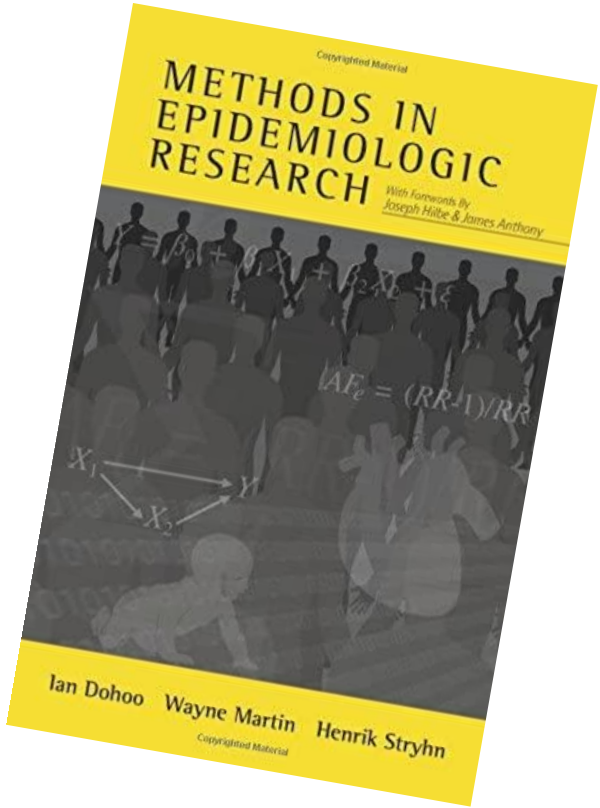
Advanced Epidemiology

(PH6201)



Dr. Ibrahim Elsohaby
(DVM, MVSc, GradCert One Health, PhD)

Assistant Professor of Public Health and Epidemiology
Department of Infectious Diseases and Public Health
City University of Hong Kong

Course Information and Outline

Teaching Team

- Dr. Ibrahim Elsohaby (Course coordinator)
  Email: ielsohab@cityu.edu.hk
  Office: 1B-404, Block 1, To Yuen Building
  Office Hours: Anytime – Afternoon is ideal!

- Dr. Casper Zhang (Scientific officer)
  Email: Casper.Zhang@cityu.edu.hk
  Office: 4/F TYB-1

Course Information and Outline

Who are you? 😊

- Name
- Where you come from
- Educational background
- Program (MSc, MVSc, PhD)
- Project details
- Biostatistics / Epi interests,
- Hopes for the course

PH6201	PH8002	PH6204	PH6205
Advanced Epidemiology	Infectious Disease Epidemiology	Public Health Surveillance	Intermediate Statistics for One Health

1st **Introduction to Biostatistics in One Health**

2nd **Principles of One Health and Epidemiology**

3rd **Applied Public Health Projects**

4th **Time Series analysis**

5th

- Statistical Data Analysis
- Spatial data analysis
- Statistical Modelling for data mining

CityU Jockey Club College of Veterinary Medicine and Life Sciences
香港城市大學
City University of Hong Kong
in collaboration with Cornell University

Course Information and Outline

Learning Outcomes

At the end of the course, you will be able to:

- Distinguish between experimental and observational studies.
- Describe the key characteristics of experimental, cohort, case–control, cross-sectional, and hybrid studies regarding subject selection, data collection, and analysis.
- Identify the design of a particular study.
- Discuss the factors that determine when a particular design is indicated.
- Understand the principles of environmental, occupational, molecular, social epidemiology and outbreak investigation.

Course Information and Outline

Teaching and Learning Activities

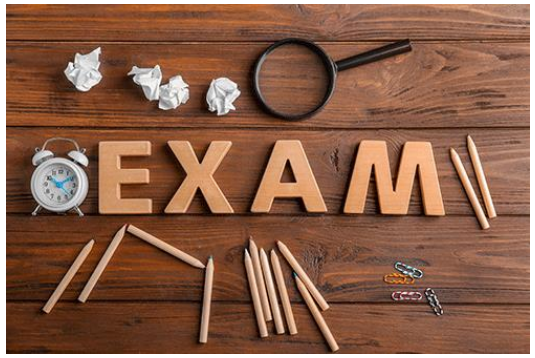
- **Lectures**
  - 13 sessions (including illustrative Examples and exercise)

- **Hands-on practical exercises**
  - 4 sessions
  - Hands-on, problem-based activities (review and critical appraise of papers)
  - Students will be divided into groups.

Course Information and Outline

Assessment Tasks/Activities

- **Classroom assessment**
  - 10%
  - Student's class participation
- **Midterm examination**
  - **February 28**
  - 40%
  - Topics covered in **Week-1 to Week-6**
- **Final examination**
  - 50%
  - Topics covered in **Week-7 to Week-13**



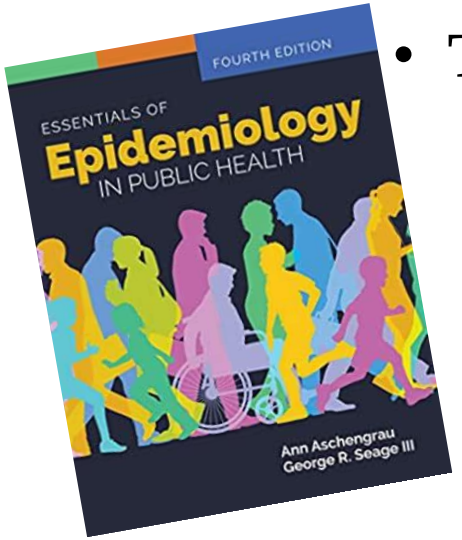
Course Information and Outline

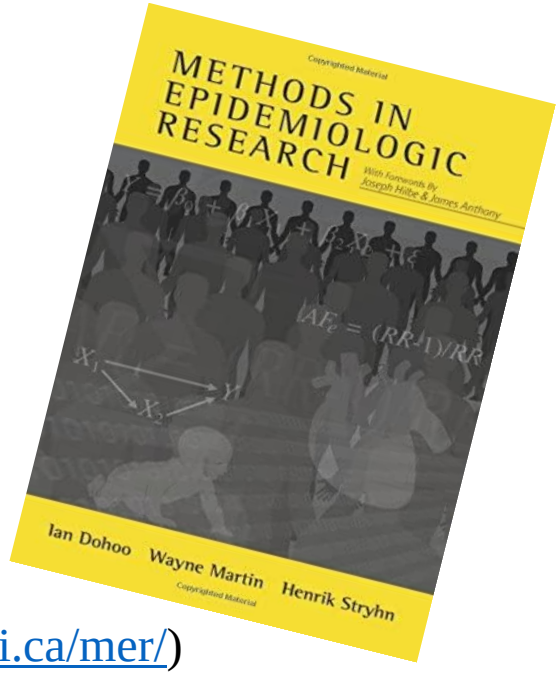
Schedule – Syllabus

Date	Week	Time / Location	Topics	Duratio	Building / Room	Instructor
Jan 10	Week1	7:00 – 10:00 PM	Introduction A comprehensive review of the principles of epidemiology	3 hrs	YEUNG / **Y4702**	Ibrahim
Jan 17	Week 2	7:00 – 10:00 PM	Descriptive studies Cross-sectional studies	3 hrs	**Y4702**	Ibrahim
Jan 31	Week 3	7:00 – 10:00 PM	Cohort studies	3 hrs	**Y4702**	Ibrahim
Feb 7	Week 4	7:00 – 10:00 PM	Case-control studies	3 hrs	**Y4702**	Ibrahim
Feb 14	Week 5	7:00 – 10:00 PM	Interventional studies / Hybrid studies	3 hrs	**Y4702**	Ibrahim
Feb 21	Week 6	7:00 – 10:00 PM	Systematic reviews and metanalysis	3 hrs	**Y4702**	Ibrahim
Feb 28	Week 7	7:00 – 10:00 PM	Midterm Exam A practical guide to epidemiological data analysis	1 hrs 2 hrs	**Y4702**	Ibrahim Omid
March 7	Week 8	7:00 – 10:00 PM	Environmental and occupational epidemiology	3 hrs	**Y4702**	Ming
March 14	Week 9	7:00 – 10:00 PM	Prevention for non-communicable disease	3 hrs	**Y4702**	Ming
March 21	Week 10	7:00 – 10:00 PM	Molecular epidemiology	3 hrs	**Online**	Ioannis
March 28	Week 11	7:00 – 10:00 PM	Social epidemiology	3 hrs	**Online**	Ioannis
April 4	Week 12	7:00 – 10:00 PM	Outbreak Investigation	3 hrs	**Online**	Ioannis
April 11	Week 13	7:00 – 10:00 PM	Course warp-up	3 hrs	**Y4702**	Ibrahim
		7:00 – 9:00 PM	Final Exam	2 hrs		

Course Information and Outline

Textbook – Notes and datasets





- **Textbook**
  - Methods in Epidemiologic Research (https://projects.upei.ca/mer/)
  - Veterinary Epidemiologic Research (https://projects.upei.ca/ver/)
  - Essentials of Epidemiology in Public Health (https://www.jblearning.com/catalog/productdetails/9781284128352)

- **Notes and datasets**
  - CityU canvas (Link: https://canvas.cityu.edu.hk/courses/52711)

Week 1

- **In Week 1:**
  - ➢ Basic research process
    - What is research?
    - Objectives & questions of interest
    - Hypotheses
    - Sections of a scientific report
    - Systematic search (reminder)
  - ➢ Basic “methodology”
    - Sampling
    - Study types
    - Causality?
    - Variable types
    - A general guide to statistical tests of hypothesis

Learning Objectives:

Review the epidemiological principles including:

1. Describe and differentiate different types of sampling, study types, variables and bias.

I. Research process

- **What is “research”?**
  - Detailed study of a subject, especially in order to discover (new) information or reach a (new) understanding (Cambridge Dictionary)
- **Two main goals**
  - States of the nature
  - Relationships between variables
- **Logical steps**
  1. Identify the problem
  2. Formulate a hypothesis
  3. Develop the research plan/protocol
  4. Collect & analyse the data
  5. Interpret the results, make conclusions, and report
- **Types (from different perspectives)**
  - https://www.indeed.com/career-advice/career-development/types-of-research-methods

Types of research methods

- Observation
- Questionnaires and surveys
- Interviews
- Focus groups and case studies
- Experiments
- Secondary data analysis
- Mixed methods

Recommended sources:

✓ https://www.indeed.com/career-advice/career-development/types-of-research-methods

✓ https://www.youtube.com/watch?app=desktop&v=V8ndAyxkxtA

Formulate clear objectives, questions, hypotheses!

