HARVARD T.H. CHAN SCHOOL OF PUBLIC HEALTH

HEIDELBERG INSTITUTE OF GLOBAL HEALTH

UNIVERSITY OF KWAZULU-NATAL ™
INYUVESI YAKWAZULU-NATALI

MUHAS

MAKERERE UNIVERSITY

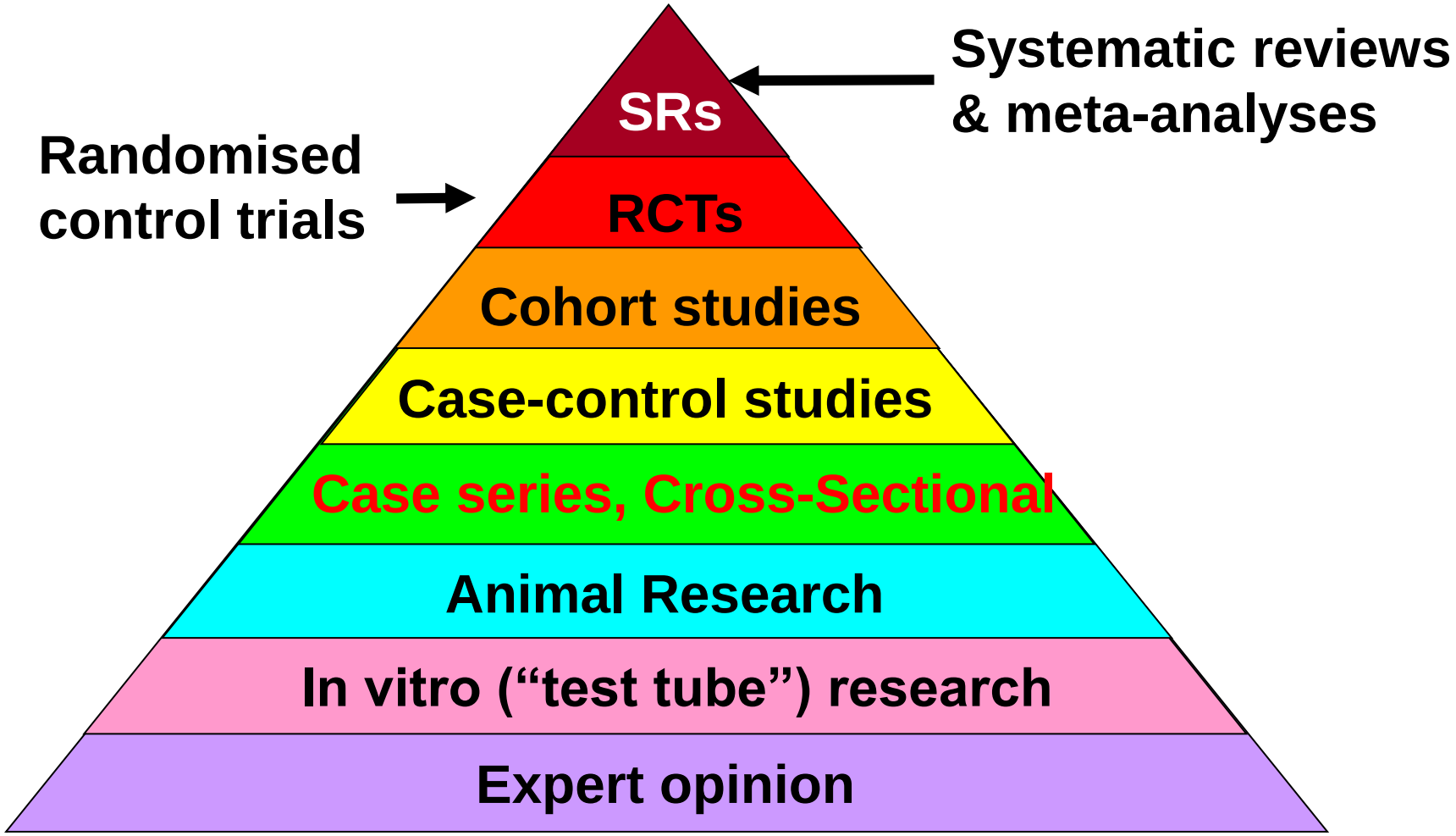
UNIVERSITY OF GHANA

WASHA Takwimu

Introduction to Study Design

Saloshni Naidoo

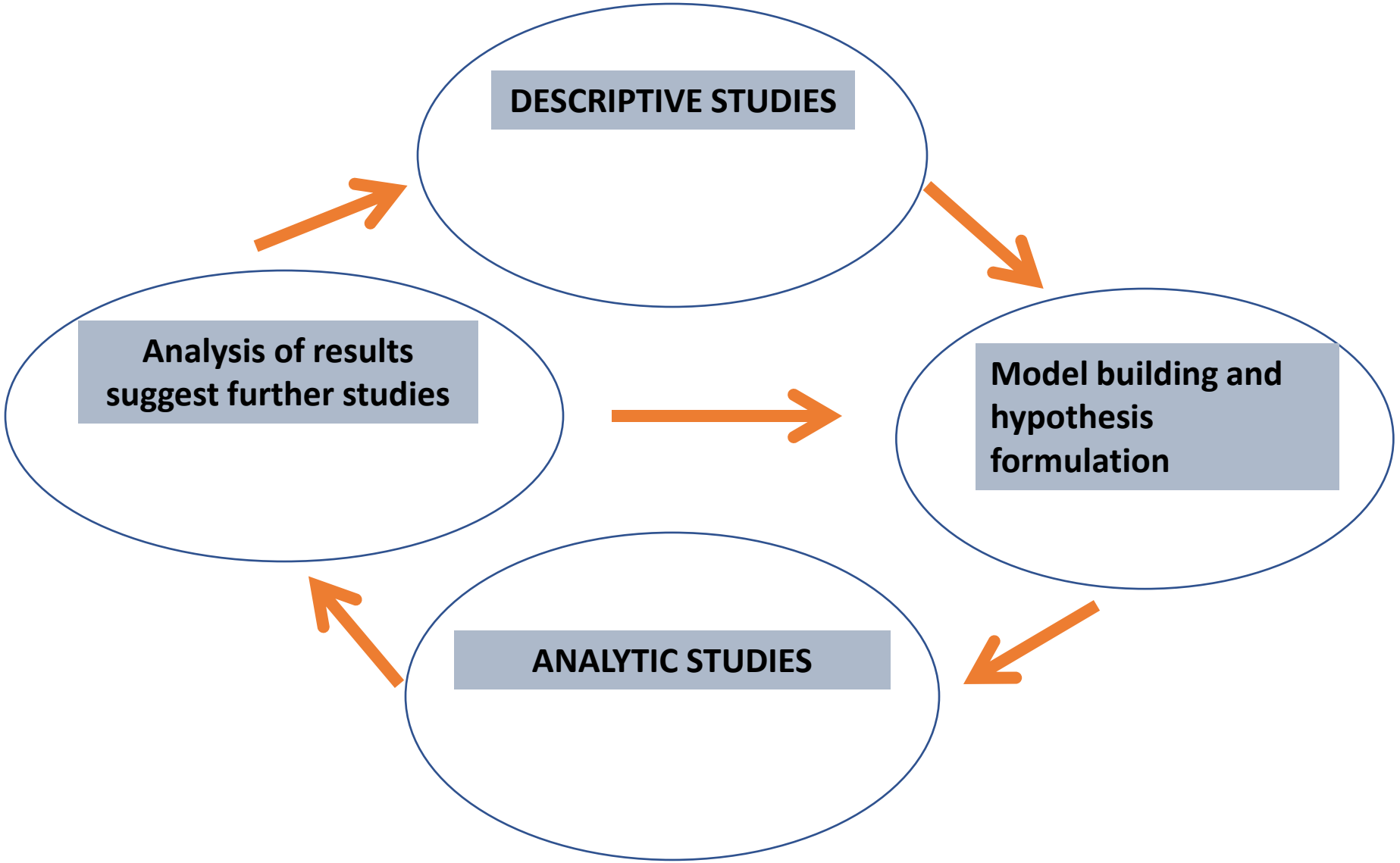
The evidence pyramid



Study Type	Observation Unit	Measure of Occurrence	Measure of Association
Observational Studies			
Descriptive			
Descriptive	Individuals	Frequency	None
Cross-sectional	Individuals	Prevalence	None
Analytic			
Cross-sectional	Individuals	Prevalence	PR, POR, PD
Case-control	Individuals	None	OR
Cohort	Individuals	Incidence (risk or rate)	RR, RD
Ecological	Groups/ Populations	varied	Correlation coefficient
Experimental studies			
Randomised controlled trials	Individuals	Incidence (risk or rate)	RR, RD
Cluster randomsied trials	Groups/ populations	Incidence (risk or rate)	RR, RD

Ehrlich & Joubert 2014

STUDY DESIGN CYCLE



Case – control Study

exposure	disease	person
?	*	+

Cohort Study

exposure	disease	person
*	?	+

Cross sectional Study

exposure	disease	person
?	?	+

UNIVERSITY OF KWAZULU-NATAL ™

INYUVESI YAKWAZULU-NATALI

CROSS-SECTIONAL STUDIES

2020

SALOSHNI NAIDOO

naidoos71@ukzn.ac.za

Acknowledging Dr S Knight & Dr N Tathia

Cross-sectional studies

- Describes health of populations
- Measures prevalence of health outcomes or determinants of health in a population at a point in time or over a short period of time
- Sample of subjects from defined population is selected
- Information on exposures & outcomes collected simultaneously at a SINGLE POINT IN TIME e.g. DHS
- Selection of sample is vital
  - Random sample avoids selection bias

OBSERVATIONAL STUDIES

