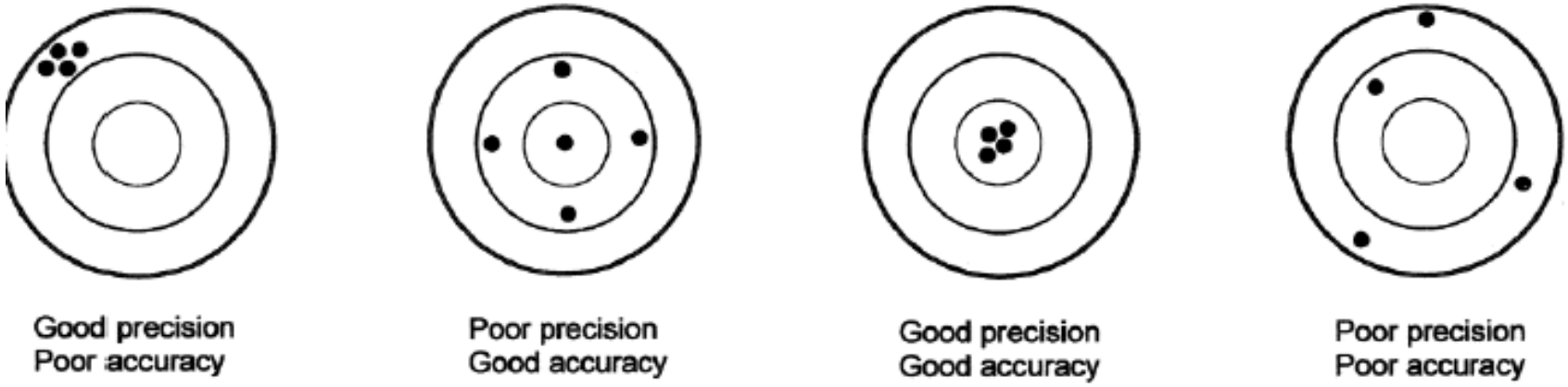
Good precision
Good accuracy



Poor precision
Poor accuracy

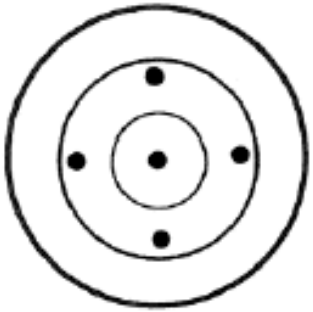
Accuracy

- Definition: the degree to which a variable actually represents what it is supposed to measure.
- Example: If the scale is not at zero with no weight, then readings will not be correct.

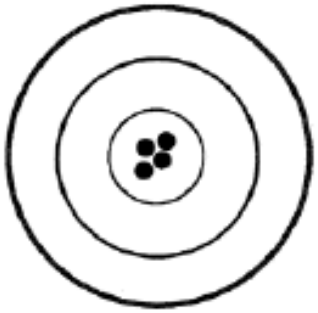


Precision and accuracy

Good precision
Poor accuracy



Poor precision
Good accuracy



Good precision
Good accuracy

Poor precision
Poor accuracy

Infectious Disease Epidemiology

Definition

Epidemiology focuses on the distribution and determinants of disease frequency in human populations.

The purpose of Epidemiology is to prevent illness. To do so, health care providers must be able to both identify persons who, because of personal characteristics or their environment, are at high risk for illness AND to intervene to reduce that risk.

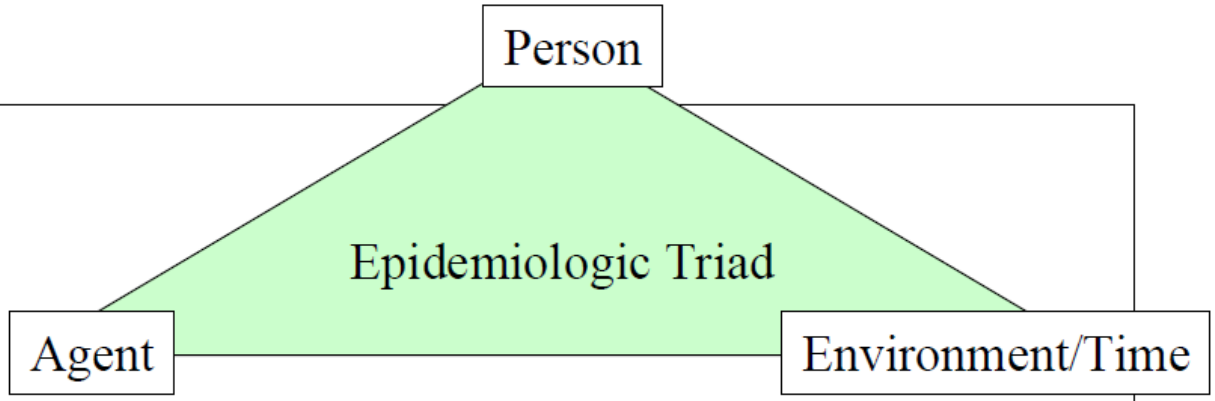
Classification of Infectious Diseases

Epidemiology Point of View

Transmission	Characteristics
Contact	Direct or indirect contact with an infected person Direct = open sore, inoculation Indirect = fomite
Food- or Water-borne	Ingestion of contaminated food or water
Airborne	Inhale infectious particle
Vector-borne	Invertebrate transmission
Perinatal	Contact occurs in utero

