Study Design

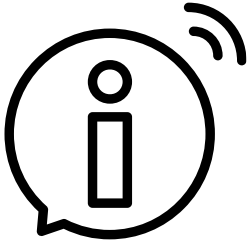
Presented by Dr Ali Green
Statistical Consulting Unit
Sydney Informatics Hub
Core Research Facilities

sydney.edu.au/sydney-informatics-hub

THE UNIVERSITY OF SYDNEY

Slides available here

Acknowledging SIH



- All University of Sydney resources are available to researchers free of charge. The use of the SIH services including the Artemis HPC and associated support and training warrants acknowledgement in any publications, conference proceedings or posters describing work facilitated by these services.

- The continued acknowledgment of the use of SIH facilities ensures the sustainability of our services.

Suggested wording for use of workshops and workflows:

- *“The authors acknowledge the Statistical workshops and workflows provided by the Sydney Informatics Hub, a Core Research Facility of the University of Sydney.”*

What is a workflow?

- Every statistical analysis is different, but all follow similar paths. It can be useful to know what these paths are.
- We have developed practical, step-by-step instructions that we call ‘workflows’, that can you can follow and apply to your research.
- We have a **general research workflow** that you can follow from hypothesis generation to publication.
- And **statistical workflows** that focus on each major step along the way (e.g. experimental design, power calculation, model building, analysis using linear models/survival/multivariate/survey methods).

Statistical workflows

- Our **statistical workflows** can be found within our workshop slides.
- **Statistical workflows** are software agnostic, in that they can be applied using any statistical software.
- To access these statistical workflows and more, visit our Workshops and Workflows page.
- The broader aim of our workshop training is to increase understanding and application of statistical concepts, to facilitate higher impact research.

Software workflows

- There may also be accompanying **software workflows** that show you how to perform the statistical workflow using particular software packages (e.g. R or SPSS). We won't be going through these in detail during the workshop. If you are having trouble using them, we suggest you attend our monthly Hacky Hour where SIH staff can help you.

- Our software workflows contain:
  - R code and comments.
  - SPSS syntax as well as screenshots of the point and click procedures and written methods.
  - Screenshots of the point and click procedures and written methods for other bespoke software.

Using Artificial Intelligence (AI) responsibly

- USyd supports the responsible use of AI in research, aiming to balance innovation with academic integrity. USyd publishes its AI guardrails and endorsed AI tools, and currently allows Microsoft Copilot and Cogniti to be used with public and protected data. Highly protected data must never be entered into these tools.

- The Statistical Consulting Unit (SCU) at SIH also supports AI use, but like with any tool, it should be used responsibly. The SCU recommends:
    - Being careful with handling sensitive information.
    - Verifying the accuracy of all AI-generated content.
    - Ensuring the output generated is not false, misleading, or misrepresenting findings.

- This aligns with the Statistical Society of Australia's statement on the use of generative AI in statistics and data science.

- For further AI training, check out the series of GenAI for Research workshops hosted by SIH's Data Science team.

Use AI to enhance, not replace statistical collaboration – it cannot match the nuance of a statistician who understands the research question, study design and constraints.

Using AI responsibly: Tips and tricks for statistical analysis

Ask detailed questions: Align each prompt to a specific step in a statistical workflow of your choice. E.g. for linear models, see below:

Be specific: State your outcome, predictors and how they are measured (numerically or categorically).

Protect your data: Avoid uploading data into AI tools. Instead, describe data structure, or upload simulated data or dummy tables.

Validate code: Check the code is statistically correct and using up to date packages and syntax. Understand what code does before running it.

Think critically: Always review, question and verify any content produced by AI.

Model Fitting Workflow

Step 0) Clean and check data.

Step 1) Pick a suitable model to fit to the data via Exploratory Data Analysis (EDA).

Step 2) Fit the Model.

Step 3) Check Model Assumptions via Diagnostics: Residual Analysis.

Step 4) Goodness of Fit: Plots and Statistics.

Step 5) Interpret Model Parameters and reach a conclusion.

Step 6) Reporting.

Understanding: Before relying on AI, make sure you understand the statistical foundations yourself. Or you won't know when it's wrong.