Case – control Study

exposure	disease	person
?	*	+

Cohort Study

exposure	disease	person
*	**?**	+

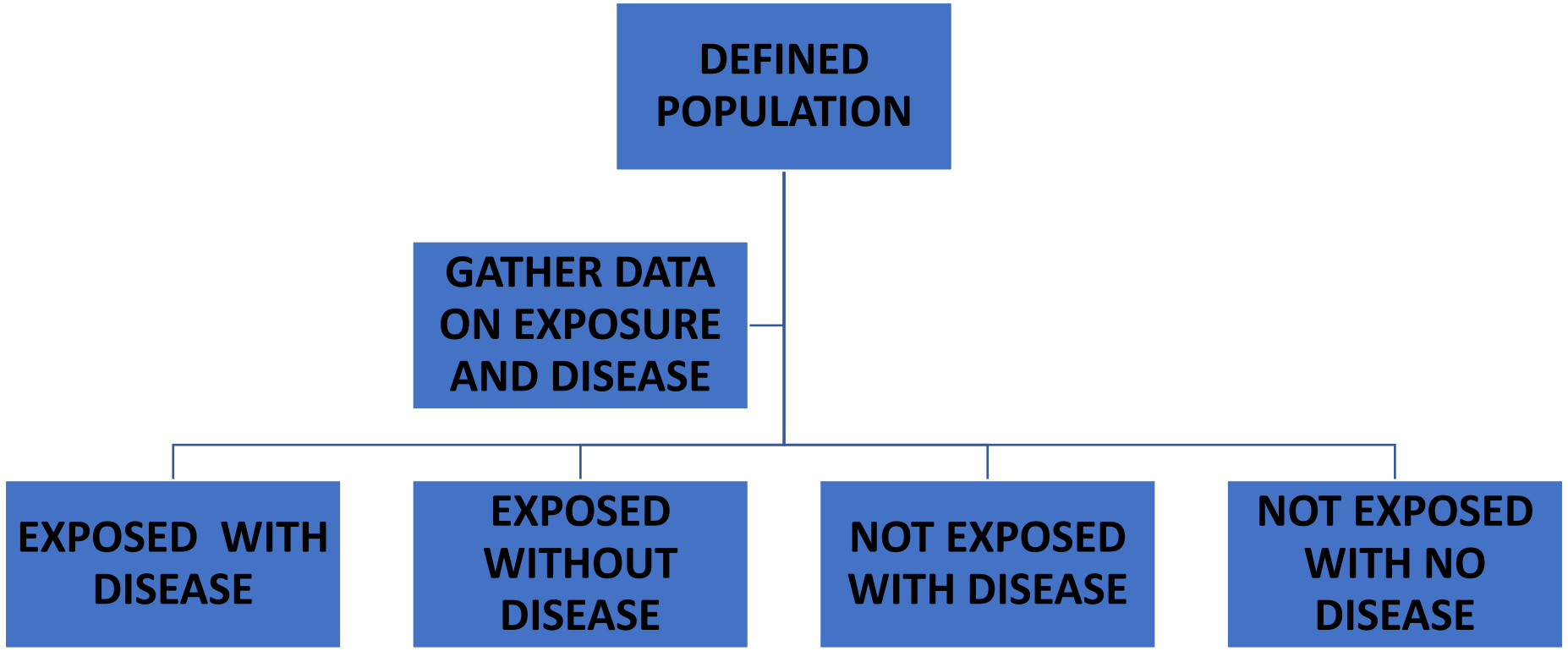
Cross sectional Study

exposure	**disease**	**person**
?	**?**	**+**

EXAMPLES OF CROSS-SECTIONAL STUDIES

- Global
  - National Health and Nutrition Examination Survey (NHANES-USA)
- South Africa
  - Demographic Health Survey in South Africa (DHS) -up-to-date estimates of basic demographic and health indicators.
  - National Income Dynamics Study (NIDS)

DESIGN OF CROSS-SECTIONAL STUDY



	Disease	No Disease
Exposed	**a**	**b**
Not Exposed	**c**	**d**

	Disease	No Disease
Exposed	**a**	**b**
Not Exposed	**c**	**d**

	Disease	No Disease
Exposed	**a**	**b**
Not Exposed	**c**	**d**

Measures to consider in a cross-sectional study

- Prevalence
  - all cases / total population at risk at a point in time
- Prevalence differences
  - prevalence (exposed) – prevalence (unexposed)
- Prevalence ratios
  - analogous to a risk ratio in a cohort study
- Prevalence Odds ratios
  - analogous to an odds ratio

	Disease	No Disease
Exposed	a	b
Not Exposed	c	d

Prevalence of exposure in Disease and No Disease

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

	Disease	No Disease
Exposed	a	b
Not Exposed	c	d

Prevalence of disease in Exposed compared to Not Exposed

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

Example

- 1998 SA Demographic and Health Survey
- Collected information of alcohol dependence (using a validated questionnaire) and gender among 13 826 randomly sampled people over the age of 14 living in South Africa
- Result:
  - SA men 2.8 more likely than women to be alcohol dependent (prevalence ratio)
  - SA men 3.5 greater odds than women to be alcohol dependent (prevalence odds ratio)