Categorization of Disease Patterns

- Person
- Place
- Time
- Infectious Agent
  - Reservoirs: Humans, Animals, Environment
  - Portals of Entry
  - Transmission
  - Host Immunity

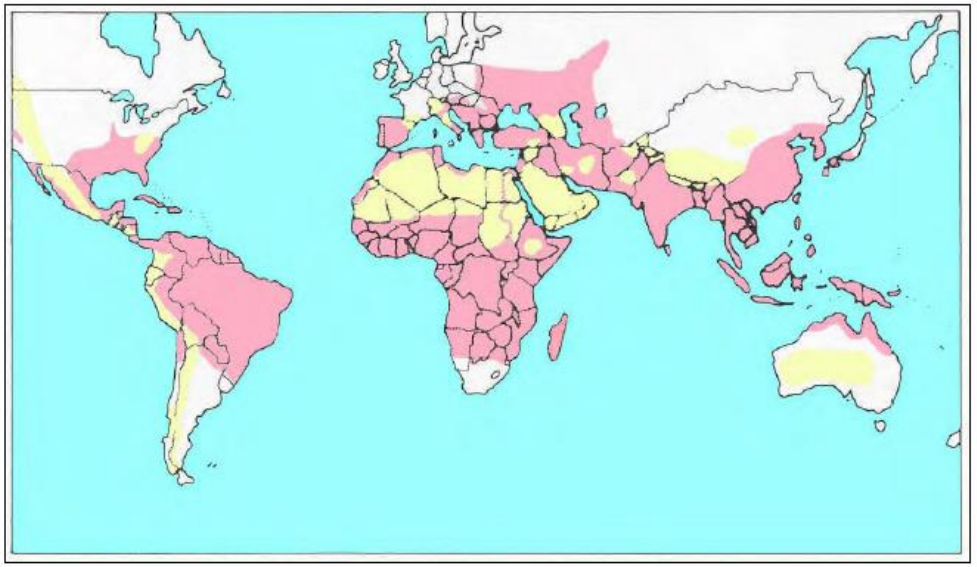


Categorization of Disease Patterns

- Person
  - Demographic characteristics (gender, age)
  - Social and economic factors (income, occupation)
  - Personal habits (tobacco, diet, exercise)
  - Genetic characteristics (blood groups, HLA type)
  - Biological characteristics (antibody titer)

Categorization of Disease Patterns

- Place
  - Temporal vs Tropical
    - Ecology
  - Urban vs Rural



Categorization of Disease Patterns

- Time
  - Cyclic Changes
    - Recurrent Alterations in the Frequency of Disease
  - Epidemics
    - Level of Disease Above What is Expected

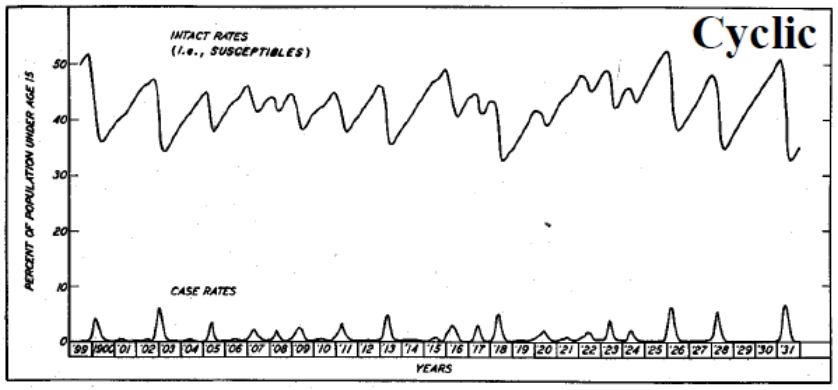
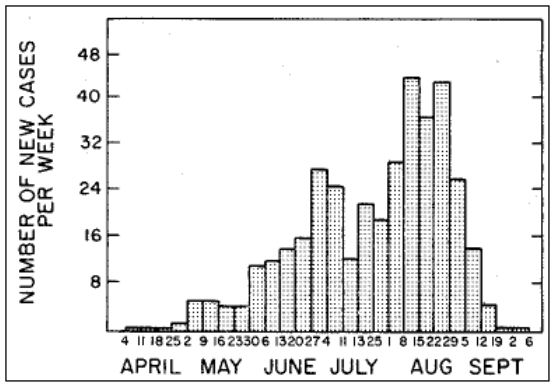
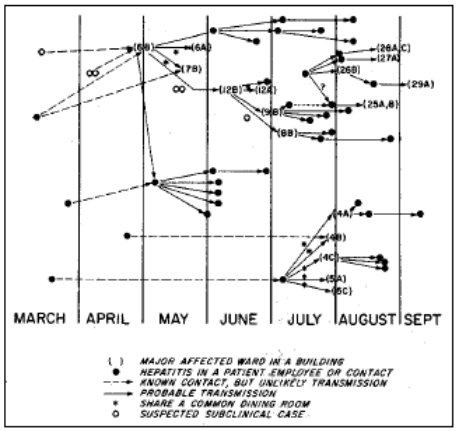


