Quantitative

Ann Mwangi, PhD

Moi University

annwsum@gmail.com

December 1, 2021

Overview

1. What are study designs
2. Types of study designs
3. Ethical issues
4. Appraising Research Methodology

What are study designs

1. Structured approaches to address specific research questions
2. Overall scheme or program of the research.
3. Includes outline of what the investigator will do,from writing the hypotheses and their operational implications to the final analysis of data.
4. Provide general guidelines for thinking about specific aspects of study conduct
    - Sampling
    - Systematically collecting measurements
    - Analysing data

Why design is important

- The trustworthiness of any research study is predicated initially on several major elements:
  - The suitability of the proposed research design or methodology to address the specific questions posed by the study;
  - The scientific rigor by which the methodology is applied;

Why design is important

- Poor design may not answer question
- Choice of design determines the type of analysis
- Incorrect design leads to waste of resources
- Poor design is unethical if patient exposed to danger for little return

Right study design

1. Relevant; Allows to answer research question
2. Novel; Design allows meaningful contribution to existing knowledge
3. Feasible; Design allows study to be done within available time and funds
4. Simple; Avoid unnecessary complexities

1. Ethical choice of study designs
2. Is the design methodologically sound
3. Have expertise to do it. Pilot?

Broad study designs

1. Experimental; Randomised control trials
2. Observational; Investigator has no control over the exposure

Observational study Designs

1. Descriptive; Prevalence of condition Z in a specific population
   - Case Reports and Case series (Clinical)
   - Cross sectional (Epidemiologic)
2. Analytic; What are the factors associated with condition Z. Is X a risk factor for Z
   - Cohort
   - Case Control
   - Ecological
3. Diagnostic; How good is test Q in detecting condition Z