Check assumptions: AI rarely checks statistical assumptions unless explicitly asked. Check model diagnostic plots.

Statistical method should match study design: E.g. is your model accounting for clustering or repeated measures?

Why, not just what: Ask AI to explain why certain methods are appropriate for your data.

Using AI responsibly: Tips and tricks for R coding

Tip 1: Use **headings and subheadings,** as well as the **document outline** in R Studio to break down the problem into a workflow and think systematically.

- The example below uses subheadings to outline a workflow for data cleaning and processing.
- Breaking the problem into smaller parts leads to simpler questions for AI, that are easier to check.

```
# ... ----
# 3 DATA CHECKING AND CLEANING ----
## 3.1 Unique Identifier ----
# Expecting a unique identifier in the data export.
# We should not have duplicate values as there is one response per respondent.

unique.id <- combined_all # creating a save point!
names(unique.id) # check for the correct variable name and update the code below

# CHECK: the unique identifier is unique
table(duplicated(unique.id$UID))
```

0 Setup
1 Import Data
  1.1 Read Data
  1.2 Import checking
2 Combining data files
  2.1 Joining
  2.2 Bind rows
3 Data Checking and Cleaning
  3.1 Unique Identifier
  3.2 Data of Interest
  3.3 Unexpected characters
  3.4 Convert to numeric
  3.5 Create Factors
  3.6 Check Factor Levels
  3.7 Missing Data
  3.8 Duplicated Data
  3.9 Flatliners
  3.10 Outliers
  3.11 Final clean data
4 Create variables
  4.1 Top box

Tip 2: Under each heading/subheading write out the **purpose of the section**. Include AI prompts, as well as ideas for later (coding tends to be iterative).

Using AI responsibly: Tips and tricks for R coding

Tip 3: Every section has a **check**!

- This can be as simple as just having a look at the output, dimensions, or crosstabulations.
- Keep in mind you are responsible for your work and need to ensure your code does exactly what you want it to do. This is especially important when using AI as it can hallucinate!

E.g. Should the number of rows change after joining a second data file (number of participants)?

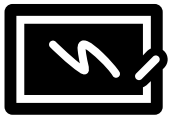
```
> # CHECK: no additional rows
> nrow(raw_data_metro) # original n = 100
[1] 100
> nrow(combined_metro) # combined file should be the same!
[1] 100
```

E.g. Did all values get recoded correctly (were any NAs generated)?

```
> # CHECK: crosstab of original variable and new numeric variable
> # Looks good
> convert.numeric |>
+   count(Q3, Q3_numeric) |>
+   arrange(Q3_numeric)
                        Q3 Q3_numeric  n
1         Strongly Disagree          1 15
2                  Disagree          2 45
3 Neither Agree or Disagree          3 65
4                     Agree          4 42
5            Strongly Agree          5 27
```

During the workshop

- Ask short questions or clarifications during the workshop (either by Zoom chat or verbally). There will be breaks during the workshop for longer questions.



- Slides with this blackboard icon are mainly for your reference, and the material will not be discussed during the workshop.

- Challenge questions will be encountered throughout the workshop.

This workshop

1. What is study design and why is it so important?

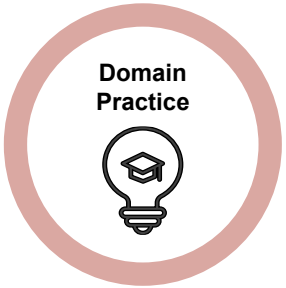
2. A workflow for study design:
    1. Step 1: Defining your sample
    2. Step 2: Considering your analysis
    3. Step 3: Designing your study

Themes of this workshop

Internal Validity
Finalising your research question, types of study designs, bias, error, confounding, dropouts, randomisation, blinding.

External Validity
Generalisability of findings, ensuring you have a representative sample, and real-world vs. controlled settings.

Replication and Sample Size
Sample size and statistical power for reducing random error and improving reproducibility, ensuring model stability and robustness.



Domain Practice
Common study designs in your field, common analysis techniques, best practice reporting guidelines and standardised methods.

Feasibility and Impact
Ethical, logistical and financial considerations that shape your study design, social considerations and translating your research into practice.

What is study design?

And why is it so important?