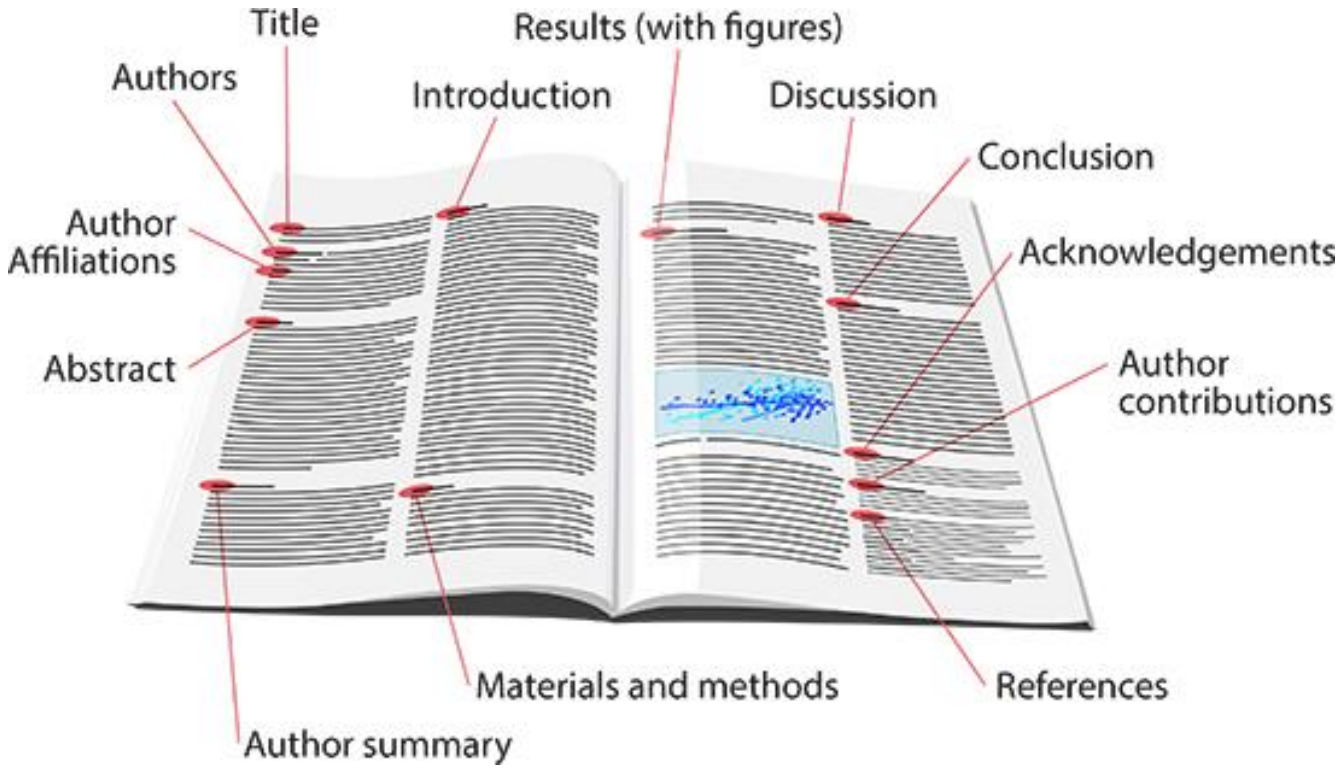
- **Objective(s)**
  - In general, research objectives describe what we expect to achieve by a project
  - https://www.indeed.com/career-advice/career-development/research-objectives

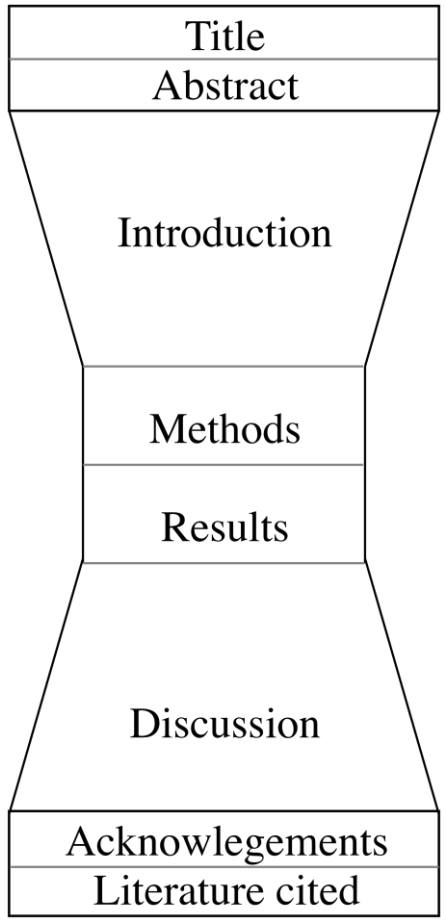
- **Question(s)**
  - A question that a study or research project aims to answer
  - Your topics – revisit ‘Week 14: EBM and critical appraisal’
  - https://research.com/research/how-to-write-a-research-question

- **Hypothesis**
  - A hypothesis (plural: hypotheses) is a testable statement about the relationship between two or more variables or a proposed explanation for some observed phenomenon
  - Corresponding to each question of interest (expected outcome)
  - https://www.enago.com/academy/how-to-develop-a-good-research-hypothesis/

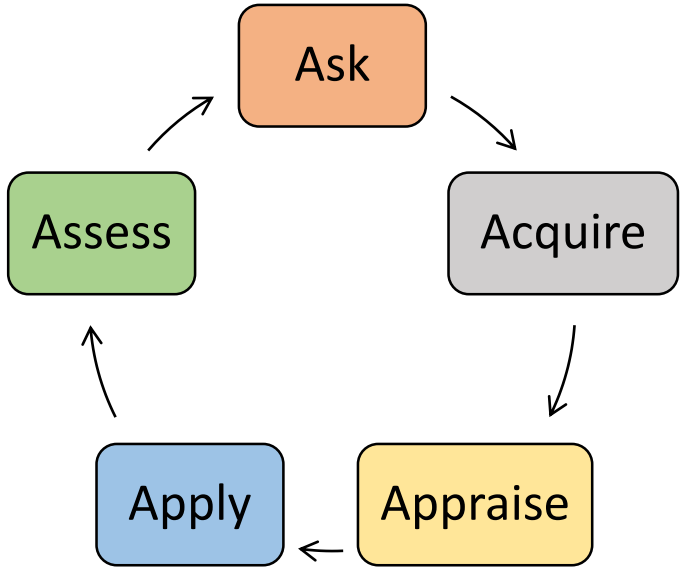
Structure of a scientific report/journal article

- Must follow the exact guidelines for authors provided by any journal or for thesis preparation
- In general





Reminder of EBM process (5 As)



Ask

✓ Defining a clinical question that is of interest and answerable

Acquire

✓ Finding the best available evidence to answer the question

Appraise

✓ Assessing the quality of the relevant evidence found

Apply

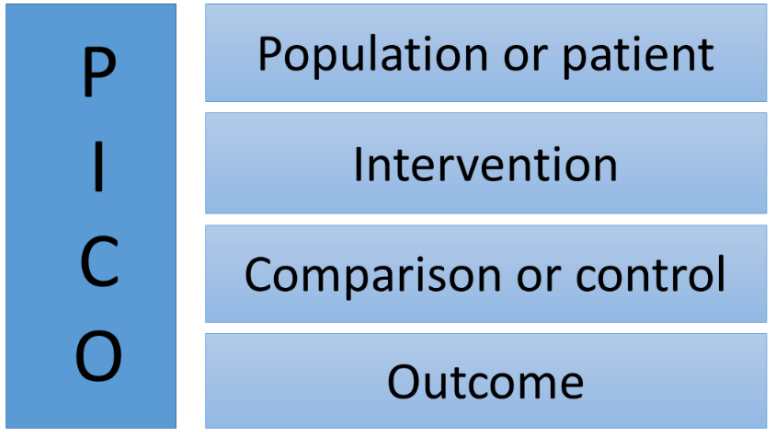
✓ Implementing the evidence into clinical practice where appropriate

Assess

✓ Evaluating the impact of the implementation and changes in clinical practice

1. Formulating an answerable question (ASK)

- The first fundamental skill required in EBM is asking ‘well-formulated’ clinical questions.
- Refine the clinical questions so the best evidence can be found (guide your search).
- Ask “**PICO**”
  - Patient or problem
    - Define the patient’s age, breed, problem, population
  - Intervention (/exposure)
    - Diagnostic or therapeutic intervention or exposure
  - Comparison/control
    - What is the main alternative (if applicable)?
  - Outcome
    - What do you hope to accomplish? Appropriate time frame? Cost-benefit analysis?

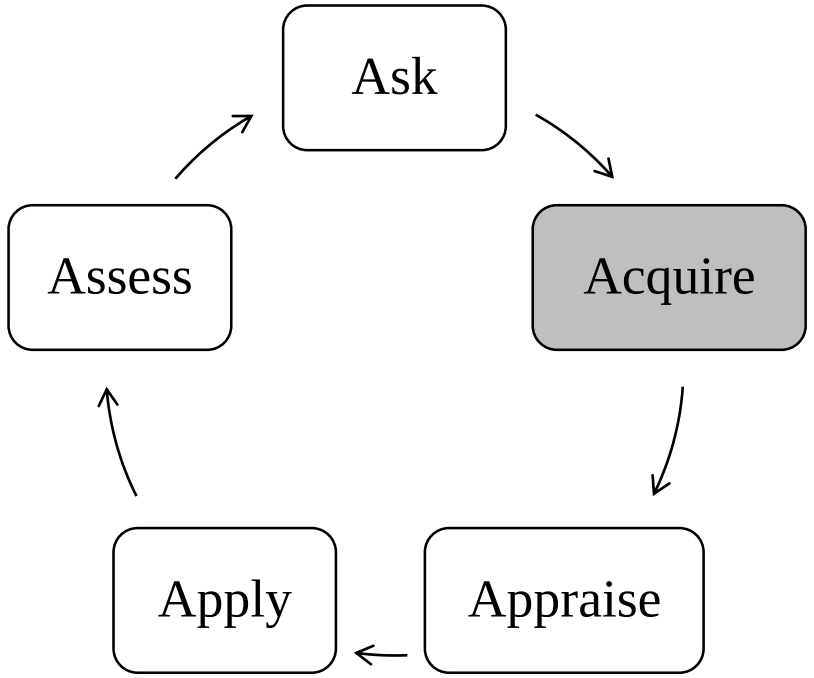


- **Example:** In pregnant women, does smoking (versus non-smoking) increase the risk of low birth weight?

2. Finding the best available evidence to answer the PICO (Acquire)

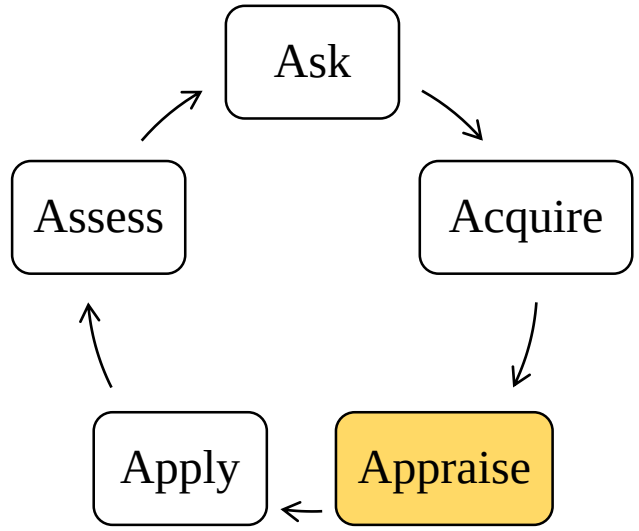
- Search strategy/string
    1. Identify key words & synonyms
    2. Truncation
    3. Combining key words

- Relevant resources/databases
    - Library sessions?



