Top Statistical Tips for Embarking on a Clinical Research Project

Presenters : Raja Dhungana & Darren Rajit

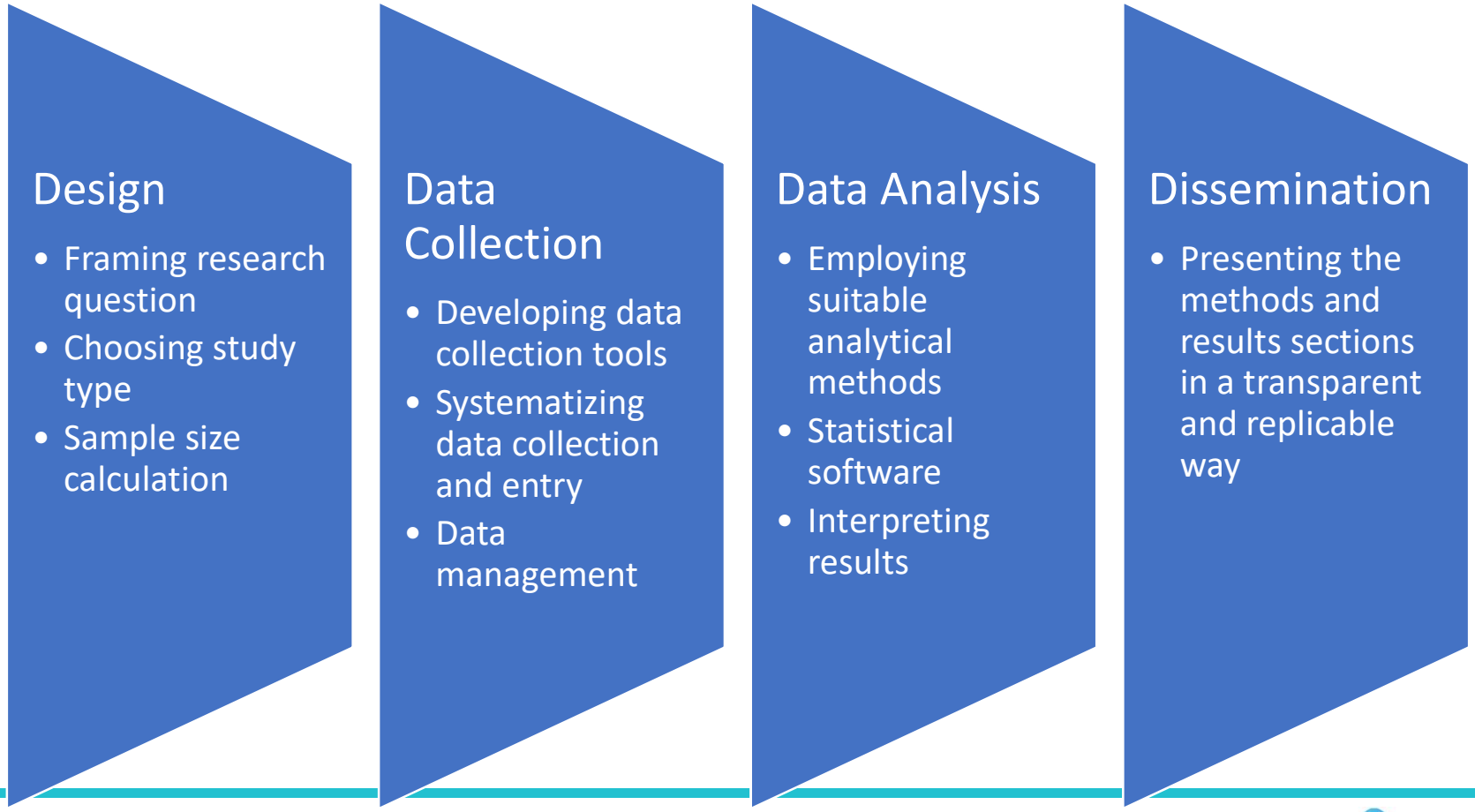
Monash centre for Health Research and implementation, Monash University
✉ info.mchri@monash.edu / joanne.enticott@monash.edu
✆ +61 3 8572 2600
https://mchri.org.au/research-support-services/biostatistics/

MCHRI Monash Centre for Health Research and Implementation
MONASH University
Monash Health

Aims

- Emphasize the value of integrating statisticians early and thoroughly in clinical research.
- Explore key statistical considerations, from design to interpretation, using a case study.
- Identify and address common statistical pitfalls in clinical research.
- How to reach out to us?