"To call in the statistician after the experiment is done may be no more than asking him to perform a post-mortem examination: he may be able to say what the experiment died of".

Ronald Fisher
Godfather of Statistics

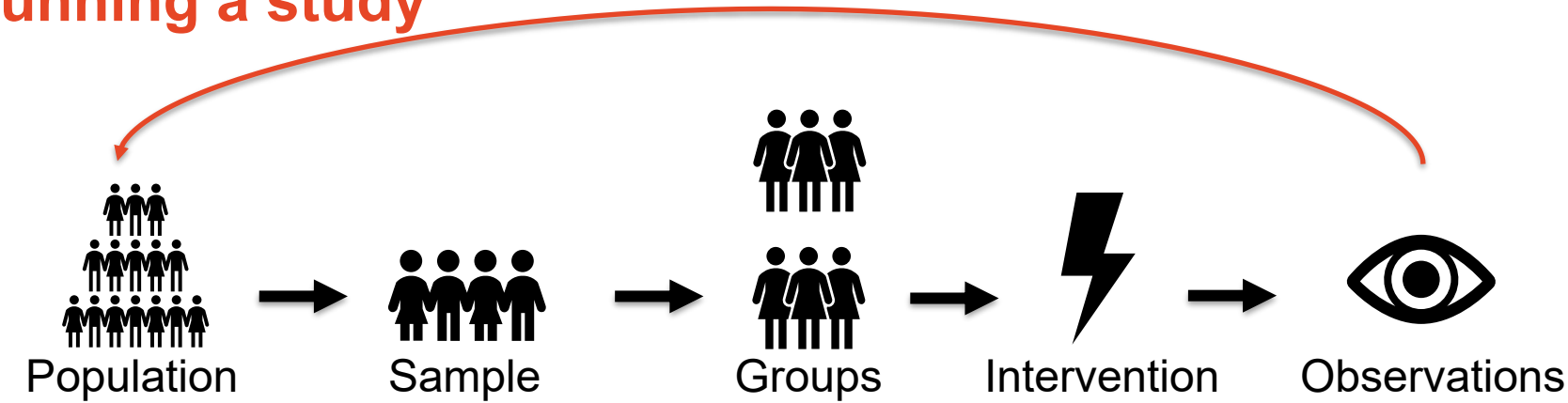
What is study design and why is it so important?

- Study design involves the methods and procedures used to plan, collect and analyse data in your research. It is the framework for systematically investigating a research question or hypothesis, ensuring that the data are reliable, valid and appropriate for drawing conclusions.

- By thinking about study design, you can ensure that you maximise the validity, reliability and reproducibility of your findings, whilst minimising bias.

Reproducibility isn't just a challenge, it's an opportunity to improve the quality of how we do research!

Running a study



- We want to find out (infer) something about the members of our population so we run a study.

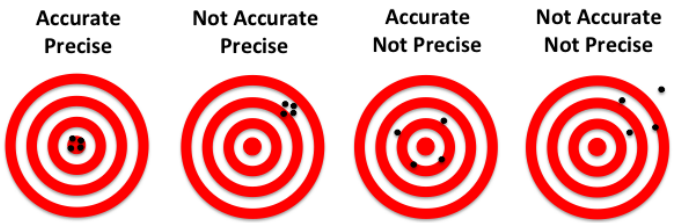
- Despite the labour involved, or perhaps because of it, many do not spend enough time thinking about designing a study so that it has the best chance of finding something out (valid statistical inference – the red arrow above).

- Study design is about being aware of the challenges to statistical inference and using the best tools available to overcome them.

"If you cook a large pan of soup, you don't need to eat it all to find out if it needs more seasoning, you can just taste a spoonful, provided you have given it a good stir".