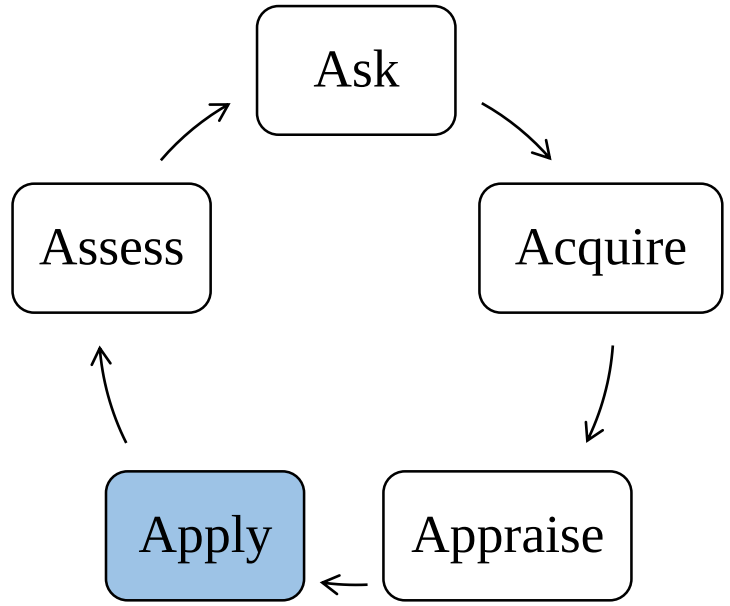
3. Assessing the quality of the relevant evidence found (APPRAISE)

- To evaluate the quality, validity & applicability of the literature found:
  - ✓ Is the paper relevant to my clinical question, my population or my patient?
  - ✓ Which level of evidence does the paper provide?
  - ✓ Is the quality of the paper "good enough" to help me answer my question?
  - ✓ Does the paper have the right study design to answer my clinical question?
  - ✓ The type and level of potential biases in the study?
- As the reader, it is up to you! Everything you read is not true!
- We have equipped you with a basic toolbox for "critical appraisal" so far



4. Apply

- Now, you need to determine whether the answers you have generated can be applied to your “circumstances”:
    - Country
    - Location (clinic)
    - Case in front of you
    - Availability of therapies
    - Owner
    - Cost
    - etc.

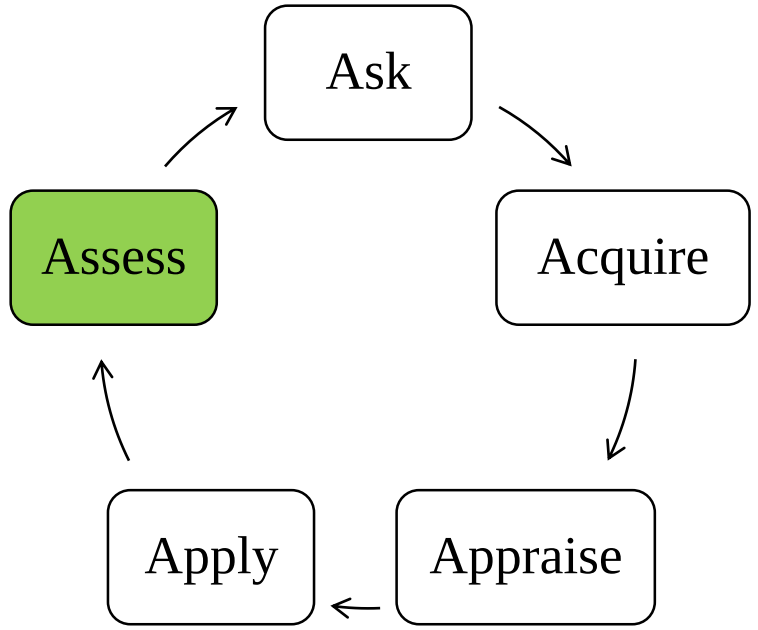


- The application of evidence into practice can sometimes be challenging!

5. Assess

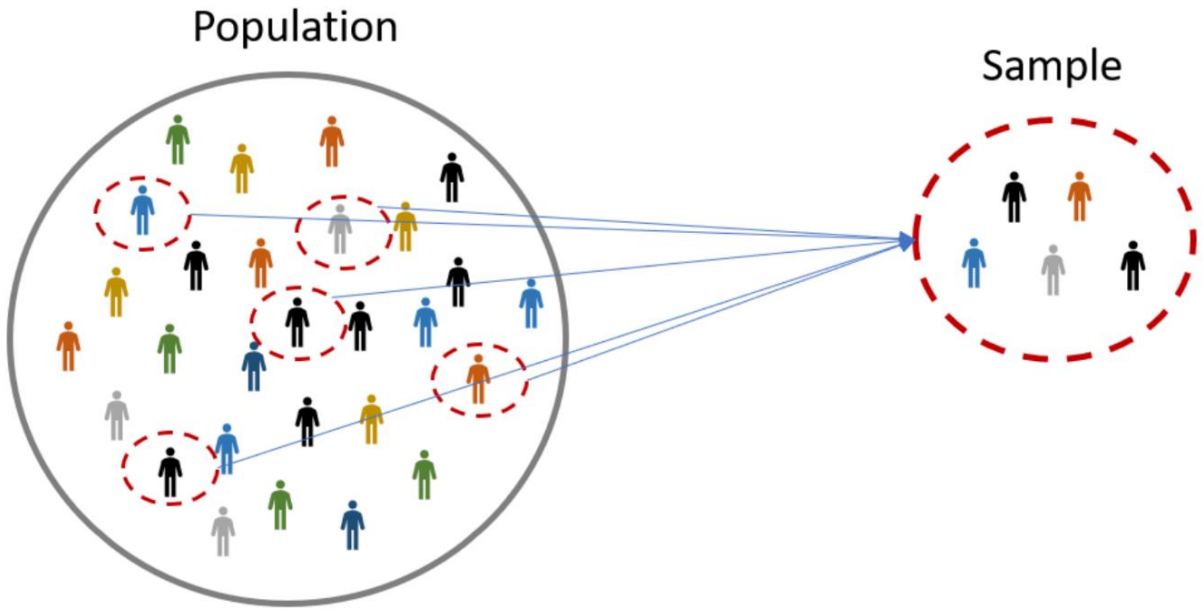
- Using your skills to establish if the EBM process and the evidence have made a difference to your practice, patients health, improved care, etc.

- EBM starts in practice (Ask) & ‘Assess stage’ ensures EBM stays in practice!



II. Basic “methodology”

- **Sampling populations**

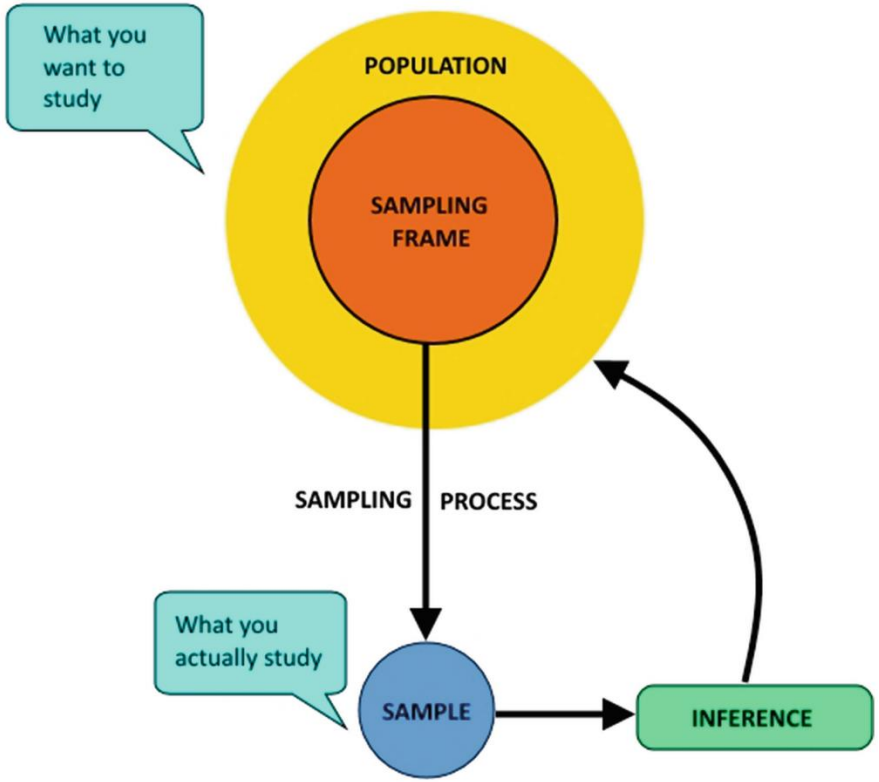


Why sampling?

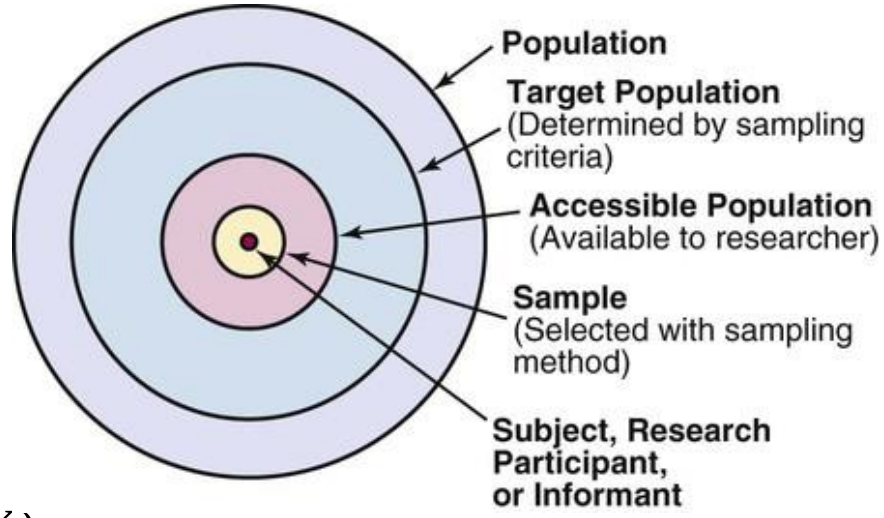
- Clinical/epidemiological research require that we gather information about populations
  - Disease prevalence
  - Detection of a disease
  - Freedom from a disease
  - Explore associations & causality

- To collect data/information: ‘census’ or ‘sampling’
  - Practical/possible
  - Logistics (cost, time, labour, etc.)

- To estimate a population parameter
  - Representative samples
  - Inferences about population
  - Attributes or relationships



Sampling terminology

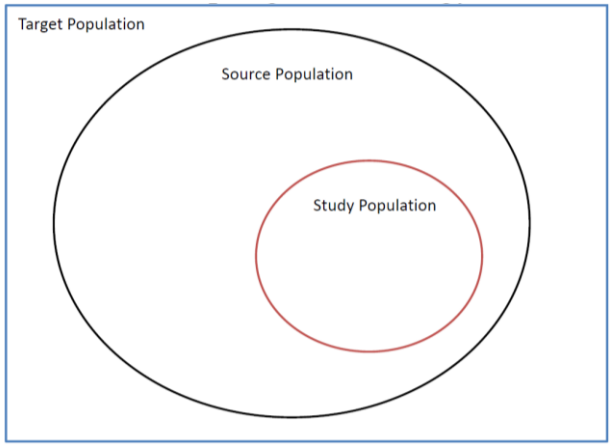


- **Population -** collection of "units" that have a common characteristic (e.g., pet cats / patients)

- **Study objective:** (e.g., the prevalence of FIP in pet cats in HK?)

- **Target population -** to make inferences about (e.g., pet cats in HK).