When to Bring Statisticians In?



MCHRI
Monash Centre for Health Research and Implementation

What's in a Minimum Viable Study?

- Research Question
  - PICO and variants –
    - be clear on focus and what you wish to investigate
- Study Type
- Statisticians can help below:
  - Sample Size & Power
  - Sampling and data collection strategy,
  - Outcome Measures
  - Analysis Plan

SELECT A FRAMEWORK BY QUESTION FOCUS

Question frameworks are usually referred to by their acronyms. Click the acronym in the table to go to the page with the full explanation for that framework.

Focus of your question	Associated frameworks
Clinical questions about the effectiveness of interventions / treatments, and the impact of exposures.	• PECO (PECOT, PECODR, PEO) • PerSPECTIF • PICO (PIO, PICOC, PICOS, PICOT, PICOTS, PICOTT)
Diagnostic test evaluation	• PICO for diagnostic tests
Economic evaluation / cost effectiveness	• PICOC
Evaluating experiences of a specific phenomenon	• CHIP • PEO • PICo • SPICE • SPIDER
Policy evaluation	• CIMO (CIMOS, CIMOT) • CLIP • ECLIPSE • PIFT • SPICE
Practice guideline evaluation	• PIPOH (PIPOS)
Prevalence of a condition incidence of a condition, disease, symptom, health condition, problem	• CoCoPop
Prognosis issues / Determining prognosis	• PFO
Questions about a client's welfare (arising from daily practice)	• COPES
Questions about complex interventions	• ProPheT
Questions about rehabilitation therapies (in speech pathology, occupational therapy, physiotherapy, ...)	• PESICO
Service evaluation / improvements	• CLIP • ECLIPSE • PICOC • SPICE
Theories / methodologies	• BeHEMoTh

Source : https://libguides.library.cqu.edu.au/c.php?g=949210&p=6880841#s-lg-box-22084512

MCHRI
Monash Centre for Health Research and Implementation

Research Question to Study Design

A Few Statistical Notes

Study Design	Research Questions	Study Type	Example Statistical Considerations
Randomized Controlled Trials (RCTs)	Efficacy or effectiveness of an intervention, comparison of two or more treatments	Parallel, factorial, crossover, cluster etc.	Randomization, controlling for confounding variables, sample size calculations, intention to treat vs per-protocol analysis
Observational Studies	Risk factors for disease, prognosis of disease	Cohort studies	Selection of exposed and non-exposed groups, controlling for confounding variables, relative risk calculations, survival analysis
	Risk factors for rare diseases, identifying causes of an outbreak	Case-control studies	Selection of cases and controls, matching, odds ratio calculations
	Prevalence of a condition, associations between variables	Cross-sectional studies	Survey design, survey weight calculation, prevalence calculations, association between variables

Sample Size & Power

- Trade offs Between
  - sample size,
  - power,
  - Desired effect size, and
  - cost

$$N = 2 \times \left(\frac{z_{1-\frac{\alpha}{2}} + Z_{1-\beta}}{\delta} \right)^2 \times s^2$$

Type I and Type II error		
Null hypothesis	**True**	**False**
Rejected	Type I error (α) (False positive)	Correct decision (1-β)
Not rejected	Correct decision	Type II error (β) (False negative)