Example

	Step 1: select study sample and determine who has exposure and outcome			
	Alcohol Dependent	Not Alcohol Dependent	Totals	**Step 2: calculate prevalence of outcomes and compare two groups**
Men	1587 a	4082 b	5669	1587/5669
Woman	816 c	7341 d	8157	816/8157
Totals	2403 e	11423 f	13826	

Prevalence = e/total
= 2403/13826
= 0.17
Expressed as a percentage =17%

Prevalence ratio = [a /(a+b)] ⁄ [c /(c+d)]
= (1587/5669)/(816/8157)
= 0.279/0.1
= 2.79

Example

	Step 1: select study sample and determine who has exposure and outcome			
	Alcohol Dependent	Not Alcohol Dependent	Totals	**Step 2: calculate prevalence of outcomes and compare two groups**
Men	1587 a	4082 b	5669	1587/5669
Woman	816 c	7341 d	8157	816/8157
Totals			13826	

Prevalence odds ratio = ad/bc

= (1587 X 7341) / (4082 X 816)

= 3.49

Potential problems and mitigation

- Sampling bias
  - Use a large sample so that inferences can be made regarding the whole population
- To avoid sampling error
  - Researcher needs to select a representative sample of the population
  - An appropriate sampling frame should be identified
  - All subjects within that sampling frame should be included in the sample
  - If not done, this causes sampling errors

Potential problems and mitigation

- Selection bias
  - Incidence-prevalence bias – Neyman's bias
  - Is a type of selection bias
  - Mix of incident and prevalent cases in a cross sectional study
  - Subjects who survive longer are more likely to be selected into the cross sectional study and are thus over represented in the study
  - Healthy worker effect is a type of selection bias where healthy subjects are more likely to be included in the study
- Misclassification
  - E.g. Social desirability bias – over or under reporting of information depending on its desirability
- Confounding
  - Presence of a 'third variable'
- Lack of information on exposures
  - Recall bias – exposure that happened in the past may be forgotten or inaccurately reported

Potential problems and mitigation

- To avoid non-response
  - Identify the potential non-responders and reasons for non-response
  - Improve response by repeated contact and addressing specific issues around non-response e.g. sensitive questions
  - Adjust for non-response in the prevalence estimate using modelling estimates

Strengths of Cross-Sectional Studies

- Primary method of estimating prevalence
- Relatively inexpensive
- Logistically efficient
  - Relatively fast (no follow-up required)
  - Can enroll large numbers of participants
- Large surveys can be used for many exposures & diseases
- Often generalizable
  - BUT smaller subpopulations can be over sampled

Limitations of Cross-Sectional Studies

- Large numbers needed for rare exposures / outcomes
- No information on timing of outcome relative to exposure
  - Limits causal inference
- Includes only those individuals alive at the time of the study
  - Prevalence-incidence bias (see Case-Control Studies)
  - Cases also often defined at the time of the study

OBSERVATIONAL STUDIES

Case – control Study

exposure	**disease**	**person**
?	*****	**+**

Cohort Study

exposure	disease	person
*	**?**	+

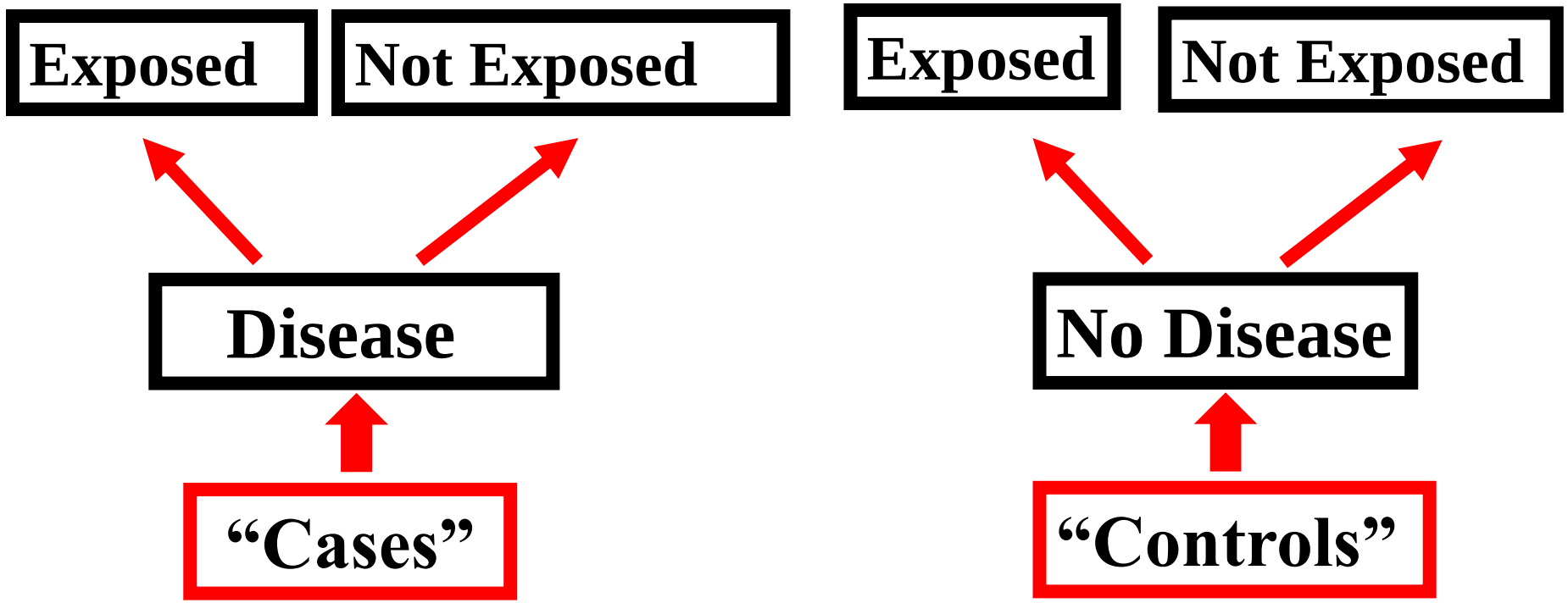
Cross sectional Study

exposure	disease	person
?	**?**	+

CASE CONTROL STUDY

- Case-control studies work backwards – start with outcome and look backwards for the exposure, e.g. lung cancer and smoking, neurodevelopmental delays and prenatal chemical exposure
    - Well-suited to study rare diseases
    - Cases (with disease) & Controls (no disease)
    - Obtain data on past exposure in both groups
    - Compare rates of exposure in both groups
    - Measure of association = ODDS RATIO
    - When the disease is rare the Odds Ratio approximates the risk ratio