George Gallup
Pioneer of survey sampling

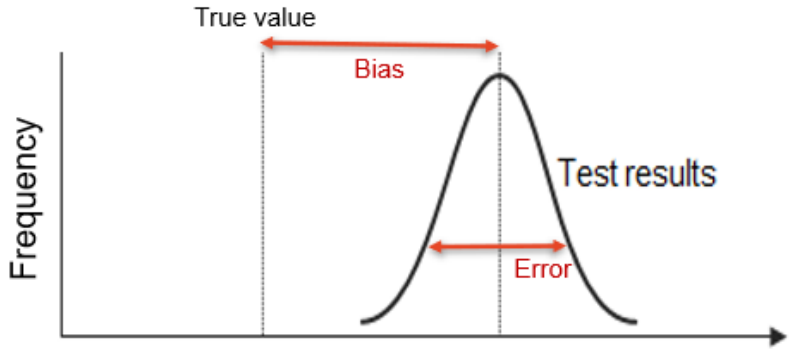
The role of bias and error in study design



Just like throwing a dart, using our experimental sample to estimate some population statistic is subject to randomness. Unlike throwing a dart – we don't know where the bullseye is!

- When we use a sample to estimate a population value, our estimate will generally differ from the true value.
- This difference can arise from:
  - **Bias (accuracy):** a systematic tendency to over- or under-estimate the true value.
  - **Error (precision):** random variation that causes estimates to differ from one sample to another.
- A useful analogy is a dartboard:
  - The bullseye represents the true population value.
  - Each dart represents an estimate from a study.
  - Bias determines how close the average position of the darts is to the bullseye.
  - Error determines how tightly the darts cluster together.
- Good study design aims to minimise bias and reduce random error so that estimates are as close to the true value as possible.

Why are our estimates biased?



If we ran our experiment many times, and plotted the frequency of our estimates, our best guess of the true value would be the mean of all of our estimates. The distance from this best guess to the true value is our bias. No matter how many times we ran the experiment, or how large our sample was, we would end up with this level of bias.

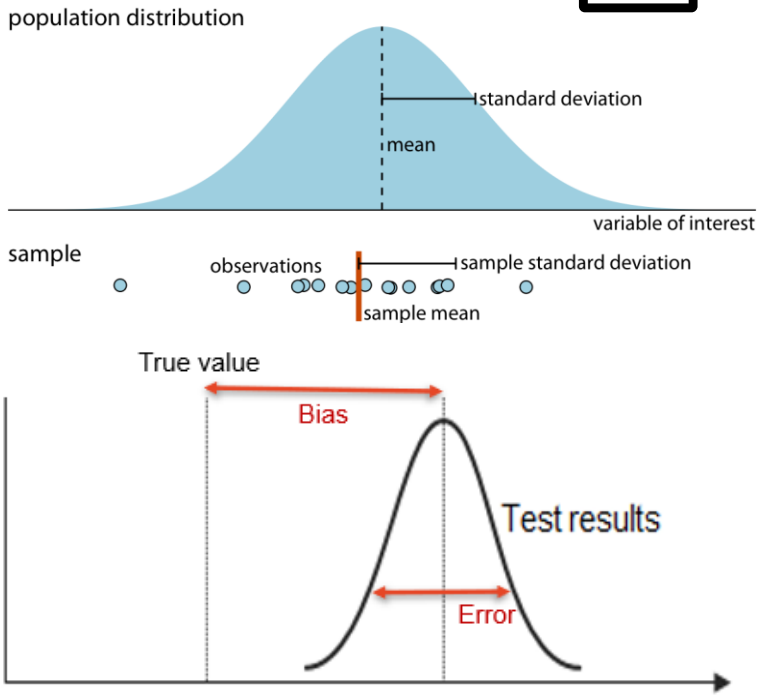
- Let's start by thinking about bias in a statistical way: it is *systematically* over- or under-estimating our population property.

- Bias may be introduced in the way we **sample**, **conduct our experiment, measure,** or even **analyse** our data.

- Some level of bias is inevitable, but too much bias can be fatal: we would reach the wrong conclusion *most of the time* in a poorly designed experiment!

Why are our estimates imprecise?

- ‘Error’ in statistical terms is about how close together our estimates are – it’s ‘error’ in the sense of measurement, not making a mistake.
- Error exists because the individuals in our population vary. Each sample is different, and so our estimates will vary from study to study*.
- Again, error is influenced by the way we sample, measure, conduct or analyse our experiment.
- We can increase the precision of our estimates by having larger samples. Choosing an appropriate sample size is a vital and non-trivial part of controlling error in experimental design: see our Power and Sample Size Calculation workshop for more.



It’s important to realise that although variability in the population affects the variability of your estimates – there’s another important ingredient: your sample size. We use variation in our sample population to quantify the uncertainty in our estimates.