Power = 1 - β

MCHRI
Monash Centre for Health Research and Implementation

Sample Size & Power: an example

Outcome: systolic blood pressure

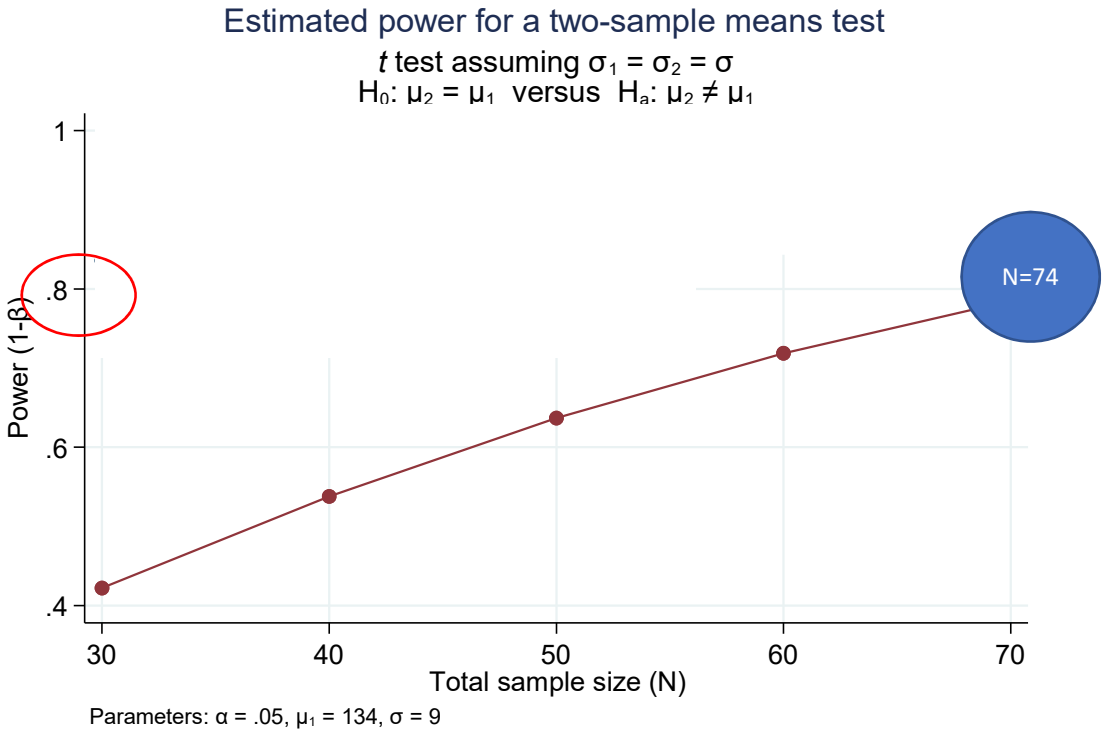
Population or control group mean (μ): 134 mmHg

Scenario 1

Experimental group means (μ2): 128 mmHg

Scenario 2

Experimental group means (μ2): 124 mmHg



MCHRI
Monash Centre for Health Research and Implementation

Data Collection

- Type of data collected (Nominal, ordinal, continuous etc.) will influence the statistical tests

- Clear protocols on collecting, defining and organizing the data will minimize potential bias, missing data, data entry and statistical errors

Statistical Analysis

Outcome variables					
Nominal	**Categorical**	**Ordinal**	**Quant. Discrete**	**Quant. Non-normal**	**Quant. Normal**
χ^2	χ^2	χ^2	Mann-Whitney U test	Mann-Whitney U test	Student's t test
Logistic regression	Logistic regression	Kruskal-Wallis test	Kruskal-Wallis test	Kruskal-Wallis test	ANOVA
	Poisson regression	Spearman rank test	Spearman rank test	Spearman rank test	Pearson and linear regression
	Other advanced techniques	Other advanced techniques	Other advanced techniques	Linear regression and other advanced technique	Other advanced techniques

Source: https://www.bmj.com/about-bmj/resources-readers/publications/statistics-square-one/13-study-design-and-choosing-statisti

Organizing data

- be consistent (variable name and data),
- put just one thing in a cell,
- organize the data as a single rectangle (with subjects as rows and variables as columns, and with a single header row),
- create a data dictionary

Organizing data: An example

Example of Excel data that is *unsuitable* for analysis

URN	Date Of Birth	Patient Age	Gender	start date	Current Smokers	NYHA	Systolic	Blood Pressure Pre	Blood Pressure Post		Marital status
9722-1	12/05/63	41YRS	Male	19/07/04	No	I	120	115	75		1
0651312	14/09/26	78	F	26/01/04	No	n/r	n/a	=90	50		2
0454545	7/12/33	70	M	n/a	N	II	.	140	70		3
0001111	21/05/35	69	m	n/a	N	III	?	130	80		2
0011111	5/02/44	60 +3months	F	29/07/04	N	II	<90	140	80		3
0106574	10/11/36	67	F	2/01/04	N	II	70 (under	120	70		1
1066329	19/09/46	58	f	n/a	N	III	>170	170	100		3
0537720	1/09/51	53	F	n/a	Y	II	115	120	80		2