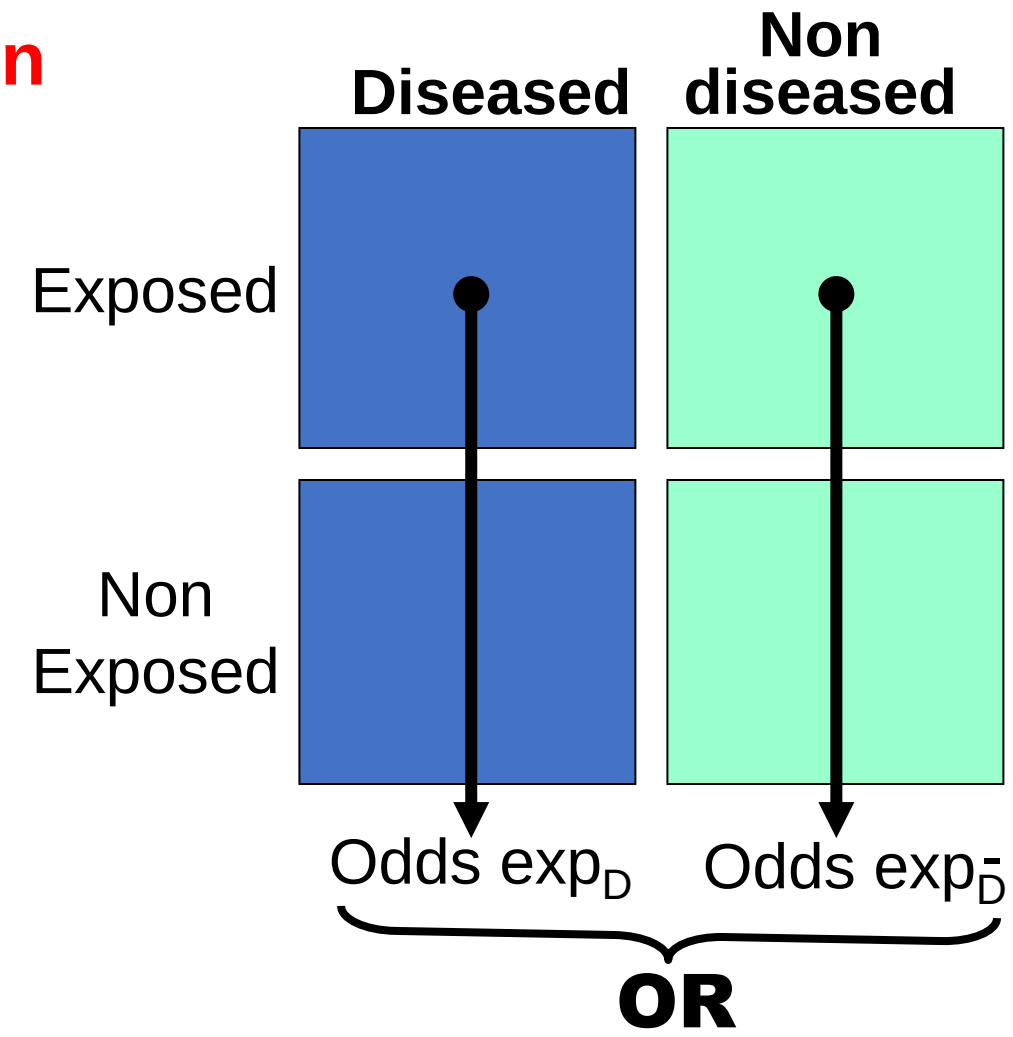
CASE CONTROL STUDY



Case-control Studies

Population selected based on outcome status

Exposure status assessed



Measure of Association

- Odds Ratio

	cases (D+)	controls(D-)
yes Exposed	a	b
no	c	d

Odds Ratio

Numerator

- Odds of exposure in cases **(a/c)**

Denominator

- Odds of exposure in controls **(b/d)**

Measure of Association

- OR

$$\frac{\frac{a}{c}}{\frac{b}{d}} = \frac{a \times d}{b \times c}$$

Measure of Association

- Estimates the magnitude and direction of an association between exposure and disease
- >1 indicates increased odds of exposure among the diseased
- <1 indicates decreased odds of exposure among cases as opposed to controls

Example B: Odds Ratio

Research question: Is smoking cigarettes during pregnancy a potential cause of offspring attention-deficit hyperactivity disorder (ADHD)?

Sample:

- 500 10-year-old children in Pholeleni who are seeking care for hyperactivity
- For each child we find with ADHD, we select two children of the same age from the same clinics who present for routine well visits (do not have ADHD) – a **purposive sample**
- **Case control study**

Measures: Mothers respond to questions, including whether they smoked cigarettes while they were pregnant

Example B: Odds Ratio

	Diseased	Not Diseased	**Total**
Exposed	300	503	803
Unexposed	200	497	697
Total	500	1000	1500

Example B: Odds Ratio

- **Odds** of exposure among those with ADHD: a/c=300/200

- **Odds** of exposure among those without ADHD: b/d=503/497
- Odds Ratio in the case control study:

  axd / bxc= 300x497/503x200= 1.48

Odds Ratio: why we use it

- In the **case control study**, we estimated the odds of exposure among the diseased and the odds of exposure among the non-diseased.
- When we **select our cases and controls correctly,** we get an unbiased estimate of the exposure odds even though we estimate the disease odds.
- This Odds Ratio is approximately equivalent to the Risk Ratio when the **disease is rare**.

Summary: Odds Ratio

- Cannot estimate the risk of disease directly when we sample people based on whether they have the disease or not (case control study)
- Can estimate proportion exposed among diseased and non-diseased
  - Estimate Odds Ratio for exposure
- If **disease is rare** in population, the Odds Ratio approximates the Risk Ratio from a prospective study

SELECTION OF CASES

- Sources
  - Hospitals, doctors practices, clinics
  - Registries (cancer, death, birth)
  - **Bias: need to avoid issues such as referral bias**
- Incident vs prevalent cases
  - Incident: wait for new cases
  - Prevalent: larger number available
  - **Bias: prevalent cases may reflect survival related to the disease and underrepresent those who die soon after a disease**

SELECTION OF CASES

- Establish diagnostic criteria
  - Based on information available in medical record, radiology reports etc.
  - Clear cut criteria
    - Eg MDRTB- specific definition , sensitivity reports

SELECTION OF CONTROLS

- Sources
  - Non-hospitalised in the community
    - Neighbourhood controls, school rosters, insurance company lists, random digit dialling in communities, sibling
  - Hospitalised with other diseases other than the one under study
    - Captive / clearly identified
    - Use a sample of admitted/ specific for another disease

SELECTION OF CONTROLS

- Selection criteria
  - Must not have the disease under study
  - Dictated by the characteristic and source of the cases
  - Comparable information to cases
  - Select from a variety of diagnostic groups
  - Acute conditions so that earlier exposures do not influence
  - Avoid patients with multiple conditions
  - Avoid diagnoses related to the risk factor under study
- More than one control
  - Hospital and community controls

STRENGTHS OF POPULATION- BASED VS HOSPITAL- BASED CONTROLS

POPULATION-BASED	HOSPITAL-BASED
Source of the population is better defined	Subjects are more accessible
Easier to make sure that cases and controls derived from the same source population	Subjects tend to be more cooperative, aware of exposures
Exposure histories more likely to reflect those of persons without the disease of interest	Background characteristics if cases and controls are more balanced
	Easier to collect exposure information and biologic specimen results

WEAKNESSES OF POPULATION- BASED VS HOSPITAL- BASED CONTROLS

POPULATION-BASED	HOSPITAL-BASED
More costly, time consuming	More likely to share similar exposures with cases
Population lists not always available	Decreased generalisability
Recall of exposures not at same level as cases	
Less motivated to participate	