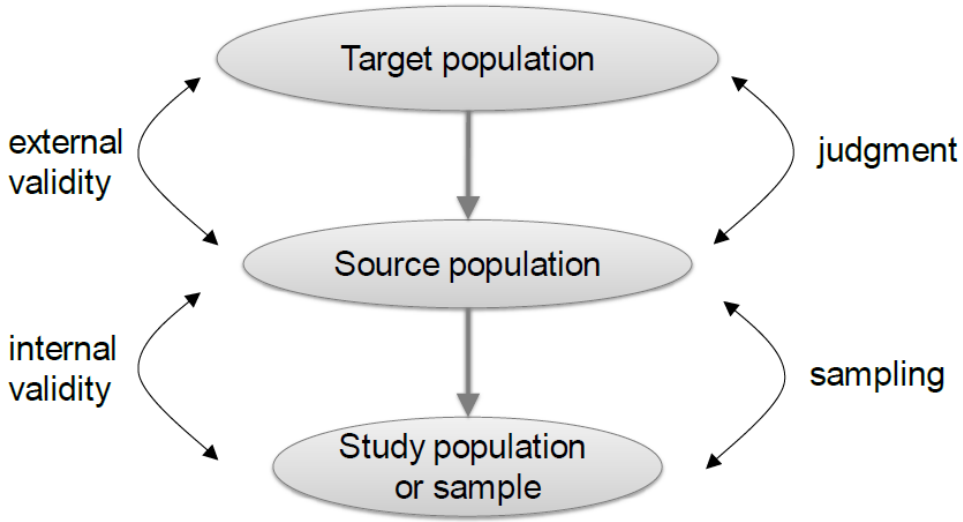
- **Source population -** accessible subset of the target population from which samples are drawn (e.g., pet cats referred to specific vet clinics in HK)
  - **Sampling frame -** lists all sampling units in the source population
  - **Sampling unit -** unit of interest to be sampled (e.g., a pet cat)

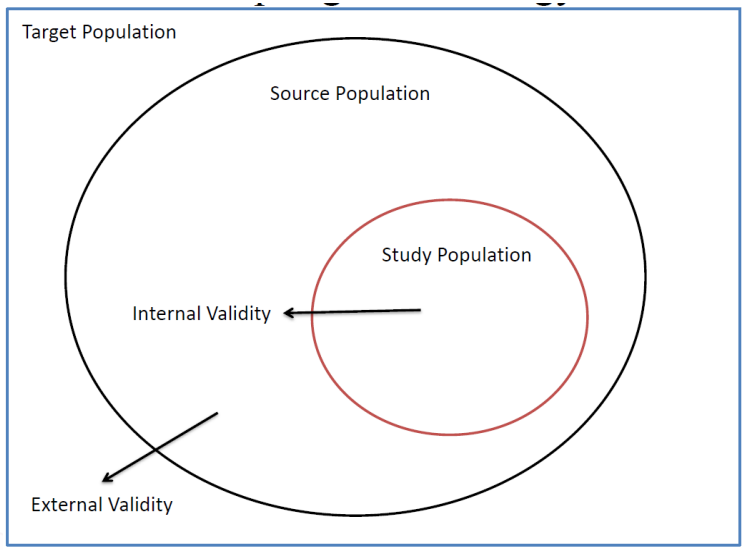


- **Study population/group -** individuals that end up in the study
- (e.g., selected pet cats for the study/test for FIP)
- **Study unit (unit of concern)** - could be = sampling unit or not (e.g., herds)
- **Sampling fraction -** study population size/source population size (n/N)

Internal & external validity

- **Internal Validity**
  - Whether or not the study results (obtained from the study group) are valid for members of the source population
- **External Validity**
  - How well the results can be 'generalized' to the target population
  - Assessment of whether or not the source population is broadly representative of the target population

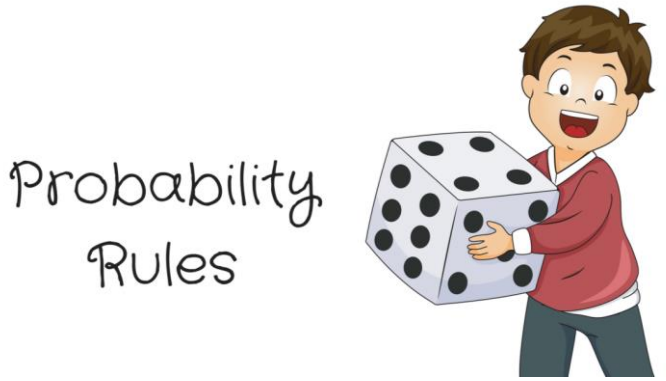




Sampling process

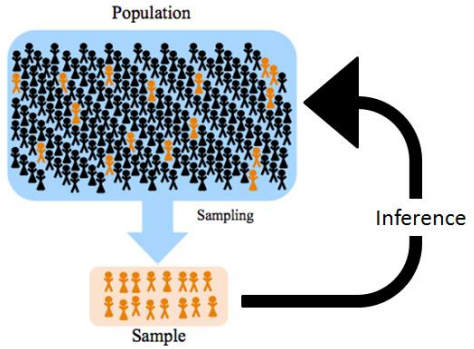
- Define & identify target/source population
- Define & identify study population
    - Ideally representative of the target population
- Produce the sampling frame
    - Identify every unit of interest (sampling unit)
- Specify the appropriate sampling strategy
- Estimate the number of sampling units required
- Select sampling units

Sampling methods



- **Aim of sampling process:**
  - To draw a sample which is a true representation of the source & ideally the target population.
  - Leading to estimates of population characteristics of an acceptable accuracy.

- Two general approach:
  - **Probability (random)**
    - Every element/unit in the source population has a known, non-zero probability of being included in the sample.
  - **Non-probability**
    - No explicit method for determining an element's probability of selection
    - 'representativeness' cannot be quantified!



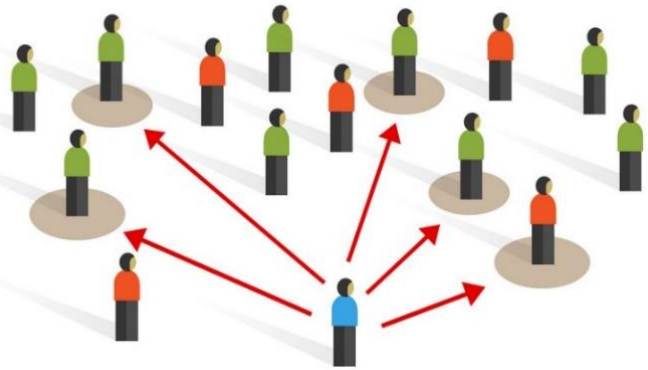
Main sampling strategies

- Probability (Random)
  - Simple
  - Systematic
  - Stratified
  - Cluster
  - Multistage
  - Others…
- Non-probability
  - Convenient
  - Purposive/targeted
  - Judgment

Non-probability sampling

- **Convenient**
  - Taking the most easily obtainable observations.

- **Purposive**
  - Targets specific risk groups (e.g., sick people).



- **Judgement**
  - Selecting what is regarded to be a ‘representative’ sample.
  - Depend on the judgement of the investigator or researcher.

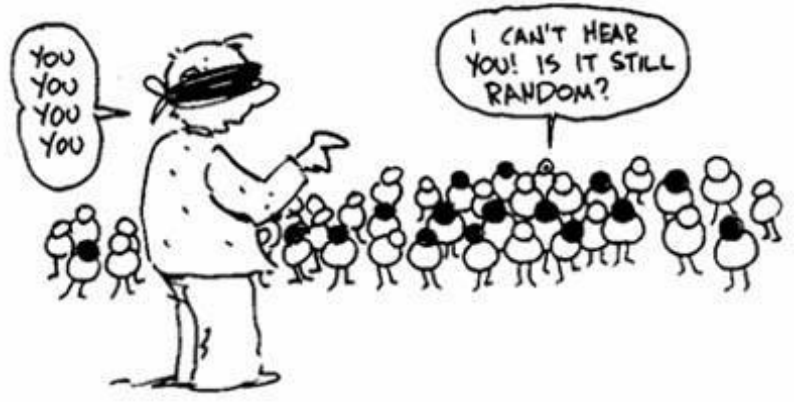
- **Disadvantages**
  - Not possible to know the probability of each “unit” being selected.
  - Cannot estimate the sample size or standard errors.
  - Biased population estimates; no way to quantify the extent of bias.

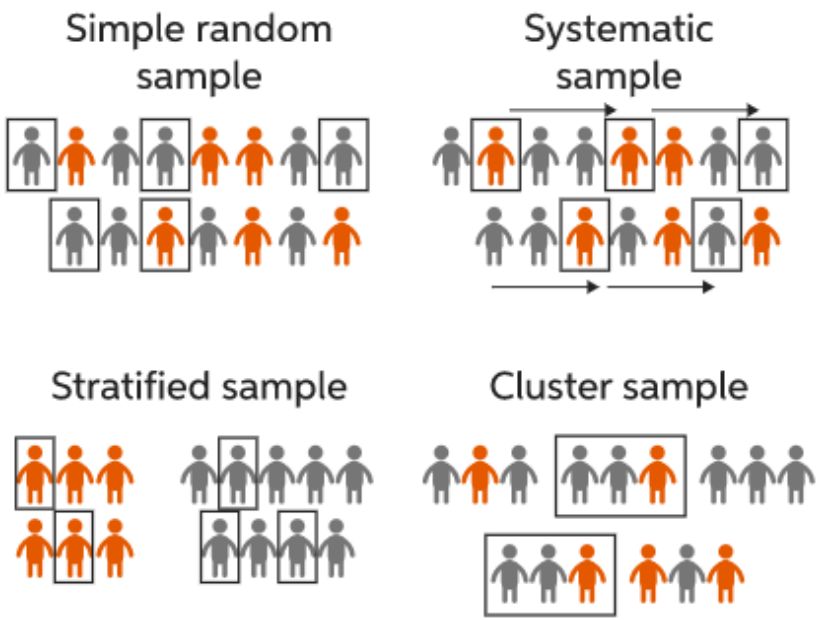
Probability sampling strategies

- Every element in the population has a known, non-zero probability of being included in the sample
  - Equal chance of being sampled (or not).

- Selection is based on a formal random procedure
  - To ensure representativeness of the sample

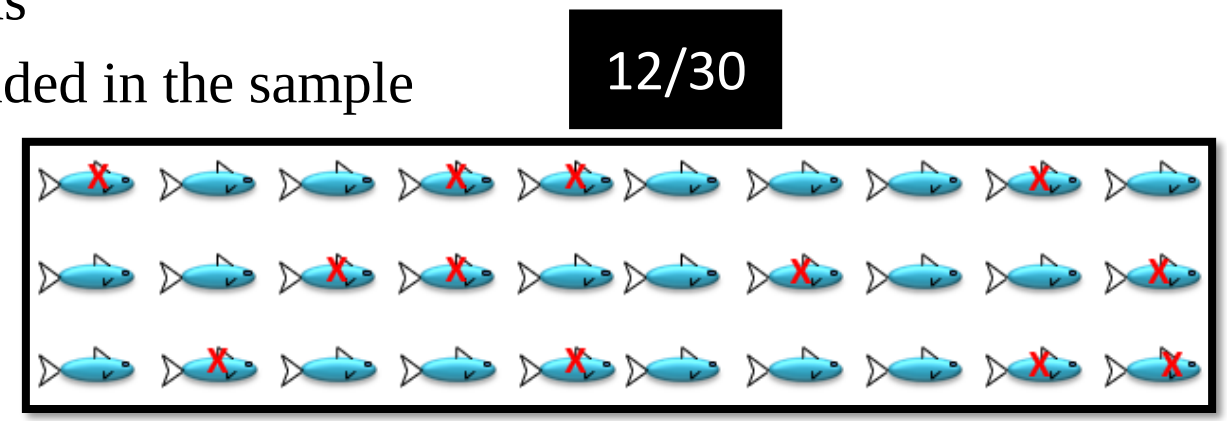
- Applicable to 'sampling unit'
  - ***individual-level*:** simple random, systematic, or stratified sampling
  - ***Group-level*:** cluster, or multistage sampling





Simple random sampling (SRS)

- SRS often forms the basis of sampling designs
- Every unit has an equal chance of being included in the sample
- Requires a list of all units/subjects
- **Procedure**
  - Generate list of all animals
  - Assign unique number to each animal
  - Use random number generator to generate a list of numbers
  - Match generated numbers with animal numbers



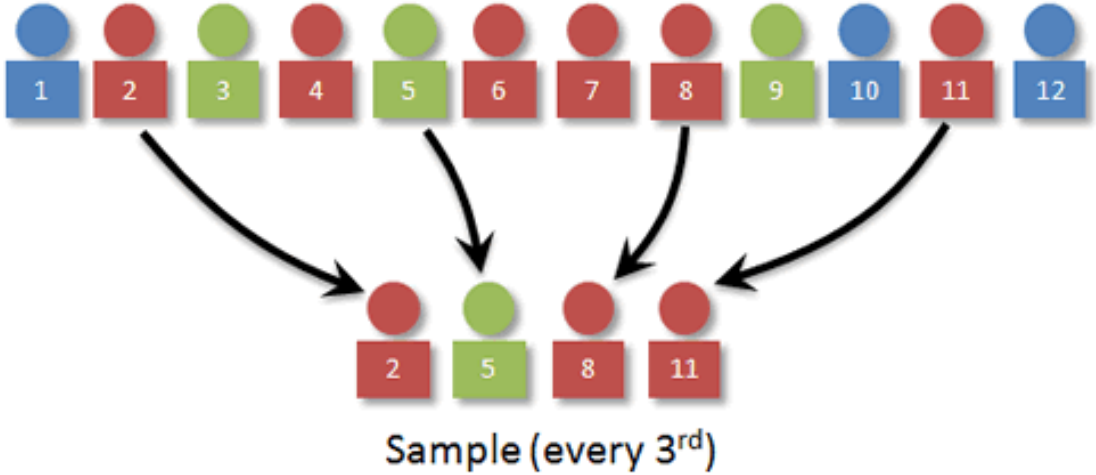
- **Limitations**
  - Often difficult to obtain a complete sampling frame (large samples).
  - Costs may be higher than other methods.
  - Heterogeneous populations may give results with poor accuracy.

