HARVARD T.H. CHAN SCHOOL OF PUBLIC HEALTH

HEIDELBERG INSTITUTE OF GLOBAL HEALTH

UNIVERSITY OF KWAZULU-NATAL ™
INYUVESI YAKWAZULU-NATALI

MUHAS

MAKERERE UNIVERSITY

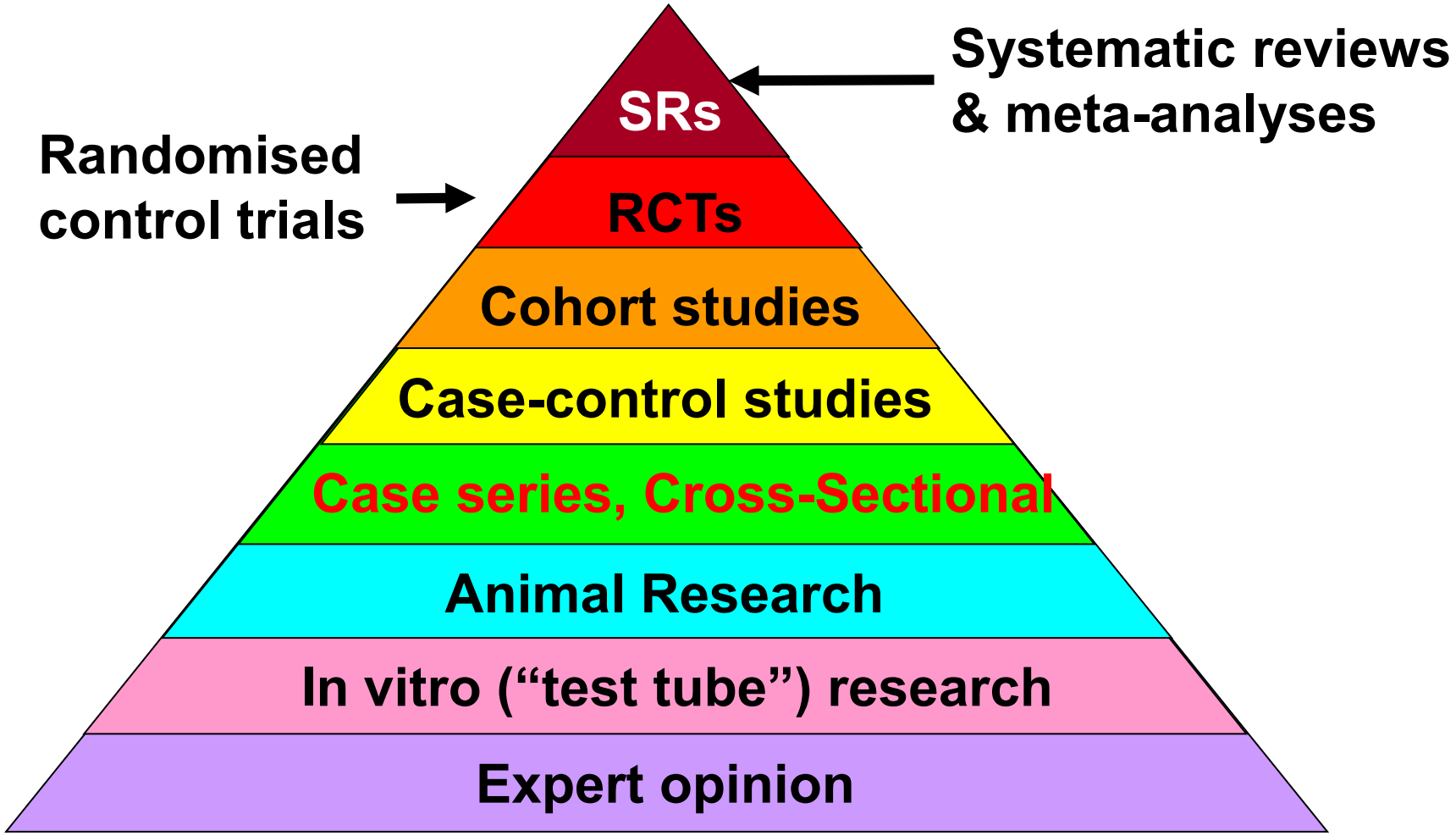
UNIVERSITY OF GHANA

WASHA Takwimu

Introduction to Study Design

Saloshni Naidoo