Figure 12–9 Estimated complete monthly attack rates from measles, and intact rates (proportions not previously attacked) for the population under age 15, Old Baltimore, Md., July, 1899–December, 1931. (From Hedrich, A. W.: Monthly estimates of the child population "susceptible" to measles, 1900–1931, Baltimore, Md. Am. J. Hyg. *17*:626, 1933.)

Epidemic of Hepatitis A in an Institution for the Mentally Retarded





Endemic vs Epidemic

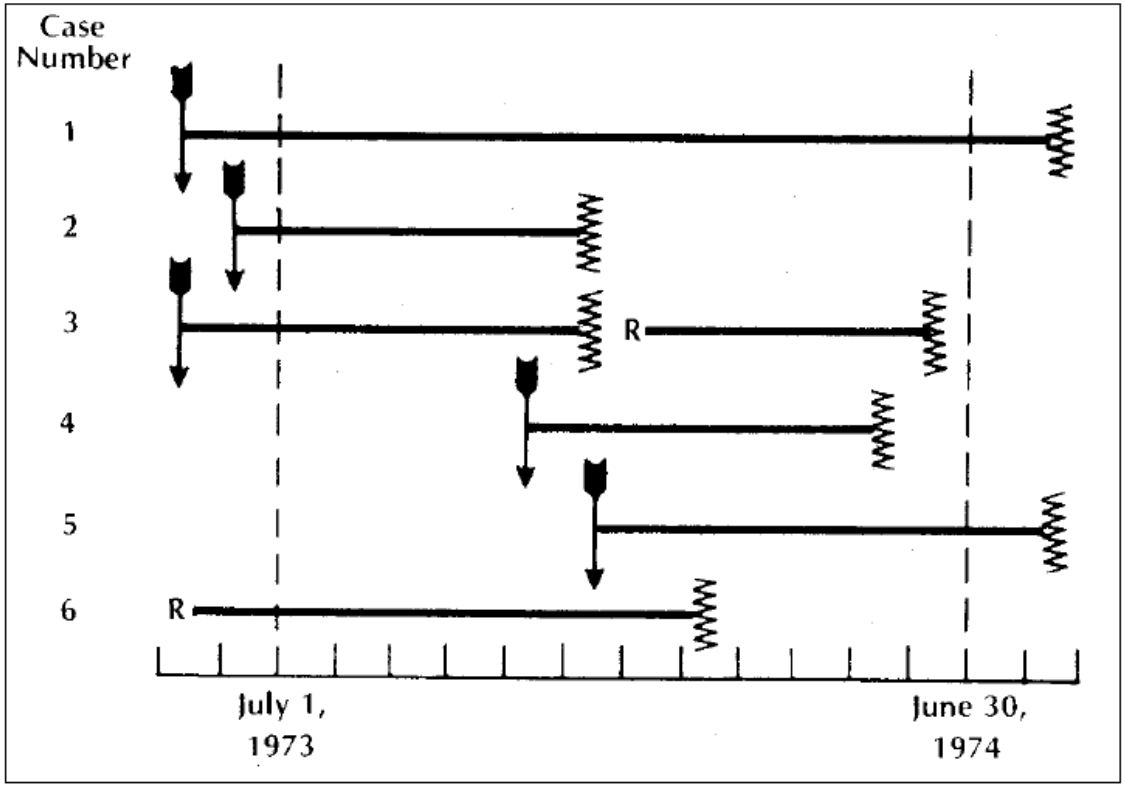
- Endemic
  - An infection by an organism that is always present in the population or environment
- Epidemic
  - An unusual increase in the number of infections by an organism in a population
- Pandemic
  - An epidemic that reaches worldwide proportions