Hierarchy of evidence

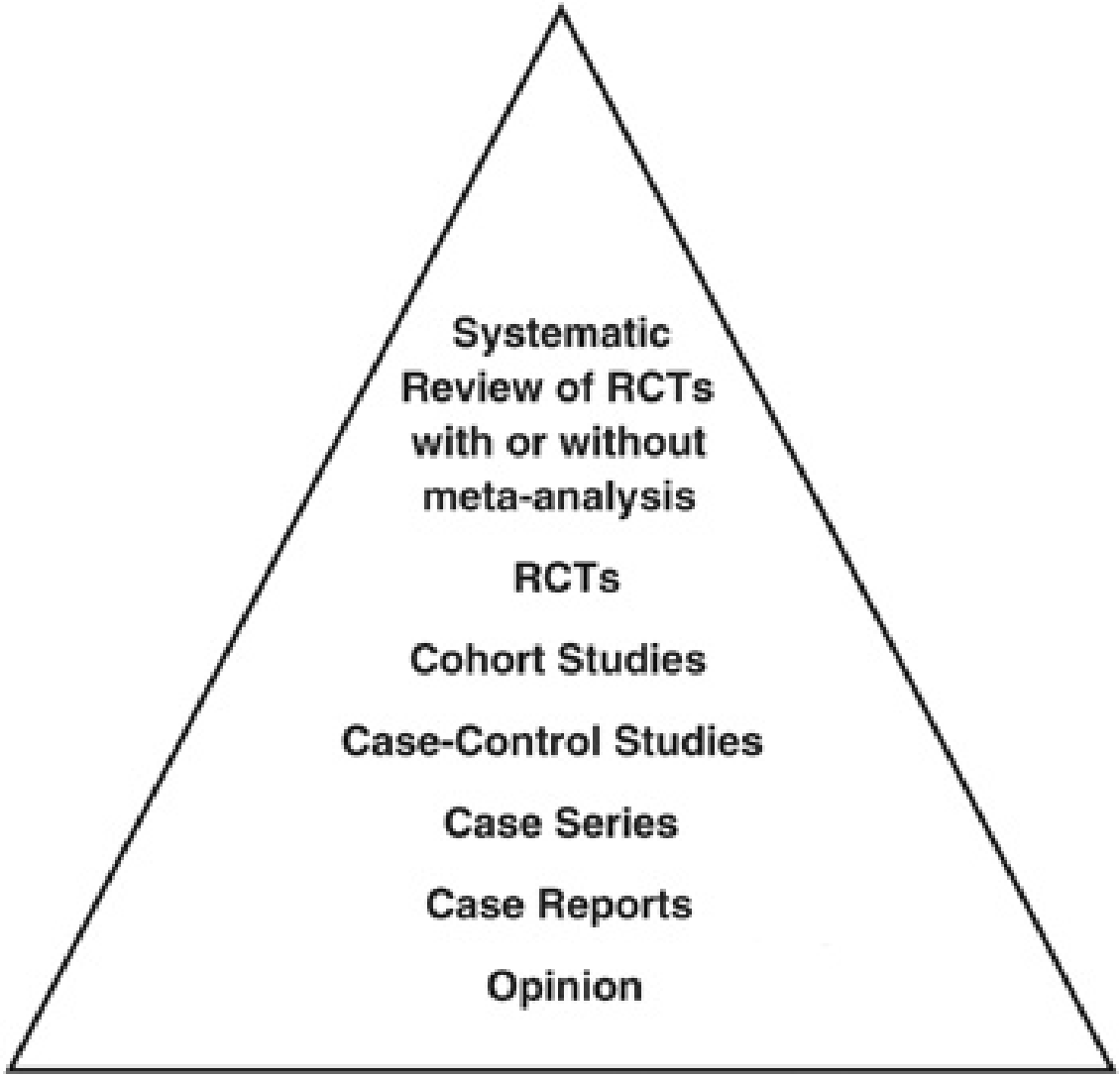


Diagram taken from Akobeng 2005

Evidence

- Design alone does not make evidence better or worse
- Details of how conducted matters
  - Rigour in design
  - Analysis
  - Sampling
  - Measurement

Study design

1. Informs sampling; exposure, disease , neither or both
2. When measurements are taken; pre or post
3. How outcome are measured; Incident or prevalent
4. Whether comparison group are involved
5. Informs appropriate measurements to collect on participants

The designing decisions happen to be in respect of:

1. What is the study about?
2. Why is the study being made?
3. Where will the study be carried out?
4. What type of data is required?
5. Where can the required data be found?
6. What periods of time will the study include?
7. What will be the sample design?
8. What techniques of data collection will be used?
9. How will the data be analyzed?
10. In what style will the report be prepared?

1. Did the study ask a clearly focused question

- Population under study
- Intervention / exposure
- Outcome (s)

1a. Study population

- Vulnerable populations
- Over researched populations
- Inclusion criteria
- Exclusion criteria

2. Study design

Is it the most suitable one for addressing the study question

3. Randomisation

- Participants should be randomly allocated to the groups
- Is the method described
- Are the groups well-balanced
- Steps taken to overcome issue of lack of balance

4.Blinding

- Preventing those involved in a trial from knowing to which comparison group (placebo/ intervention) a particular participant belong
- Is the risk of bias minimized
- Single / double blind (not always possible)
- Use of placebo / standard of care

5. Follow-up / Intention to Treat

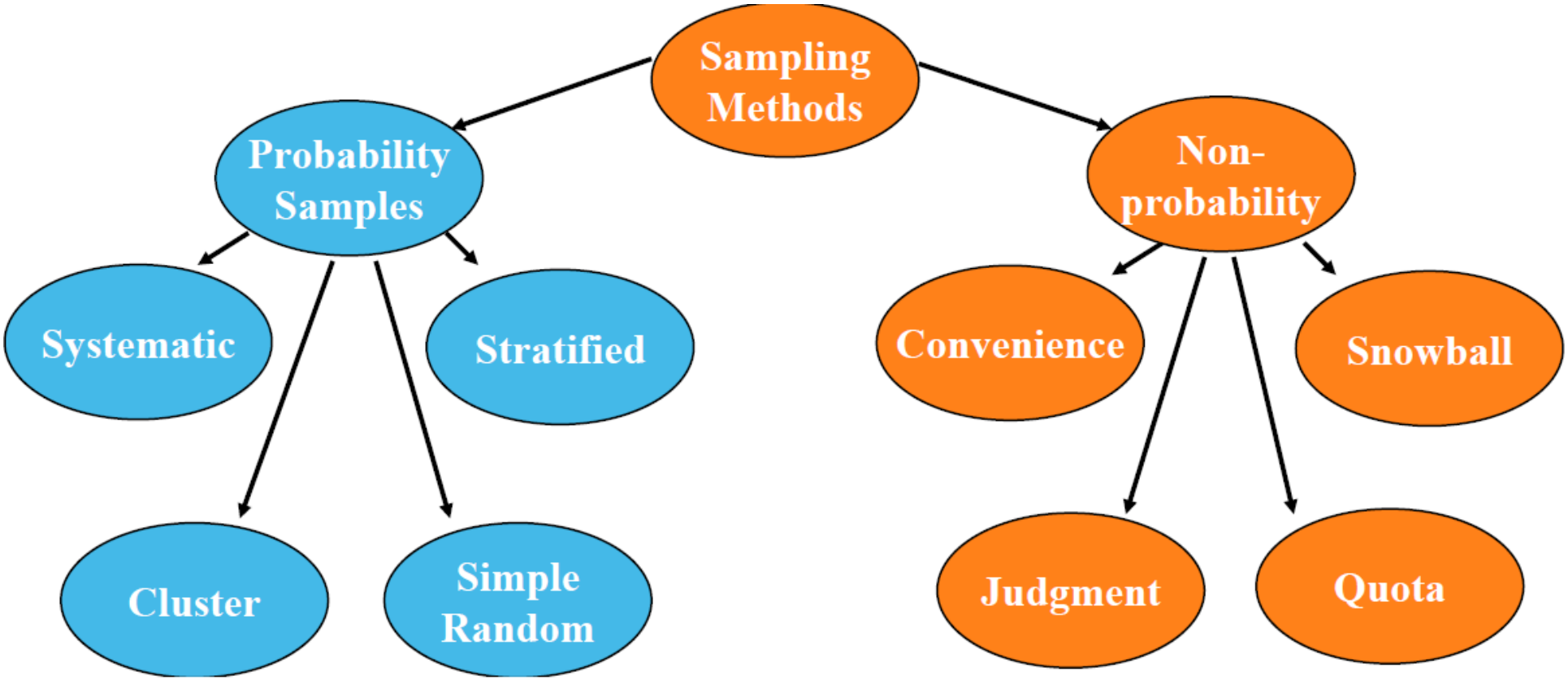
- plan to minimize lost to follow up or how will be addressed
- missing data
- ITT - Analyzing people at the end of trial in the groups to which they were randomized even if they did not receive intended intervention