MATCHING

- Select cases & controls so that they are similar on certain characteristics such as age, sex, occupation, socio-economics (potential confounder)
- Group vs individual
- Group/ Frequency
  - 25% of cases and controls are married
- Individual
  - Each case is matched to a control on the chosen characteristics

ISSUES TO CONSIDER WITH MATCHING

- Unplanned
  - neighbourhood controls- socioeconomic
  - Best friends- lifestyles
- If you match on too many characteristics then you may have problems finding a control
- Matching on a variable which is not a confounder can lead to a loss in precision (especially if the variable is strongly related to the exposure)

ADVANTAGES AND DISADVANTAGES OF MATCHING

ADVANTAGES	DISADVATAGES
Provides control for a confounder which is difficult to measure	Once matched on a characteristic then cannot study that characteristic
Eliminates the need to have a complete list	Expensive and time-consuming
Provides an appropriate control group	Matching on continuous and ordinal variables - differences may persist
Provides for more adequate control of confounding by continuous variables	
Gain in precision of estimating the ORs	

BIASES

- Selection
  - An inappropriate control group is chosen
  - Response rates are low or unequal
  - Great deal of media attention about the exposure before the study takes place
- Information
  - Recall (cases remember more)
  - Observers (know the study hypotheses)
  - Classification of cases vs controls

Strengths of Case Control Studies

- Quick, inexpensive, simple
- Evaluate diseases with long latency
- Rare diseases
- Examine multiple etiological factors /exposures for a disease
- Small sample size

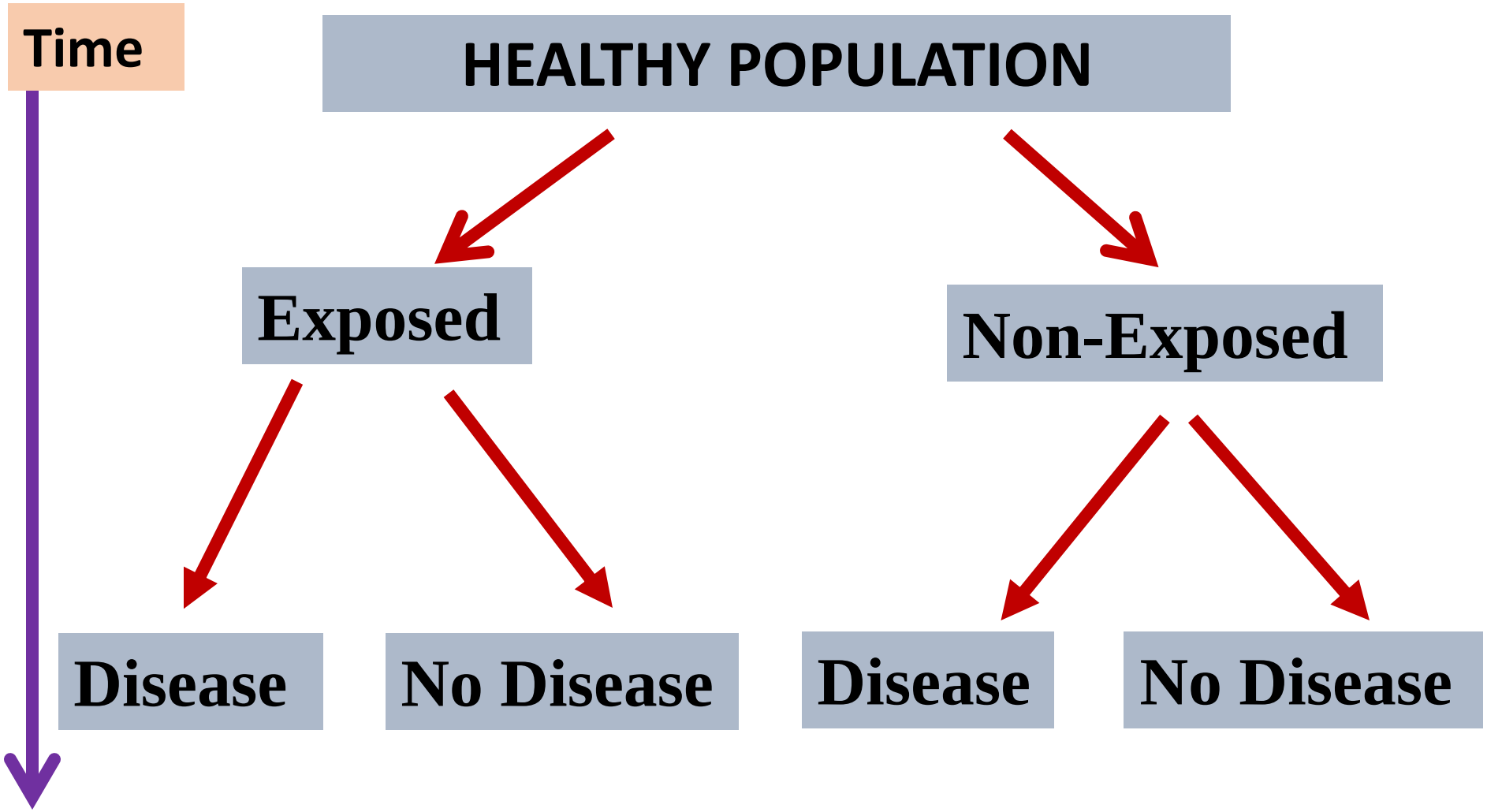
Limitations of Case Control Studies

- Inefficient for evaluating rare exposures
- Cannot compute incidence rates directly
- Temporal relationship between exposure and disease difficult
- Prone to bias
- Difficult to select appropriate controls
- Retrospectively collected exposure data
- Require large sample if exposure is rare

COHORT STUDIES

- Prospective
- Retrospective

DESIGN OF A PROSPECTIVE COHORT STUDY



MEASURE OF ASSOCIATION

- **Relative Risk**

		Diseased		
		yes	no	all
Exposed	yes	a	b	a+b
	no	c	d	c+d

Risk Ratio

- RR

$$[a/a+b)] \quad / \quad [c/(c+d)]$$

Relative Risk

- Estimates the magnitude and direction of an assoc
- > 1 indicates a harmful effect
- < 1 indicates a protective effect

Strengths of Cohort Studies

- Efficient for rare exposures
- Multiple exposures & outcomes
- Temporal relationship
- Minimal bias
- Direct measurement of disease incidence
- Cause-effect relationship

Limitations of Cohort Studies

- Expensive and time consuming
- If retrospective need records
- Large sample size
- Inefficient for rare diseases
- Validity affected by losses to follow-up
- Changes in time of criteria lead to bias

ECOLOGICAL STUDIES

- Correlation studies
- Unit of analysis is the population
- Relationships at a population level
- Example:
  - Sales of asthma medication and asthma related deaths compared between countries

Strength of a Ecological study

- Simple to conduct
- Hypothesis for other studies
- Useful when little variation in a population

Limitation

- Ecological fallacy

RANDOMISED CONTROLLED TRIALS

- A planned experiment designed to assess the efficacy of a treatment
- Comparing outcomes in a group of patients given a test treatment with those observed in a comparable group of patients given a control treatment.
- During this time both groups are enrolled, treated and followed over the same period of time