** This is called ‘sampling error’, but there is also ‘non-sampling’ error which covers everything else that causes variation between estimates, e.g. mistakes when collecting data.*

Some common types of bias

Selection bias

Recall bias

Design bias

Sample bias

Reporting bias

Confounding bias

Attrition bias

Publication bias

Observer bias

Measurement bias

The 1936 Literary Digest scandal

Literary Digest magazine predicted a landslide victory for Republican Alf Landen, however, Roosevelt won in one of the most decisive victories in U.S. history.

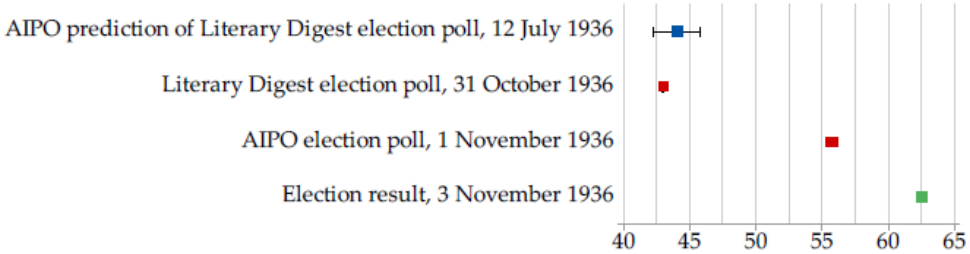
Literary Digest

- Mailed out ballot cards to obtain responses (upward of 10 million people).
- Response rates of 11.8% to 23.8%.
- Predicted that Landon would receive 57.1% of votes for the two candidates.
- Results described as “exactly as received from more than one in every five voters polled in our country... neither weighted, adjusted, nor interpreted”.

George Gallup and the American Institute of Public Opinion

- Polls conducted with mail outs and face-to-face interviews.
- Gallup’s final poll predicted Roosevelt to have 55.7% of major party votes. It also predicted that the Literary Digest would predict Landon to win.
- Gallup’s poll was out by only 6.8%, whereas Literary Digest was out by over 19%.

What went wrong?



Assumed sample sizes: AIPO prediction of Literary Digest: 3,000; AIPO election poll: 50,000.

Figure 1: Predictions of the 1936 US Presidential election result with approximate 95% confidence intervals

Source: Thinking critically about the 1936 US Presidential Election Polls, Sue Finch and Ian Gordon, University of Melbourne.

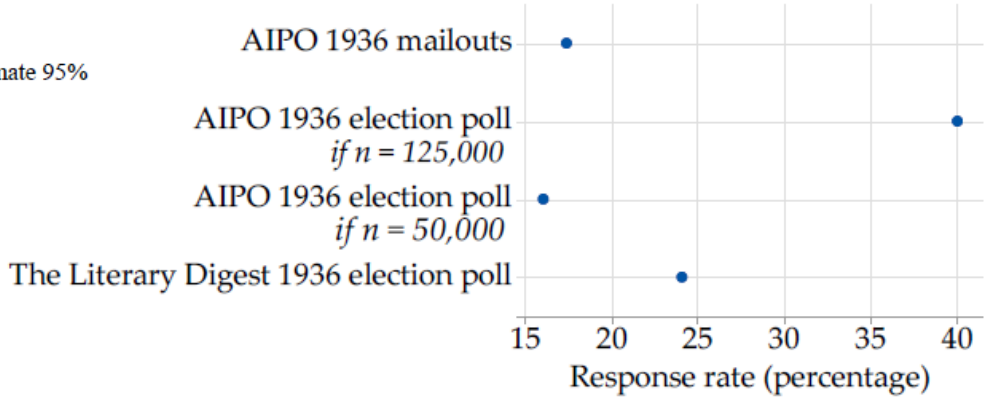


Figure 2: Response rates for various 1936 US polls

Key takeaways from the 1936 poll

Methodological flaws:

- **Sampling/selection bias:** Surveyed via car registrations and telephone directories, excluding many voters during the Great Depression. Disproportionately sampled wealthier Americans who don't reflect the target population (all Americans).
- **Nonresponse bias:** Only 25% of recipients responded, with Roosevelt voters less inclined to respond. Gallup used a quota system to ensure his poll contained voters of different demographics.
- **Bigger is not always better –** large samples don't guarantee accuracy if they are not representative! This poll relied on telephone directories and automobile registrations, excluding lower income voters who would have voted for Roosevelt.