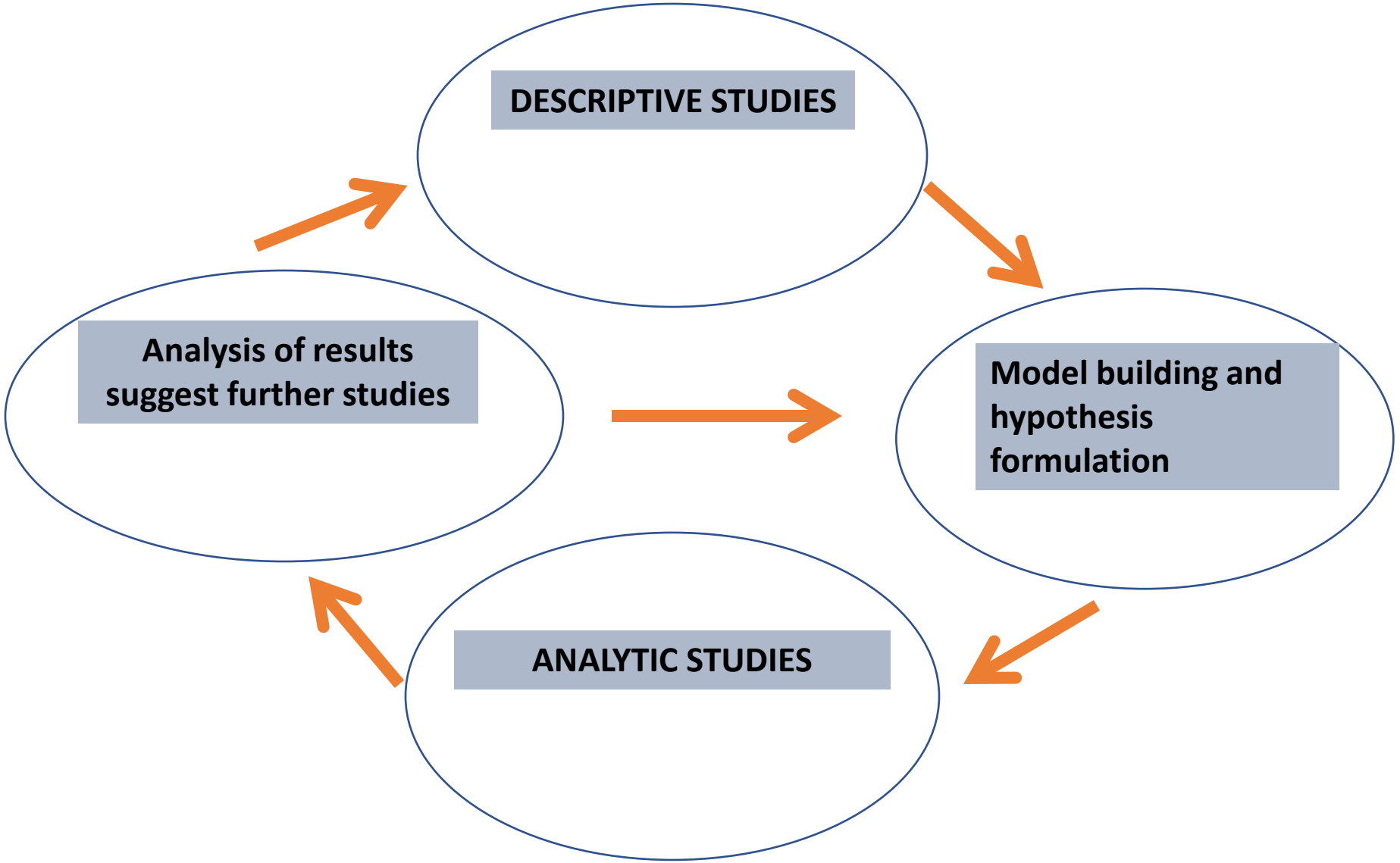
The evidence pyramid



Study Type	Observation Unit	Measure of Occurrence	Measure of Association
Observational Studies			
Descriptive			
Cross-sectional	Individuals	Prevalence	None
Cohort	Individuals	Incidence (risk or rate)	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
?	?	+