Systematic random sampling

- Does not require a sampling frame, but need all individuals/animals in a sequence.

- **Procedure**
  - Calculate ‘sampling interval’ (**j**) = number of individuals in the frame/number of samples required
  - Randomly pick a number between 1 to j, sample every **+j** individuals …

- e.g., 4 samples from 12 individuals (j = 12/4 = 3)
- Production line

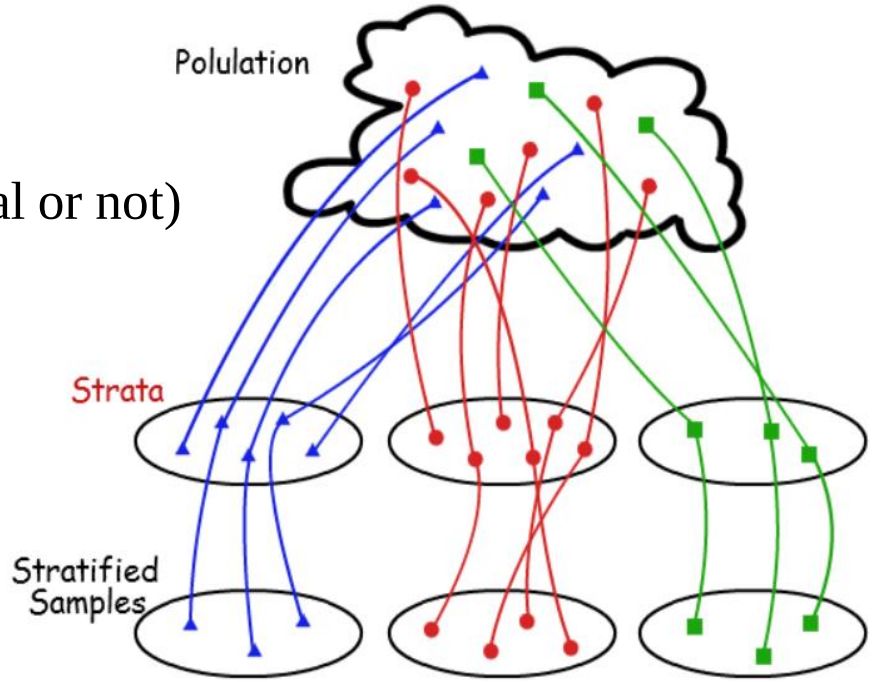


- **Limitations**
  - Subject to bias in the intervals
  - May not account for heterogeneity

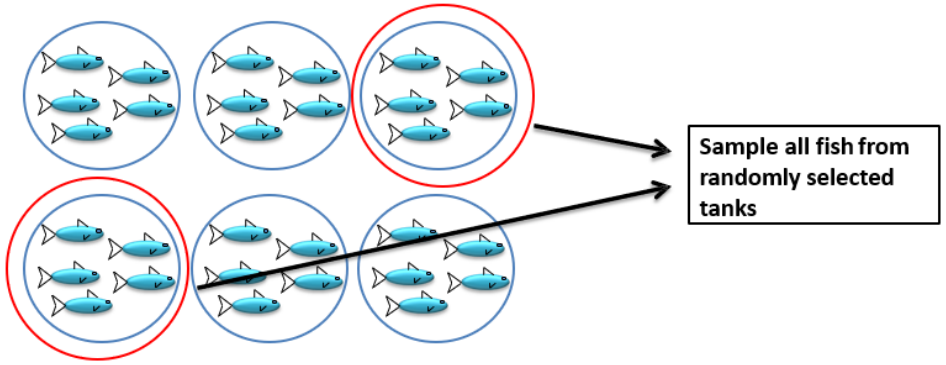
Stratified random sampling

- Sampling frame is divided into groups (strata) of defined common characteristics.
- **Procedure**
  - Random selection of animals within each stratum (proportional or not)
  - e.g., age groups for pet cats; 10 needed from 100 cats

  (20 young + 80 old)
- **Advantages**
  - Ensures that all strata are represented in the sample
  - More precise estimate can be obtained
  - Strata-specific results are available
- **Limitations**
  - Status of the sampling units with respect to the stratification factor needs to be known in advance (e.g., age)
  - Requires more complex methods to estimate sample size



Cluster sampling



- Sampling frame is divided into logical aggregations (clusters) and a random selection of clusters is performed
- All units within the selected clusters are included in the sample
- Cluster = Sampling Unit
- Elements within cluster = Units of Concern

PSU is larger than the unit of concern

Cluster may occur in time or space

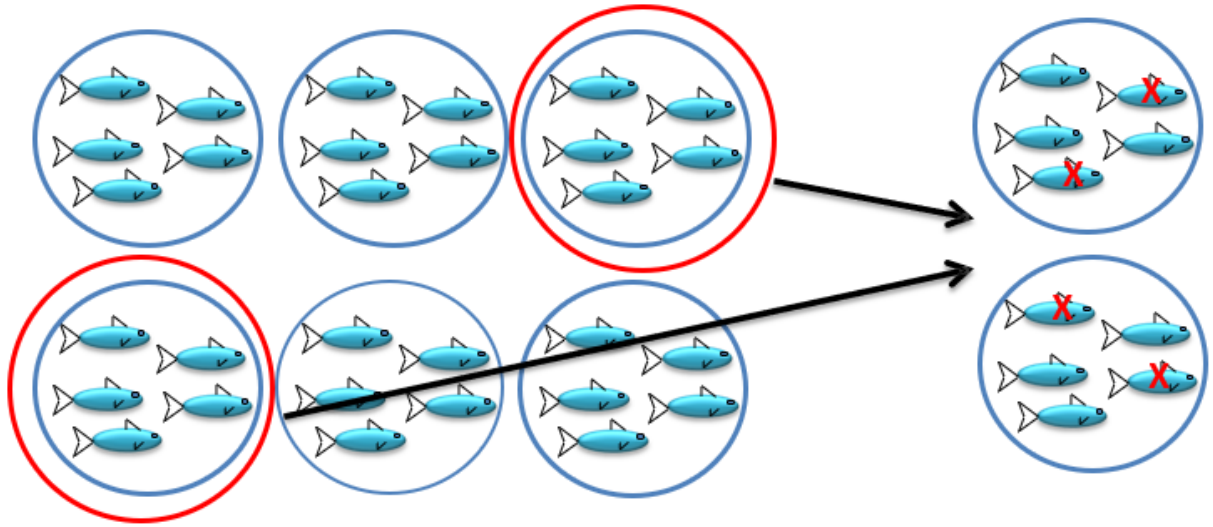
- **Advantages**
  - Only requires sampling frame for the "groups"
  - Fewer sample units (clusters) are required to achieve a certain sample size of study units
  - More economical approach than SRS

- **Limitation**
  - Higher variability than in SRS, so greater sample size is required to achieve the same precision

Multistage sampling

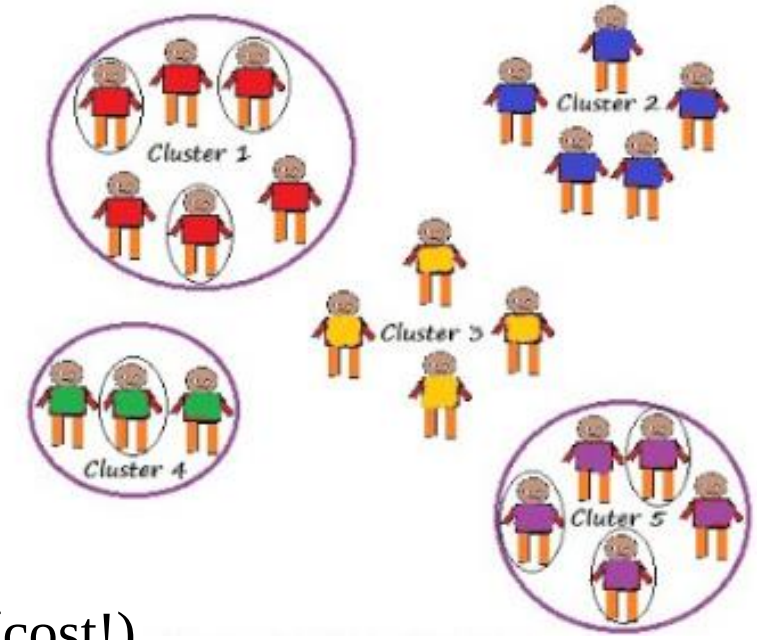
- Same as cluster sampling, but take a ‘sub-sample’ of the subjects within a selected cluster (two-stage or more)

- Why multistage strategy?
  - Too many subjects per cluster
  - Subjects are similar within cluster (little-no additional info)



- **Procedure**
  - Clusters randomly selected (PSUs)
  - Individuals randomly selected within each selected cluster (Secondary Sampling Unit = SSU)
  - …..

Multistage sampling

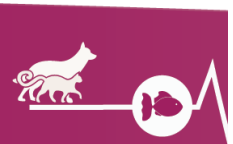


- **Advantages**
  - Only requires list of PSUs (not all individuals)
  - Most practical for large projects (national or regional level)
  - The number of PSUs & SSUs to be sampled can be adjusted
    - To take into account variability between & within PSUs (cost!)
    - e.g., if within-cluster variation is lower relative to between-cluster variation?

- **Limitations**
  - Larger sample size may be required to achieve the same precision as SRS, due to dependence among individuals within clusters
  - Complexity of statistical methods to determine sample size & variation among parameter estimates

Comparing main sampling strategies (examples)

Population characteristics	*Example*	*Appropriate sampling strategy*
homogeneous	cattle on farm; sampled to determine tuberculosis prevalence	simple random
definite strata, but homogeneous within strata	farm with 2 different dairy breeds and with similar numbers of each; sampled to determine milk production	stratified
definite strata, each stratum has proportionate ratio of number of members relative to total	farm with 2 different dairy breeds, but very different numbers of each; sampled to determine milk production	proportional stratified
groups with similar characteristics, but heterogeneous within group	veterinary laboratories in country equipped according to standard; wide variation between samples submitted to each; sampled to determine proportion of contaminated tissue samples	cluster
no sampling frame for units of interest	cattle in region; sampled to determine tuberculosis prevalence	multistage



Sample size (a brief reminder)

- Some online calculators:
  - http://www.sample-size.net
  - http://clincalc.com/stats/samplesize.aspx
  - http://hedwig.mgh.harvard.edu/sample_size/size.html#ssize
  - http://www.openepi.com/SampleSize/SSCohort.htm
  - http://powerandsamplesize.com/
  - https://epitools.ausvet.com.au/riskbasedsscomplex

Epi Info

Centers for Disease Control and Prevention

100K+
Downloads

3+
Rated for 3+

Epi Tools 12+

Epidemiology Calculation Tool

Mark Stevenson

Designed for iPhone

★★★★★ 5.0 • 1 Rating

Free

CityU

Jockey Club College of Veterinary Medicine and Life Sciences

香港城市大學
City University of Hong Kong
in collaboration with Cornell University

Apps!