MCHRI
Monash Centre for Health Research and Implementation

Contents for data dictionary

- The exact variable name as in the data file
- A version of the variable name that might be used in data analysis
- A longer explanation of what the variable means
- The measurement units
- Expected minimum and maximum values

Data dictionary: Examples

1.5 Gender

Definition	Gender is the biological distinction between male and female		
Database Name	Gender	**Collection**	Mandatory
Data type	Numeric	**Form**	Code
Field size	1	**Layout**	N
Code set	tlkpGender (reference table)		

Code	Description
1	Male
2	Female
3	Intersex or indeterminate
9	Not stated/inadequately described

Reporting guide	Gender should be retrieved from the hospital administrative dataset to ensure consistency with data collection. Gender is what the person considers themselves to be irrespective of anatomy or birth, it is usually unnecessary and may be inappropriate to ask a person their sex. Sex may be inferred from other cues such as observation, relationship to respondent, or first name. The term 'intersex' refers to a person, who, because of a genetic condition was born with reproductive organs or sex chromosomes that are not exclusively male or female and who identifies as being neither male nor female. Excludes: transgender, transsexual and chromosomally indeterminate individuals who identify with a particular sex (male or female).
Purpose	Patient identification. Service utilisation and epidemiological studies
Data Users	Data Collectors, BRANZ Staff, Reporting, Epidemiologists
Collection start	Jul-09 Note: Migrated data dates back to Jul-05, however not used in BRANZ reporting
Definition source	National Health Data Dictionary (NHDD)
Code set source	NHDD

Common Pitfalls

- Multiple testing
  - The more hypothesis tests you conduct in a study, the higher the chance of a false positive somewhere
  - Addressed via: Post hoc statistical corrections
- Bias
  - Multiple flavors exist that can affect external + internal validity of findings
  - Addressed via: Study design improvements
- Missing data (eg: participant drop out)
  - Depends on reason for missing data (eg: MCAR, MAR, MNAR) but can bias results
  - Addressed via : imputation methods, but highly contextual

Common Pitfalls

- Inadequate power
  - Power = study ability to detect a real effect. Low sample size generally = lower power
  - Addressed via: Increasing sample size ($$) or bootstrapping
- Ignoring confounders
  - Presence of a variable that effects both variables you're investigating -> Spurious association
  - Addressed via: Post hoc statistical adjustment is possible but we don't know what we don't know! Subject matter expertise important. Causal Loop Diagrams or DAGs can be helpful
- Table 2 Fallacy
  - When causal conclusions are drawn based on adjusted associations. Interpretation is important, statisticians can help

Engaging with Statistician – How to get the most bang from your buck

- **Research Question** – Most important !
- Study Design & Data organizing
- **Data Dictionary** – Most important !
- Initial ideas on Sample size
  - From your clinical perspective, what is the smallest but meaningful effect size? We can go from there
- Initial ideas on Analysis Plans –
  - Brainstorm with us
- Potential issues – Practical constraints that you know will influence the project
- Timeline

MCHRI
Monash Centre for Health Research and Implementation

How to reach out to us?

- Biostatistical Consulting Service at Monash Centre for Health Research and Implementation (MCHRI), under Monash Health and the Monash University
- Services
  - General advice on study design and grant writing
  - Advice on statistical rigour and sample size
  - Assistance with statistical analysis.
- Cost
  - The initial consultation (up to two hours) is free
  - Additional cost $125 per hour
  - Costs for small projects are generally quoted up to $10,000; medium projects up to $25,000; large projects over $25,000

MCHRI
Monash Centre for Health Research and Implementation

Resources

- Research Question Frameworks : https://libguides.library.cqu.edu.au/c.php?g=949210&p=6880841#s-lg-box-22084512
- Data Dictionary Course at Monash University : https://www.monash.edu/data-fluency/toolkit/health-research-data-dictionary
- Catalog of Bias : https://catalogofbias.org/
- Different Study Designs : https://www.cebm.ox.ac.uk/resources/ebm-tools/study-designs
- Sample Size and Power : https://emj.bmj.com/content/20/5/453