OBSERVATIONAL STUDIES

Case – control Study

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

Cohort Study

exposure	disease	person
*	**?**	+

Cross sectional Study

exposure	disease	person
?	**?**	+

UNIVERSITY OF KWAZULU-NATAL ™

INYUVESI YAKWAZULU-NATALI

Case-Control STUDIES

2020

SALOSHNI NAIDOO

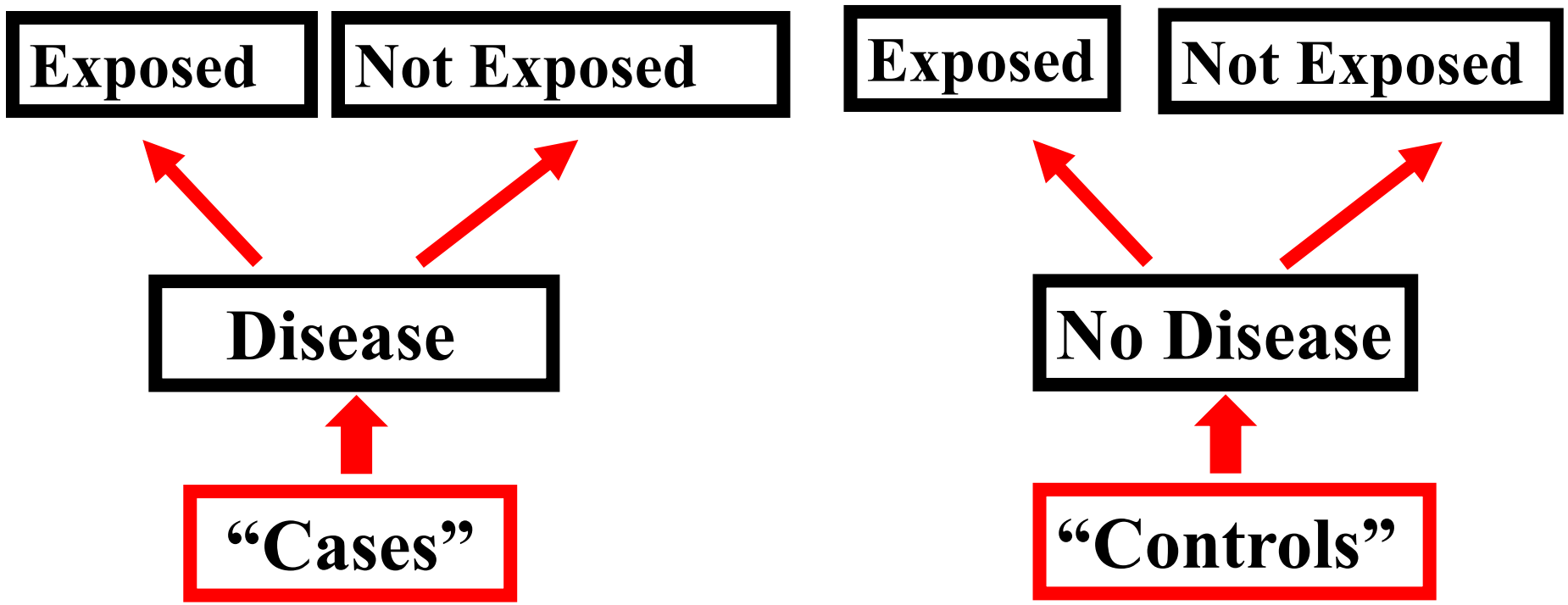
naidoos71@ukzn.ac.za

Acknowledging Dr S Knight

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

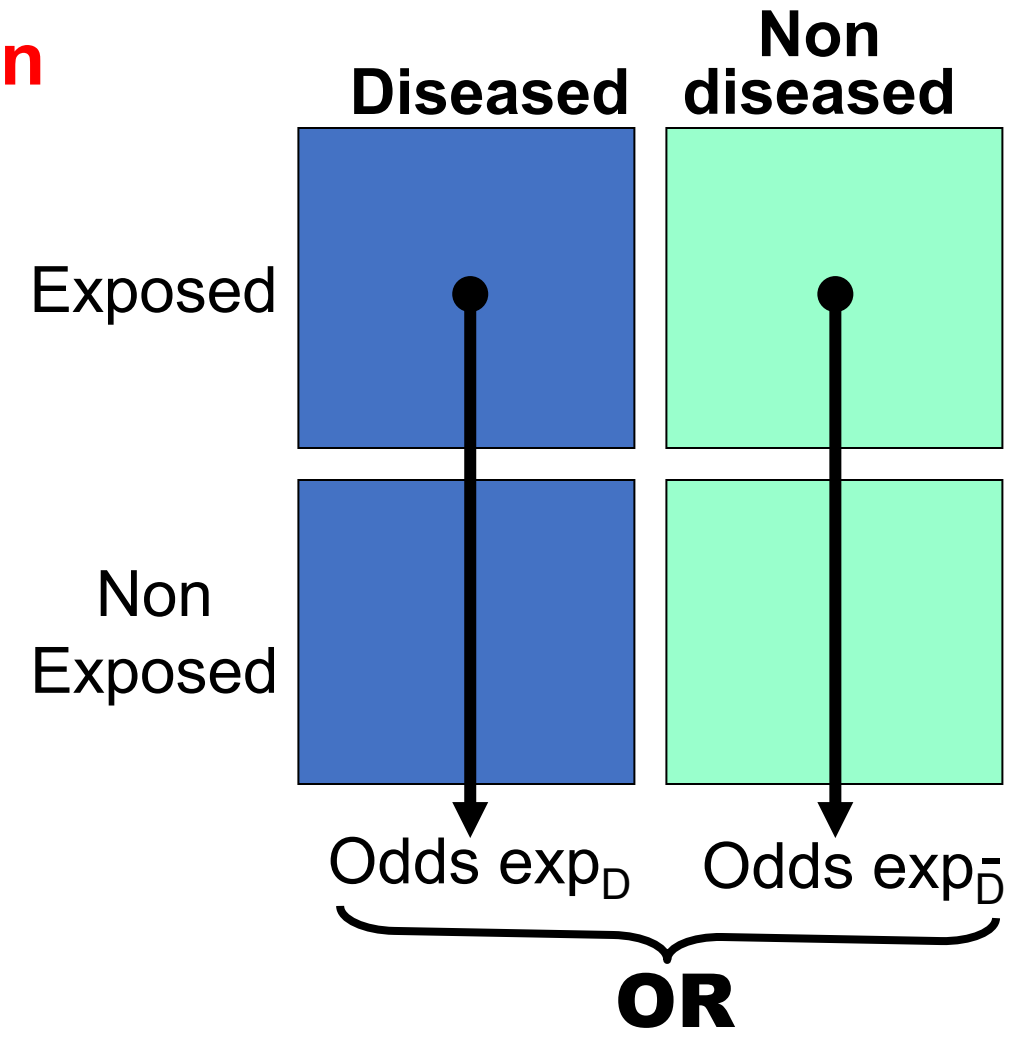
CASE CONTROL STUDY



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:

  ad / bc= 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**.
- RR = Risk in non-exposed/Risk in exposed

$$= \frac{a/(a+b)}{c/(c+d)}$$

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
    - 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

References

- Keyes, K. M., & Galea, S. (2014). *Epidemiology matters: a new introduction to methodological foundations*. Oxford University Press.
- Katzenellenbogen, J., Joubert, G., & Karim, S. A. (1997). *Epidemiology: a manual for South Africa*. Oxford University Press Southern Africa.

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