Measures of Morbidity

- Prevalence Rate
  - Number of existing cases of a disease / total population for a point in time
  - Dependant on the number of new cases and duration of illness
    - P ≅ Incidence x duration of illness
  - Intervention programs which increase life will also increase prevalence
  - Prevalence rates are useful for the planning health care workloads

Prevalence Rates

Total Population 250

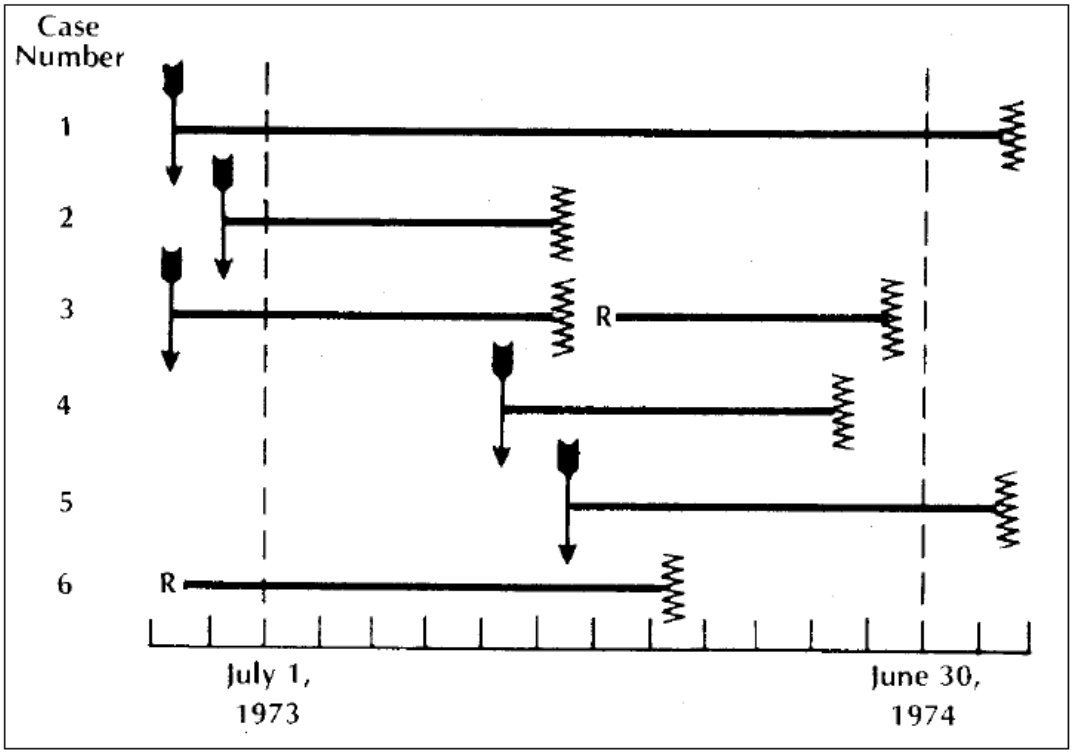


Measures of Morbidity

- Incidence Rate
  - Number of new cases of a disease / population at risk of infection for a given period of time
  - Rate at which new illnesses occur
    - Measurement of the "risk" for developing the disease
  - By comparing incidence rates for a given risk factor among two populations, the importance and magnitude of the risk factor can be determined
  - Influenced by control programs

Incidence Rate

Total Population 250



Attack Rate

Definition

- The Attack Rate for an Illness Measures the Proportion of the Population that Develops the Disease Among the Total Number of Individuals With a Specific Risk Over a Specified Time Period:

$$\frac{\text{Number of People Ill in the Time Period}}{\text{Number of People at Risk in the Time Period}}$$

Types of Epidemiological Studies

Observational Studies

- **Case-Control Study**
  - Inexpensive to conduct
  - Small number of subjects
    - Controls represent general population
  - Able to study rare diseases
    - Prevalence < 5%
  - Results obtained quickly
  - Recall bias
  - Matching of controls may mask or give a spurious association
  - Limited to 1 outcome variable
  - Can not calculate incidence rates, prevalence, excess risk

- **Cohort Studies**
  - Expensive to conduct
  - Large number of subjects
  - Disease detected in yearly check-ups
  - Study time is long
  - Establishes sequence of events
    - Subjects are classified in relation to exposure before becoming ill
  - Study several outcome variables
  - Calculate incidence rates, prevalence and excess risk

Risk Factors

Modifiable vs Non-Modifiable

- Non-Modifiable Risk Factors
  - Age
  - Gender
  - Race/Ethnic Group
- Modifiable Risk Factors
  - Smoking
  - Weight
  - Exercise
  - Diet

Strategies for Controlling Infectious Diseases

- Biological and Environmental Control
- Changing Human Behavior
- Vector Control
- Mass Drug Treatment
- Mass Vaccination