6. Eligibility Criteria

- Inclusion and Exclusion criteria should be adequately described and examined carefully

Classification of sampling methods

- Ensure representativeness of the group from which it is drawn
- Probabilistic / non-probabilistic
- Rationale for the sampling strategy especially non-probabilistic



Sampling

- Included
- Excluded
- 'Opt-in' and 'opt-out' sampling
- 'Opt-out' samples are those where participants are contacted without volunteering to take part in the research and excluded only when they say they are unwilling to participate But people may find it difficult to say no to a researcher for a variety of reasons

Example

Imagine that your boss has been asked to nominate workers to take part in interviews in a management study in your organisation. You do not want to do it - you are feeling very over-worked, and have serious concerns about the management of your workplace, and are worried that your views might be identifiable. But can you say no? Or is your boss entitled to ask you to do this, as part of your professional role?

Issues with 'Opt-In'

- Low response rate
- Representativeness- selection bias

Why sample size

- Requirement (Clinical Research Protocol, Funding Agencies, etc) in many grant application
- Budgetary Constraints
- Provide Statistical Justification
- Inference (decision) is based on it

Scientific perspective:

- Cant be sure weve made right decision regarding the effect of the intervention
- However, we want enough subjects enrolled to adequately address study question to feel comfortable that weve reached correct conclusion.

Ethical implication:

Too few subjects:

- Cannot adequately address study question. The time, discomfort and risk to subjects have served no purpose.
- May conclude no effect of an intervention that is beneficial. Current and future subjects may not benefit from new intervention based on current (inconclusive) study.

Too many subjects:

- Too many subjects unnecessarily exposed to risk. Should enroll only enough patients to answer study question, to minimize the discomfort and risk subjects may be exposed to.

Sample size

- Based on primary outcome measure
- Power calculation
  - Used to calculate the sample size necessary to detect a true diffence between outcomes in control/intervention groups
  - Allow researchers to work out how large a sample they will need to use
  - Power of 80-90 % is standard
  - Level of confidence 5%
  - Level of precision $<= 5\%$
- Citation of sources for figures used in the calculation
- Response rate

Case control

- Sources of controls
- Matching and on what variables
- Population based controls for

Data Collection

- Validity of measurements and outcomes identified in the study
- Data collection -instrument (new or existing)
- Questions should be linked to the research question

Research Method

- How well do your chosen methods fit the aims of your research? In other words, do the methods offer the most reasonably efficient way of answering the research questions?
- What are the strengths and limits of the methods?
- Have they been used before effectively in a similar context?
- Are they respectful of your respondents capacity and willingness to participate? Some methods will work better with some groups than with others.
- Are there any potential unintended consequences of your research (e.g. disclosures of sensitive information) that may arise through those methods?
- Do the methods proposed fit with the ESRC six core ethics principles as set out in the Framework for Research Ethics? If not, can the exception be justified? Is what you are proposing to do justifiable in terms of the benefits, risks and harms of your research?

Data

- How their data will be stored
- If they will have access to their data.
- How long their data will be kept for.
- Data anonymisation
- Secondary analysis: Make sure that consent covered your proposed use of the data. For example, did they agree for data to be archived, or to be used for future research?