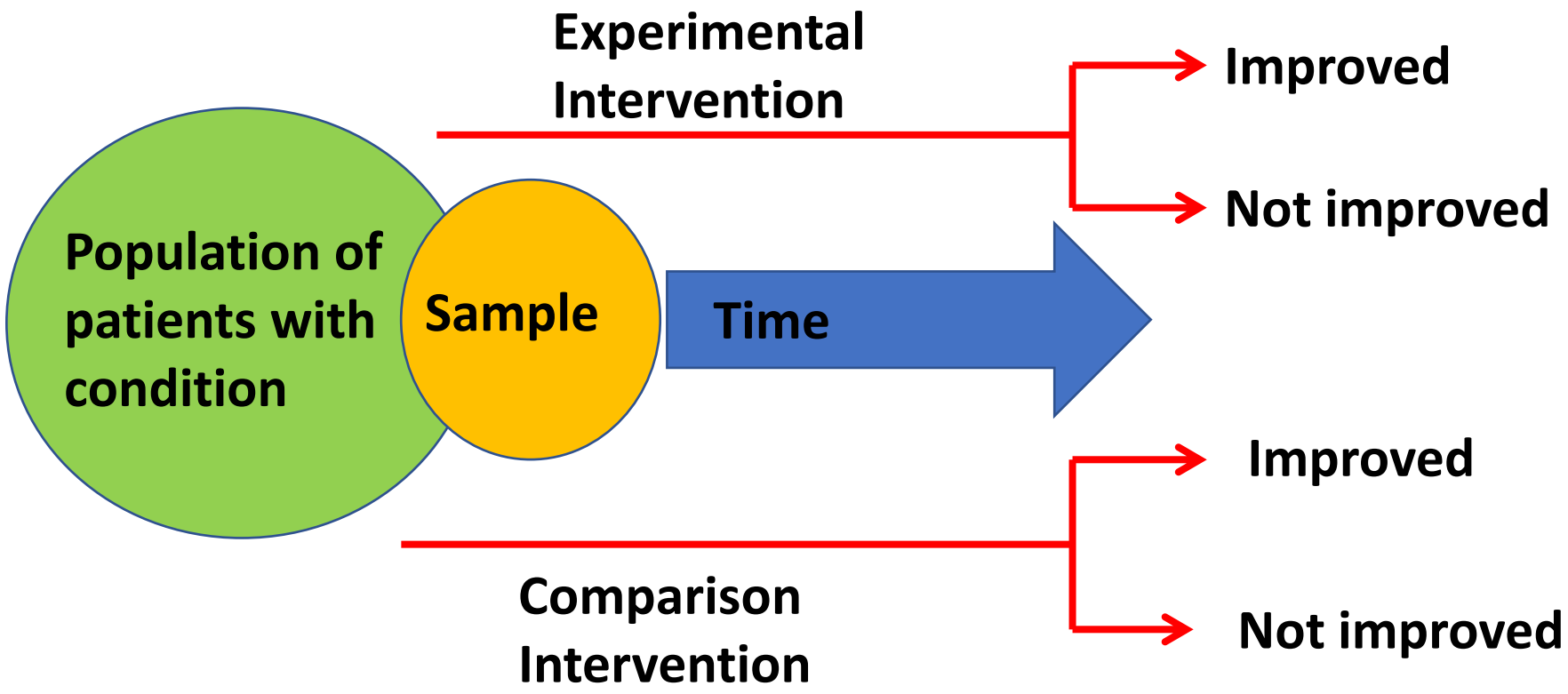
OBJECTIVES

- Evaluate new forms of therapy and prevention e.g.
  - New drugs
  - Medical health care technology
  - Methods of primary prevention
  - Screening programmes
  - Ways of organizing & delivering health services
  - Impact of new policies in health care and health care financing

TYPES OF RCTs

- Historical
  - Comparison with patients who were treated before the new drug became available
    - Information bias due to data quality
    - Differences due to support etc
- Nonrandomised
  - Concurrent study by nonrandom assignment of intervention
    - Assignment bias
- Randomised
  - Concurrent study by random assignment of intervention

STRUCTURE OF AN RCT



Fletcher et al. 1996

CHARACTERISTICS OF A RCT

- Hypothesis
    - Study design is based on prior data and biological mechanism
    - State the question that needs to be answered
- Population Selection
    - Be able to generalize
    - Provide a good cases definition. Who will be treated?
    - Inclusion and exclusion criteria
    - Recruitment & screening

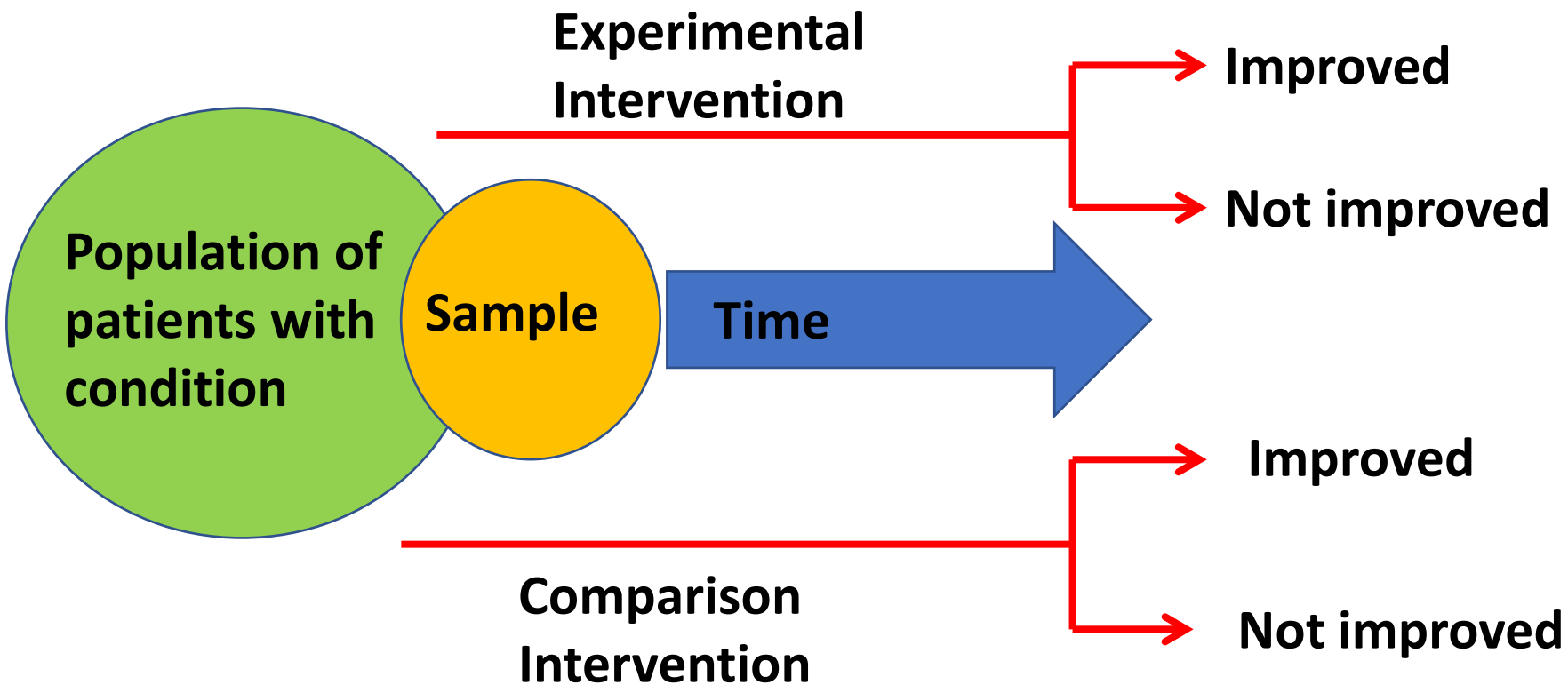
CHARACTERISTICS OF A RCT

- Treatment Protocol
    - Assignment of randomization
    - How will the intervention be applied
    - The process of standardization
    - Background treatment, supportive care and differential treatment