Epi Info™

Epi Info™ is a public domain suite of interoperable software tools designed for the global community data entry form and database construction, a customized data entry experience, and data analyses w professionals who may lack an information technology background. Epi Info™ is used for outbreak inv systems; as analysis, visualization, and reporting (AVR) components of larger systems; and in the cont analytic methods at schools of public health around the world.

Epi Info™ for Windows

Create forms, collect data, and perform epidemiologic data analysis and visualization. Appropriate for small to medium size surveillance and response activities and special epidemiologic studies.

More

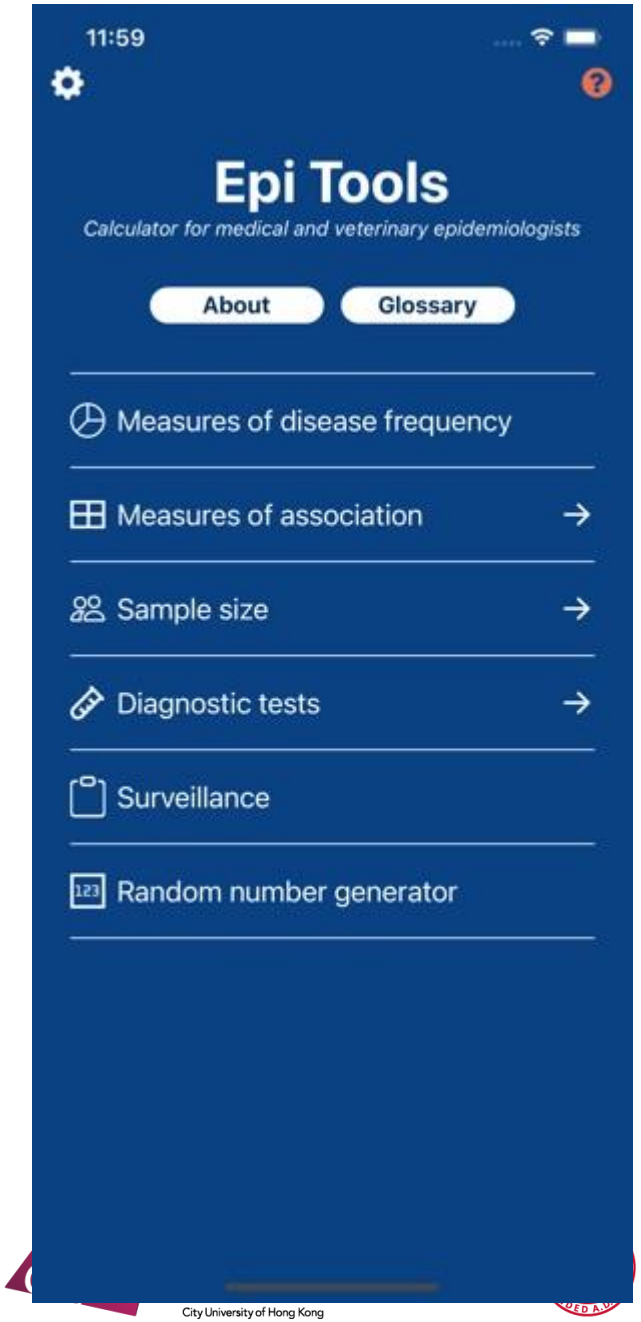
Epi Info™ for Mobile

Load your Epi Info™ forms on tablets or smart phones and conduct epidemiologic studies in the field. Appropriate for distributed data collection in locations lacking IT infrastructure.

More

Web based and cloud-optimized components for data collection, analysis and visualization. Appropriate for large scale surveillance and response activities in locations with reliable network connectivity.

More



City University of Hong Kong
in collaboration with Cornell University

Web-based tools

Sample Size Calculators
for designing clinical research

UCSF Clinical & Translational Science Institute

Explore the *Training in Clinical Research* Program at UCSF

Home

Calculators

- CI for proportion
- CI for mean
- Means - effect size

Sample Size Calculators

ClinCalc.com

Home | Academy | Blog | About

Sample Size Calculator

Determines the minimum number of subjects for adequate study power

ClinCalc.com » Statistics » Sample Size Calculator

Study Group Design

- Two independent study groups vs.
- One study group vs. population

Two study groups will each receive different treatments.

Primary Endpoint

- Dichotomous (yes/no)
- Continuous (means)

The primary endpoint is **binomial** - only two possible outcomes.
Eg, mortality (dead/not dead), pregnant (pregnant/not)

Search

Enter search term ...

Related Calculators

- Post-hoc Power Calculator

Follow Us!

Get Email Updates

New and Popular

- ASCVD Risk Calculator
- ClinCalc DrugStats 2021 Update – The Most Commonly Prescribed Drugs in the United States
- Opioid Equianalgesic Calculator
- RxHero for iOS: Educational Gaming of the Top 250 Drugs
- Acute Physiology and Chronic Health Evaluation (APACHE II) Calculator

EPITOOLS

Home | Prevalence | Freedom | Studies | Diagnostics | Sampling

Epitools - Epidemiological Calculators

This site is developed and maintained by Ausvet. The site is intended for use by epidemiologists and researchers involved in estimating disease prevalence or demonstrating freedom from disease through structured surveys, or in other epidemiological applications.

EPITOOLS

Home | Prevalence | Freedom | Studies | Diagnostics | Sampling

Sample size calculations

These utilities can be used to calculate required sample sizes to estimate a population mean or proportion, to detect significant differences between two means or two proportions or to estimate a true herd-level prevalence.

Epidemiological studies

- To estimate a single proportion
- To estimate a single mean
- Two proportions
- Two means with equal sample size and equal variances
- Two means with unequal sample size and unequal variances
- To estimate true prevalence (at animal or herd-level)
- Sample size for a cohort study
- Sample size for a case-control study

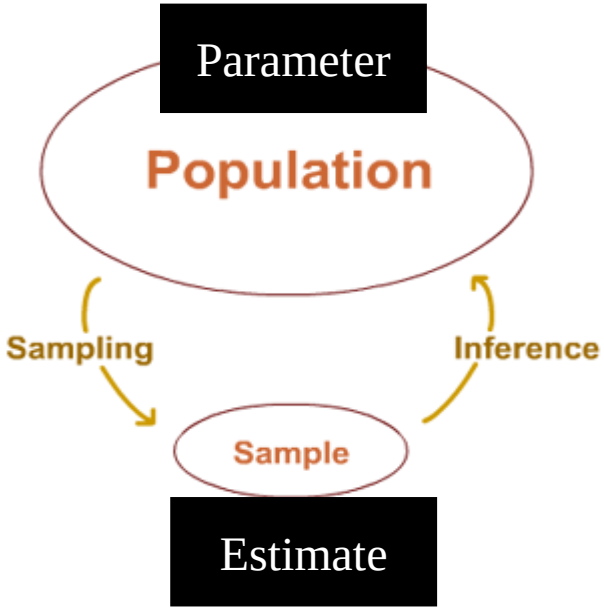
Sample size to demonstrate disease freedom

- Sample size assuming perfect test specificity
- Sample size for pooled sampling in a large population
- Sample size to achieve target confidence of freedom
- Design prevalence required to achieve target population sensitivity for given sample size
- FreeCalc sample size calculation for imperfect tests

City University of Hong Kong
in collaboration with Cornell University

Simple sample size formula

- Estimating proportion (e.g., prevalence of a disease)
- Estimating mean (e.g., average blood pressure)
- Finite vs infinite populations (FPC)



$$n = Z^2 \times \frac{p(1-p)}{L^2}$$

- n = required sample size
- p = expected prevalence (expressed as a proportion)
  - if not known use p=0.5 (50% prevalence)
- L = desired precision (expressed as a proportion here)
- Z value
  - 1.96 for 95% confidence level (1.65 for 90%, 2.58 for 99%)

$$n = \frac{Z_\alpha{}^2 * \sigma^2}{L^2}$$

Z_α= Z value for desired confidence level: 1.96 for 95%

σ= estimated standard deviation of parameter of interest

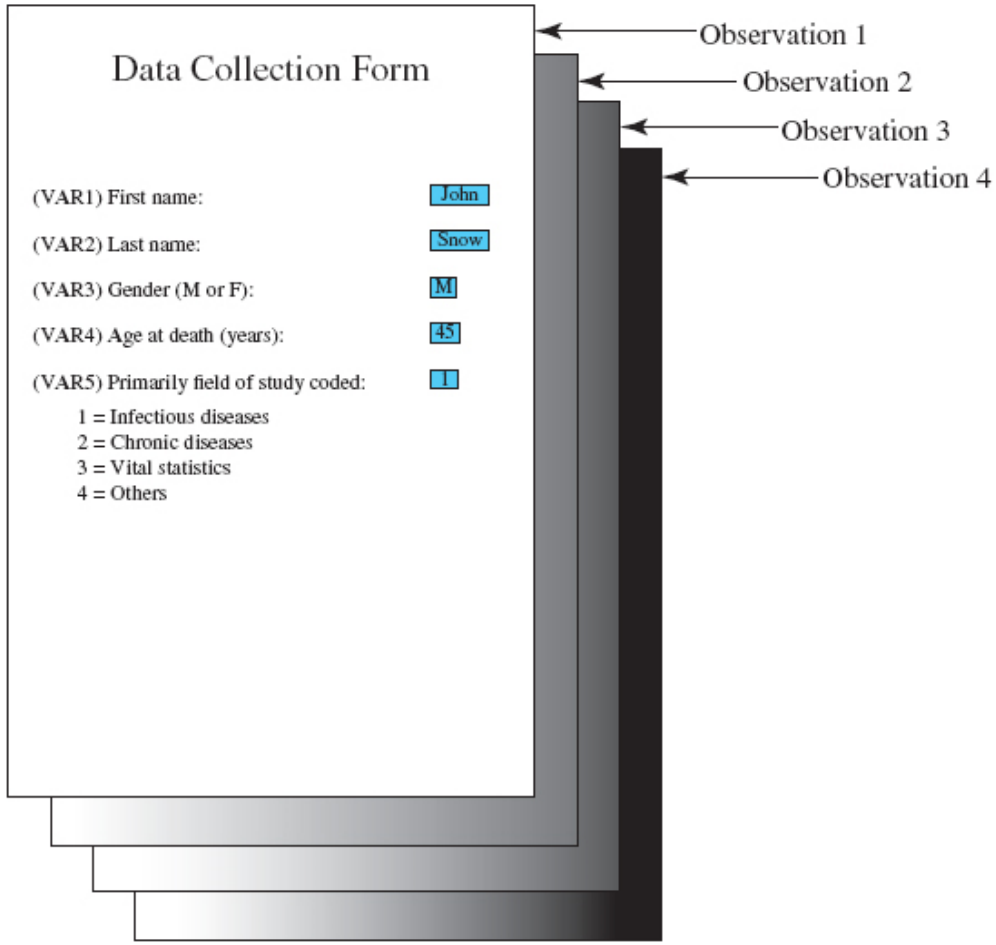
L = how accurate estimate is supposed to be (in units of parameter of interest)

Collect and Organize Data?