- Who needs to have access to hard data?
- Will these data be anonymised before they are stored? If not, why not?
- Will these data be stored separately from personally identifying data?
- Where will the key be stored?
- Could any one find it and access the data who should not?
- How will you deal with hard copies in the period between data collection and data storage?
- Capacity building

Bias

- Selection bias
- Confounding
- Measurement

Data analysis plan

An analysis plan is a document you will develop in advance to guide data analysis. The analysis plan usually contains:

- research question(s) and/or hypotheses, if any,
- dataset(s) to be used,
- inclusion/exclusion criteria (e.g., if data only for adults or only for children will be analyzed),
- variables to be used in the main analysis (the main exposure, outcome, and stratifying variables),
- statistical methods and software to be used, and,
- key table shells (univariable, bivariable, and stratified)

Sample Data Analysis: Binary Outcome

Suppose we wish to assess blood pressure control with the following objectives

1. Determine proportion of patients with blood pressure control among patients with Diabetes
2. Compare proportion of patients with blood pressure control among patients with Diabetes and without Diabetes
3. Determine factors associated with blood pressure control other than Diabetes

Sample Data Analysis: Binary Outcome

Objective	Dependent variable	Independent variable	Test
1	BP control	-	Proportion
	Binary (Yes/No)	-	95% CI
2	BP control	DM status	Compare Proportion
	Binary (Yes/No)	Binary (Yes/No)	Z test for proportions
3	BP control	**Categorical**	Chisquare
	Binary (Yes/No)	DM status, Gender	Fishers'exact
		Education , Marital status	
		Continous	Ttest, ANOVA
		Age, BMI	Kruskal Wallis, Wilcoxo

In objective 3 to adjust for confounders, logistic regression model.

Sample size :Examples

Sample size determination is a perennial question that is asked by most researchers

Objectives: To determine sample sizes for:

- A single mean
- Two independent means
- A single proportions
- Case control studies

Sample size prevalence study

How large a sample would be necessary to estimate the true proportion of defectives in a large population within 3%, with 95% confidence? (Assume a pilot sample yields p= 0.12)

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

where e is the margin of error

Sample size one mean study

If $\sigma = 45$, what sample size is needed to estimate the mean within 5 with 90%confidence?

$$n = \frac{Z^2\sigma^2}{e^2}$$

where e is the margin of error

Sample size comparing two proportions

Imagine a doctor wanted to set up a double-blind trial of a new drug, to compare mortality after a stroke among patients using the new drug or a placebo.

- Measure: death from any cause within one year of first treatment
- Analysis: comparison of proportion of deaths amongst new drug and placebo patients, using chi-squared at $\alpha = 5$ per cent significance
- Standard treatment: 90 per cent expected to survive at least one year on placebo
- Power required: if the new drug can halve the mortality (reduce deaths from 10 to 5 per cent), this should be detected 90 per cent of time (power $= 0.9$, $\beta = 0.1$))

In summary:

- p_0 = proportion of successes on standard treatment = 90%
- p_1= proportion of successes on the new drug which indicate it as more effective = 95%
- $\alpha = 0.05$
- $\beta = 0.1$

$$n = \frac{(Z_\alpha + Z_\beta)^2 * (\pi_0(1-\pi_0) + \pi_1(1-\pi_1))}{(\pi_1 - \pi_0)^2}$$

n=580 per group

Sample size comparing two means

A clinical trial tests the preventive effect upon neonatal hypocalcemia of giving Supplement A to pregnant women. Women are randomised and given either placebo or Supplement A.

- Measure: serum calcium level of baby one week postnatally
- Analysis: Comparisons of difference between two groups of babies using an independentsamples t-test at 5% significance ($\alpha = 0.05$)
- Serum calcium in babies of untreated women 9.0 mg/100 ml, standard deviation (s) 1.8mg/100ml
- Study should detect clinically relevant increase in serum calcium of 0.5 mg/100ml, 80 per cent of the time ($\beta = 0.2$)

In summary:

- μ = Mean serum calcium level = 9.0 mg/100ml
- σ = Standard Deviation = 1.8mg/100ml
- d = difference in means $m_1 - m_0$ = 0.5mg/100ml
- $\alpha = 0.05$
- $\beta = 0.2$

$$n = 2\frac{(Z_\alpha + Z_\beta)^2\sigma^2}{(\mu_1 - \mu_0)^2}$$

n=205

References

1. University of Wisconsin-Extension, program development and extension
2. http://learningstore.uwex.edu/pdf/G3658-04.pdf
3. Research Methodology by Getu Degu and Tegbar Yigzaw
4. Akobeng AK. Principles of evidence based medicine. Archives of Disease in Childhood 2005; 90: 837-840.
5. https://extension.usu.edu/