CHARACTERISTICS OF A RCT

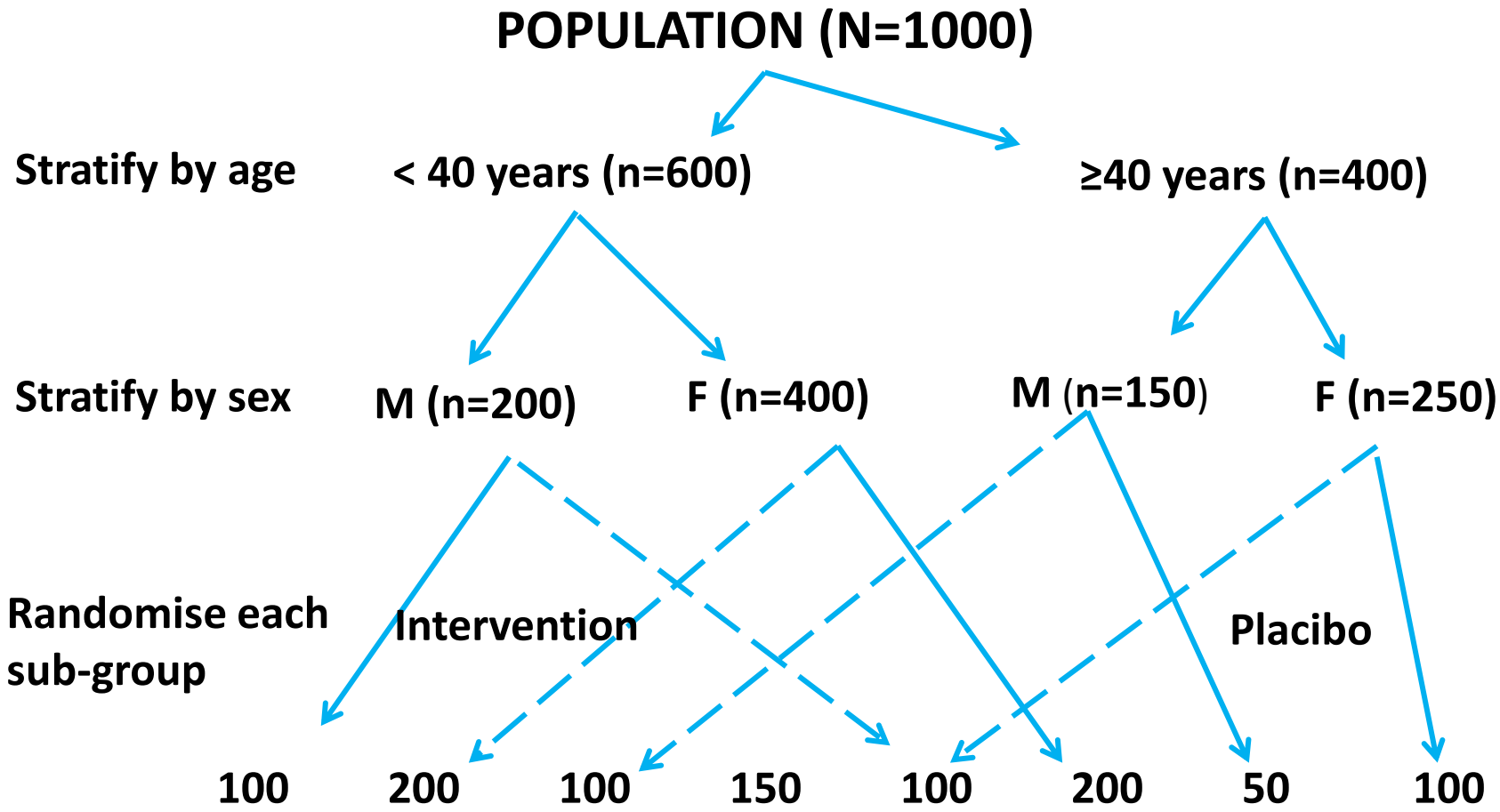
- Study Design
  - Randomised vs nonrandomised
  - Method of randomisation
  - Simple non-crossover randomisation