A quick challenge question

Would you be *most* concerned about bias, error, or both equally in the following studies?

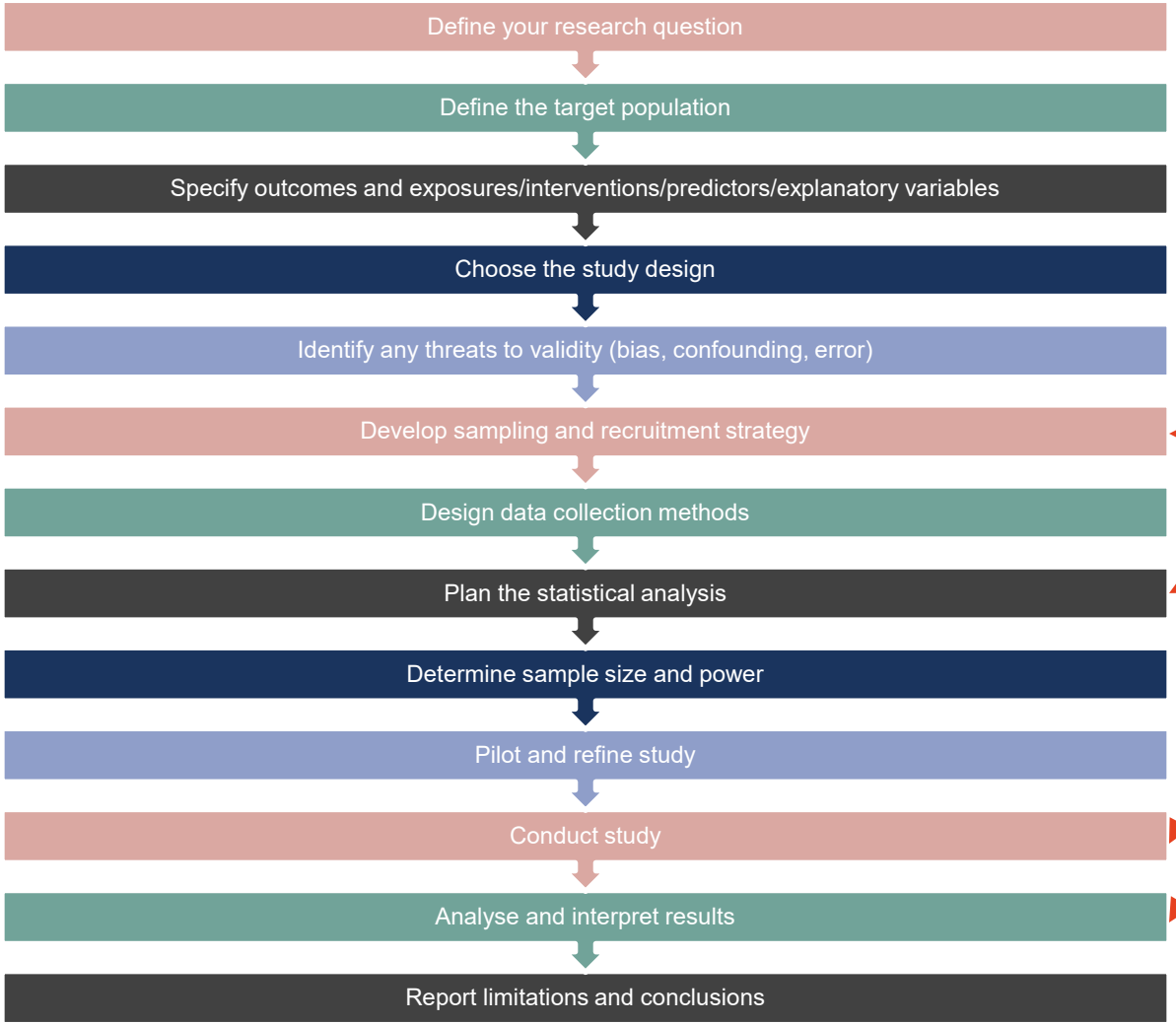
- A study on the effectiveness of a drug in killing cancer cells, where the cells were counted by a technician who knew whether each sample was drug or vehicle (no drug) treated.