- **Measurement**
  - Is how we get our data
  - Is the assigning of numbers and codes according to prior-set rules (Stevens, 1946)
  - There are three broad types of measurements: Categorical, Ordinal and Quantitative
- **Observation**
  - The unit upon which measurements are made
  - Can be an individual or aggregate
- **Variable**
  - The generic thing we measure
    - e.g., AGE of a person
    - e.g., HIV status of a person
- **Value**
  - A realized measurement
    - e.g., “27”
    - e.g., “positive”

Data Table

- Each **row** corresponds to an **observation**
- Each **column** contains information on a **variable**
- Each **cell** contains a **value**



VAR1	VAR2	VAR3	VAR4	VAR5
John	Snow	M	45	1
William	Farr	M	75	3
Joseph	Goldberger	M	54	2
Janet	Lane-Claypon	F	90	2

Measurement Scales

Variable

Categorical (qualitative)

- Categories and usually fixed.
- Blood type, gender, …

Nominal

- Have two or more categories
- **NO specific order or direction**.
- Blood type, gender, eye color

Dichotomous

- Or Binary
- Categories have two values only.
- Sex: Male/Female Yes/No or 0/1

Polytomous

- More than two possible values.
- Blood type

Ordinal

- Have two or more categories
- **Can be ordered or ranked**.
- Education level (high school, BS, MS, PhD).

Numerical (quantitative)

- Any value within a finite or infinite interval.

Discrete

- Whole numerical value
- Typically countable.
- Number of students in class.
- Number of visits to dentist.

Continuous

- Any value within a range.
- Uncountable
- Height in cm
- Weight in Kg

Interval

- Numerical scale where the order and difference of variables is known.
- Temperature scale

Ratio

- Numerical

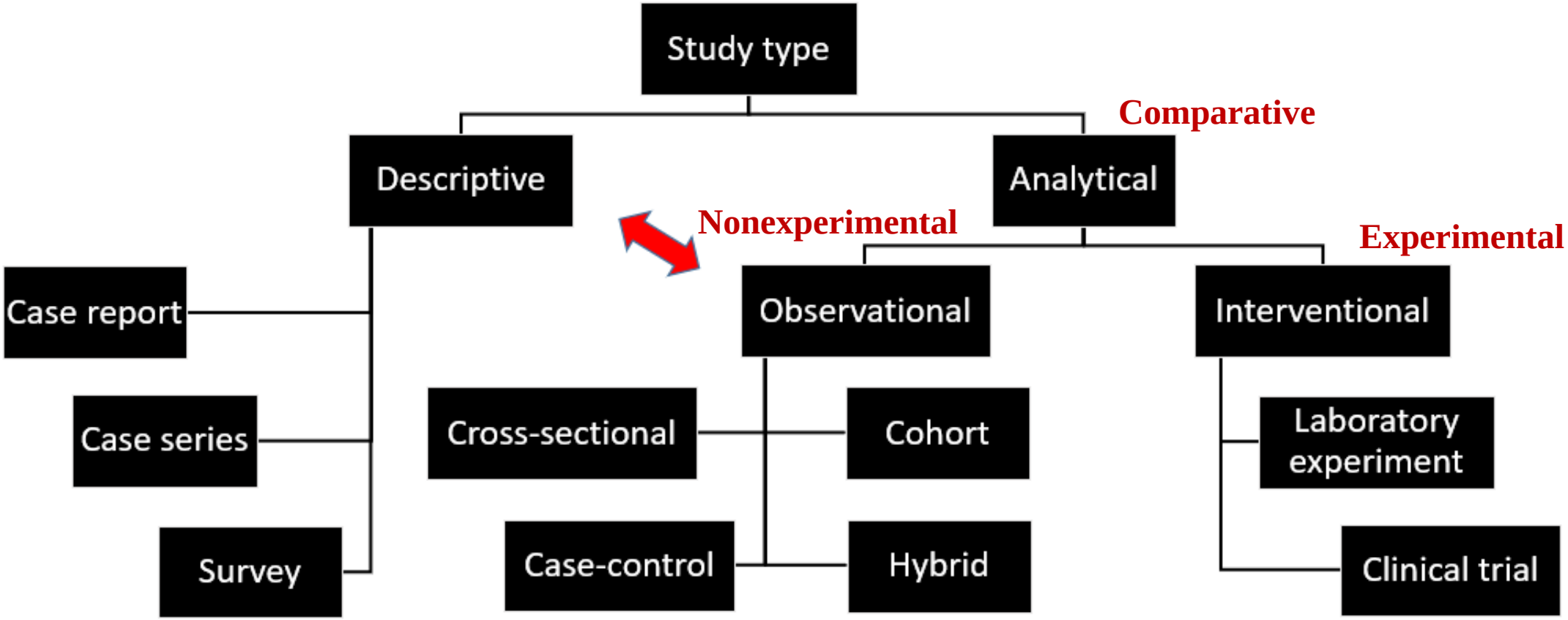
Data Quality

- An analysis is only as good as its data

- Does a variable measure what it purports to?
  - Validity = freedom from systematic error
  - Objectivity = seeing things as they are without making it conform to a worldview

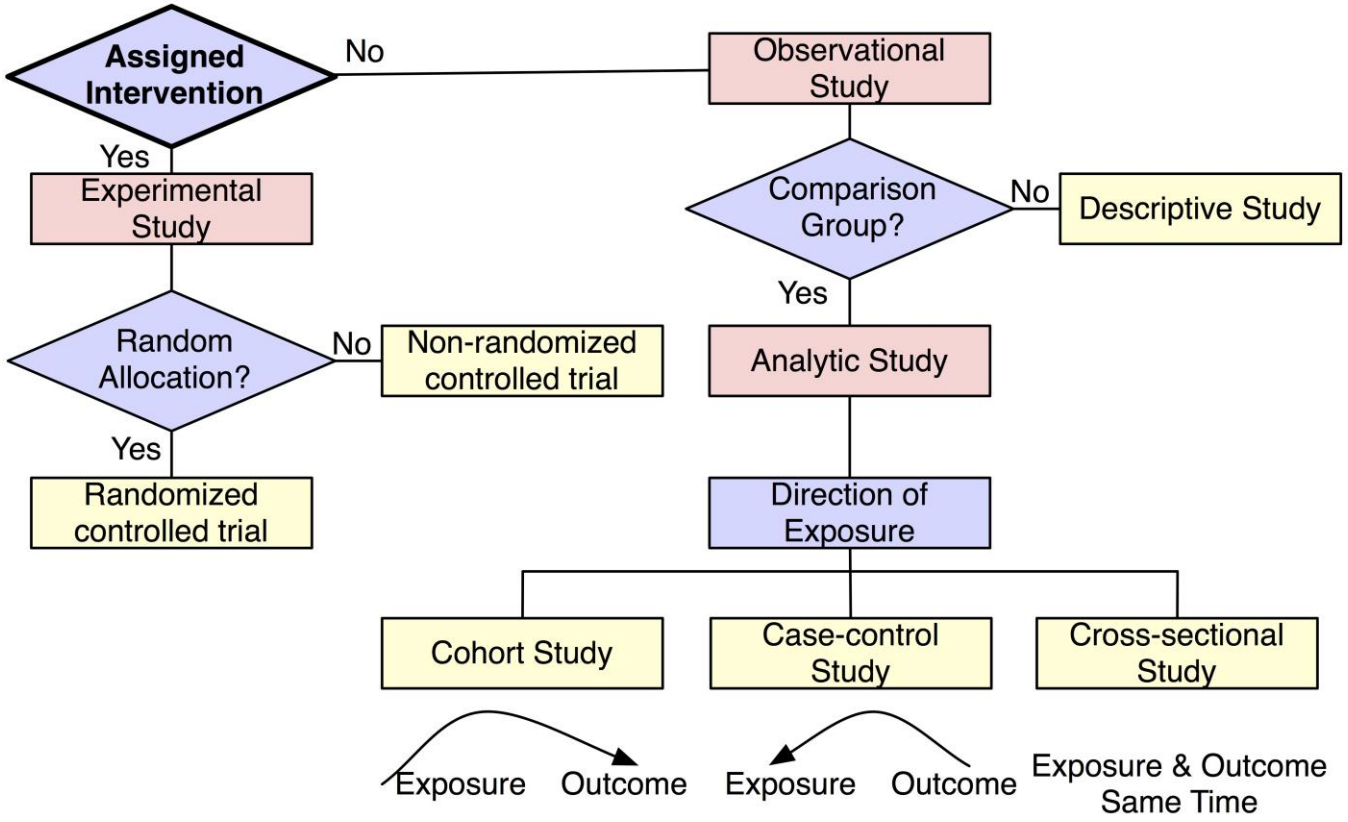
- Consider how the wording of a question can influence validity and objectivity

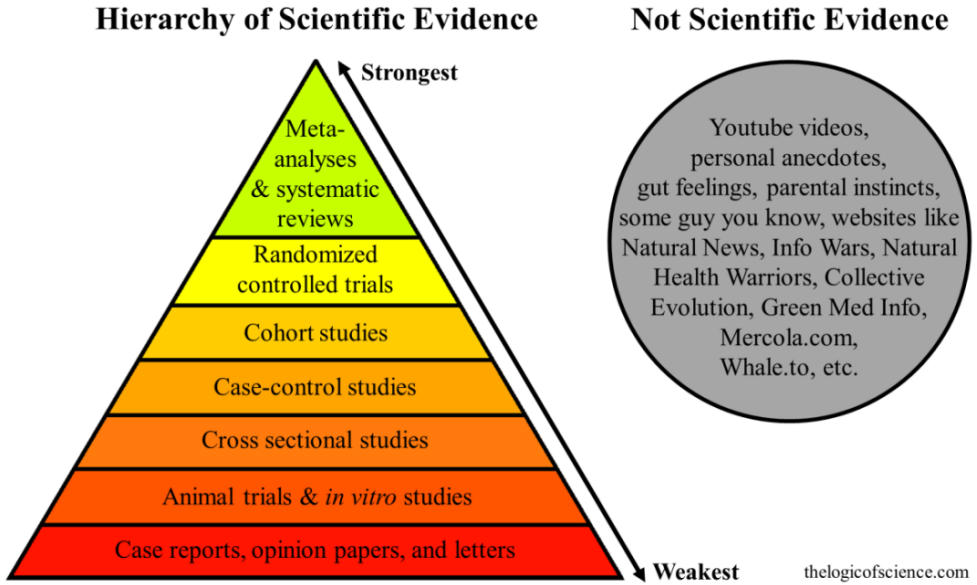
Study design/types



Strength of evidence for potential causality

Study design/types





Measures of disease frequency

- **Incidence risk (R)**

$$R = \frac{new\ cases\ in\ a\ defined\ time\ period}{population\ at\ risk\ in\ the\ onset\ of\ that\ period}$$

  - Probability that a subject will develop the outcome of interest (e.g., disease) over a defined time period.
  - Used for individual predictions.