STRUCTURE OF AN RCT

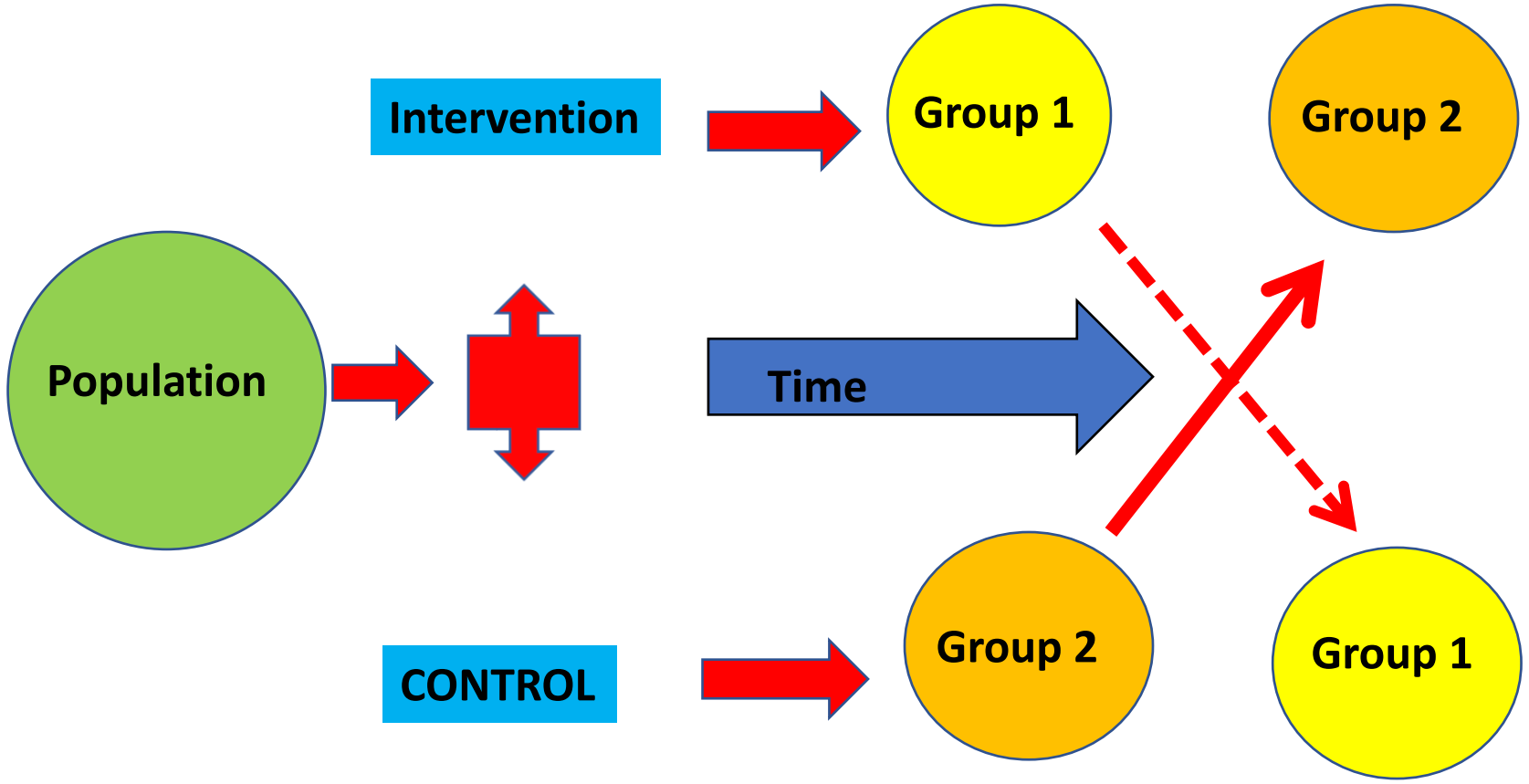


Fletcher et al. 1996

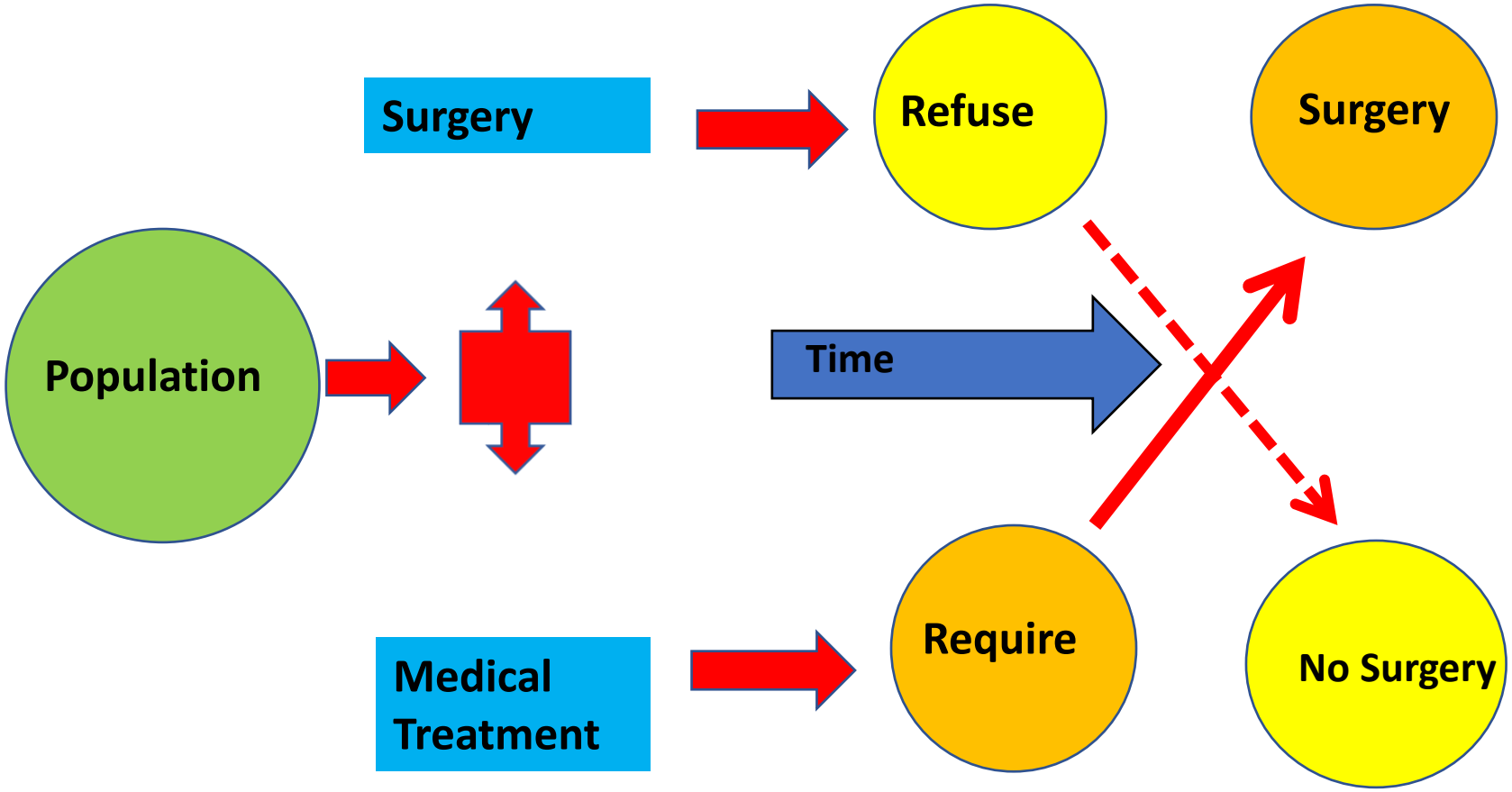
CHARACTERISTICS OF A RCT:
Stratified randomisation



CHARACTERISTICS OF A RCT: Planned crossover



CHARACTERISTICS OF A RCT: Unplanned crossover



CHARACTERISTICS OF A RCT

- Blinding
  - Single
  - Double
  - Triple
  - Placebo effect
  - Don’t blind in the ffg cases
    - Surgical intervention
    - Behavioral intervention
    - Health care intervention
-

CHARACTERISTICS OF A RCT

- Compliance
  - Characteristics of compliance
    - Adherence to study protocol
    - Reasons for noncompliance
      - Side effects
      - Difficulty of the intervention
      - Forgetting medication
      - Withdrawal
      - Crossover
      - Moving
  - Strategies to achieve compliance
    - Select motivated subjects
    - Pretest willingness to participate (bias)
    - Simple and clear instructions
    - Incentives to comply
    - Positive reinforcements
    - Communication and frequent contact
    - Pill counts
    - Limit duration of the study

CHARACTERISTICS OF A RCT

- Analysis
  - Type 1 & type 2 errors
    - Study results differ or are the same
  - Comparison of baselines
  - Comparison of outcome variables
  - Intention to treat vs on-treatment
  - Adjustment of differences in the baseline predictor variables
  - Sub-group analysis
  - Criteria for premature termination of the trial

CHARACTERISTICS OF A RCT

- Endpoints
  - Primary outcome
    - Defined in advance
    - Be able to assess in all patients
    - Participation ends when trial ends
    - Unbiased assessment
    - Complete ascertainment
  - Intervening events
  - Adverse event
  - Surrogate outcome

CHARACTERISTICS OF A RCT

- **Advantages of RCT'S**
  - Randomization – balance confounding factors
  - Uniform collection of data
  - Protocol procedures & interventions specified
  - Blinding
  - Statistical comparison based on statistical theory
- **Disadvantages of RCT'S**
  - Lack of generalizability
  - Long time for conclusion
  - Large numbers needed for study
  - High financial costs
  - Ethics
  - Compliance maybe poor
-

This training programme is supported by the Fogarty International Center of the National Institutes of Health under Award Number U2RTW012140. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health

WASHA TAKWIMU

WORKING ON APPLICATIONS FOR DATA SCIENCE AND HEALTH IN AFRICA