- **Incidence rate (I)** (speed of disease spread)

$$I = \frac{new\ cases\ in\ a\ defined\ time\ period}{sum\ of\ all\ individual's\ time\ at\ risk\ in\ that\ period}$$

  - New cases developed in a population per unit of animal-time during a given time period (unit: 1/time)

- **Prevalence (P)**

$$P = \frac{number\ of\ present\ cases\ in\ a\ time\ point}{total\ population\ at\ risk\ in\ that\ time\ point}$$

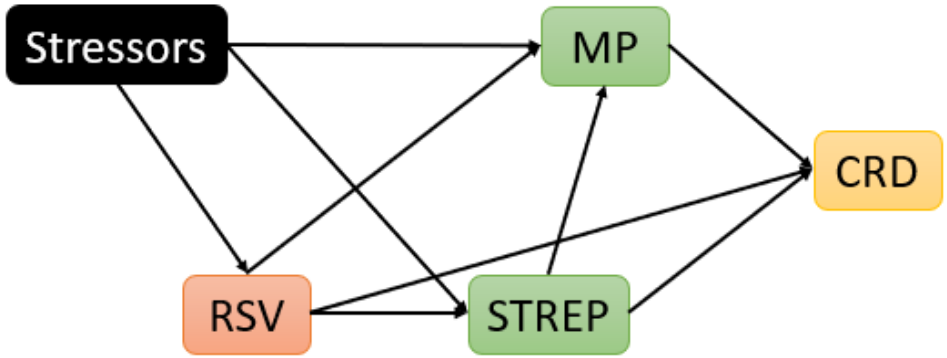
  - Probability that a subject has a specific disease at a certain point/period in time
  - Snap-shot of the population status

Measures of association and effect

- **Epi goal:** to identify causes of health /disease in populations
  - Complex
  - Factors that may be causal

- "A cause is any factor that produces a change in the severity or frequency of the outcome" (Dohoo et al., 2012)
  - Individual vs. population level
  - Single vs. multiple cause/effect

- To learn the truth about 1 potential cause, control for 'confounders'



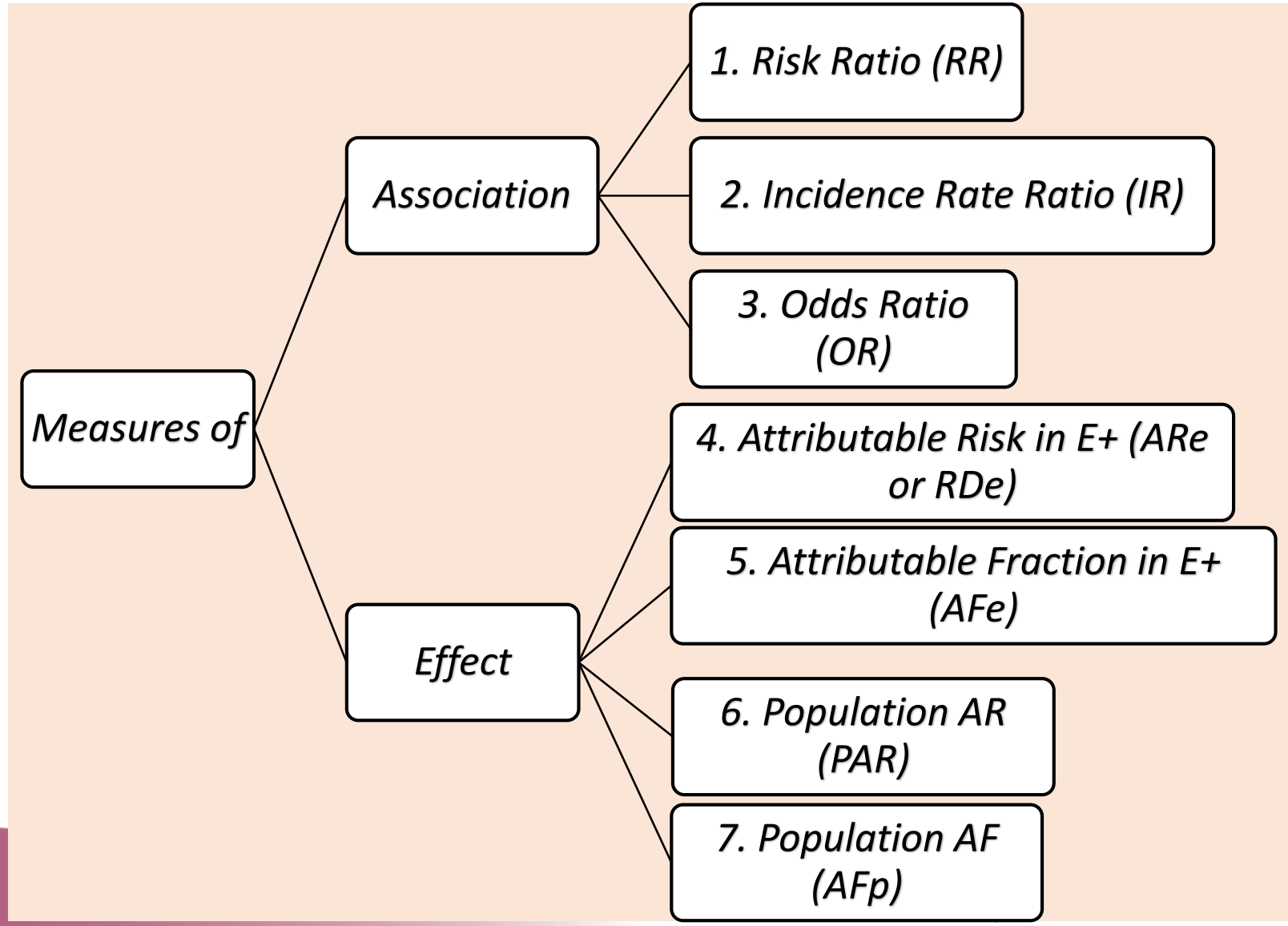
http://www.dagitty.net/

Causal-web model

- Conceptualizes how multiple factors can combine to cause disease
  - An interconnected network
  - Complements the component-cause model
- Causal diagrams (directed acyclic graphs, modified path models).
- Use factors portrayed in the sufficient-cause approach & link them temporally
- Direct & Indirect causes
  - Direct = proximal (e.g., infectious agents & toxins)
  - Indirect = mediated through "intervening" variables (stress)
- Control & prevention measures on indirect causes (e.g., John Snow)
- It can guide study design, statistical analysis approach, & interpretation of data.
- **Never start a study without one!**

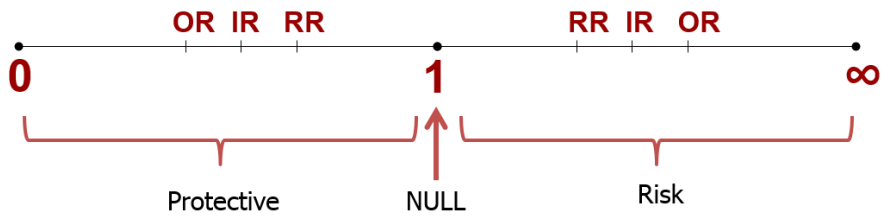
Measures of association and effect

- In studying associations (X →Y or E →D) in our course, please be reminded that:
- Exposure = independent variable = explanatory variable = X = (sometimes, the risk/preventive factor of interest).
- Outcome = dependent variable = Y = (sometimes, the disease of interest).



Measures of association (MoA)

For a simple 2-by-2 table: Exposure (/a potential risk factor) → Disease (/an outcome)



		Exposure		
Disease		+	-	
	+	a1	a0	m1
	-	b1	b0	m0
		n1	n0	n
S-time		t1	t0	t

"Measures of association"

1. **RR = Risk Ratio** = $\frac{(\frac{a1}{n1})}{(\frac{a0}{n0})}$

Risk of (developing) the disease in 'E+' group is 'RR' times higher than the risk of disease in 'E-' group.

2. **IR = Incidence Rate Ratio** = $\frac{(\frac{a1}{t1})}{(\frac{a0}{t0})}$

Rate of (developing) the disease in 'E+' group is 'IR' times higher than the rate of disease in 'E-' group.

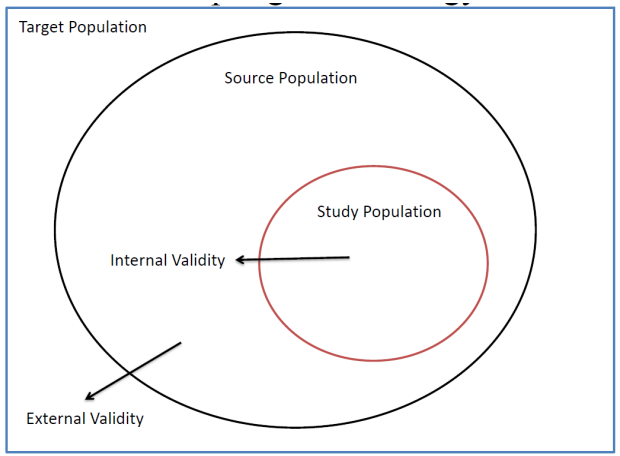
3. **OR = Odds Ratio** = $\frac{(a1 \times b0)}{(a0 \times b1)}$

Odds of the disease in 'E+' group is 'OR' times higher than the odds of disease in 'E-' group.

Or: odds of the exposure in 'D+' group is 'OR' times higher than the odds of exposure in 'D-' group (more appropriate for 'case-control' studies).

Validity in Observational Studies

- **<u>Terminology:</u>**
- **Validity**
  - Absence of systematic bias
  - (*i.e.* measures of association in study groups accurately estimates a true parameter in the source population.)
- **Internal Validity**
  - Inferences drawn from the study group(s) are correct for the source population
- **External Validity**
  - Inferences drawn from the study group(s) are correct for a target population
  - Source population completely reflects the target population
- **Generalizability**
  - Inferential step beyond external validity
  - Refers to the ability to develop scientific theories extended to broad populations.



Validity in Observational Studies

- **Bias**

- **Three Major Types of Bias:**
  - **Selection Bias**
    - Arises due to factors affecting the selection of study subjects.
    - Other factors that relate to the willingness of potential study subjects to participate in a research project.
  - **Information Bias**
    - Arises due to factors relating to obtaining accurate information on the exposure, outcome and covariates of interest.
  - **Confounding Bias**
    - Arises due to the effects of factors other than the exposure of interest on the observed measure of association.

Selection Bias

- Inappropriate procedures to ***select*** and ***keep*** subjects fully participating in the study:

1. Eligible to be selected

- Survivor Bias
- Detection Bias
- Admission Risk Bias

2. Agree to fully participate

- Response Bias
- Missing Data Bias

3. Stay in their study group(s)

- Lost to Follow-up
- Follow-up Bias

Information Bias

- Collective term for Misclassification Bias and Measurement Bias

- **Misclassification Bias**
    - Describing errors in classification of categorical variables.

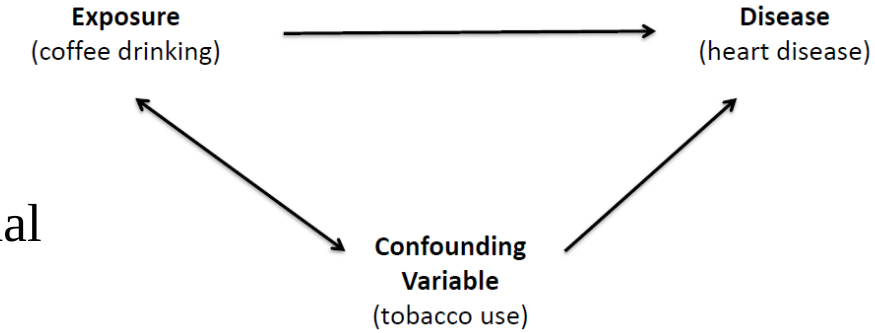
- **Measurement Bias**
    - Erroneous measure of continuous variables.

- Can alter the magnitude and direction of estimates of association.
    - Not always intuitive which direction the bias is towards.

Information Bias

- **Types of Information Bias**
  - Recall Bias
  - Blinding Bias
  - Diagnostic Bias
  - Interviewer Bias
  - Social Desirability Bias

Confounding



- Is a systematic error that results from unaccounted-for differential distributions of particular covariates.
- Confounding is a form of bias
- One of the key concepts of epidemiology!
- Confounding can have a very important influence, and may even change the apparent direction of an association
  - A variable that appears to be protective may, after control of confounding may be found to be harmful.
- Most common concern about confounding is that it may create the appearance of a cause-effect relationship that does not actually exist.
- Confounding is situation-specific, and you have to know something about the biology and logic of the situation to guess at things that should be explored as confounders.
- In most studies, you should think about the following types of variables as being potential confounders: age, sex, species, production type and physiological status.

Questions?