- A study on the effectiveness of a drug in killing cancer cells, where the drug was administered by a technician who was measuring the dose using a Pasteur pipette rather than a micropipette.

- A survey on the number of motor vehicle crashes experienced over your lifetime where you, the respondent was at fault.

- As above, where either party were at fault.

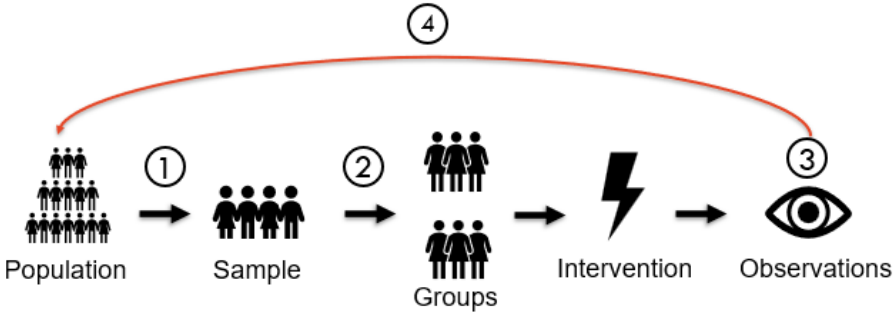
A workflow for study design



Comprehensive workflow for study design

We will now home in on these steps!

A workflow for study design

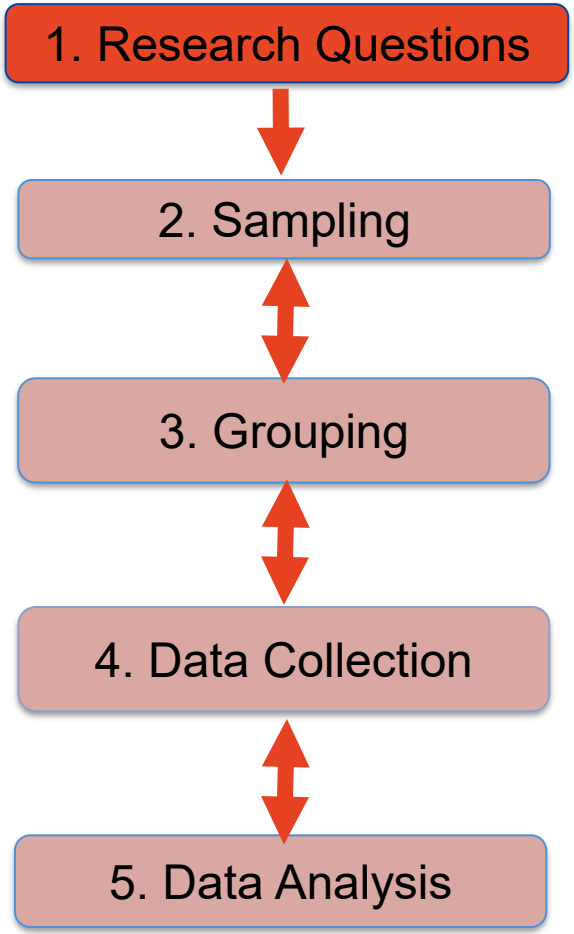


- We need to consider the steps of study design in the context of how a study is performed.

- The process of experimental design itself should be iterative, so after you have considered the following, cycle back through these steps until you are convinced the design is the best it can be.
    I. Sampling
    II. Grouping
    III. Data collection
    IV. Analysis

A workflow for study design

1. Finalise your research questions and hypotheses then choose your study type.
2. Sampling
   - Define your sampling method
   - Consider representativeness
   - Identify your units
   - Consider replication
3. Grouping
   - Choose your blocks or strata
   - Perform randomisation for treatment allocation
4. Data Collection
   - Consider blinding
5. Data Analysis/Inferential Analysis
   - Identify your method



Study design template

- After attending this workshop (and potentially other stats consulting workshops) you should be able to fill in all these boxes as they apply to your research.

- You can bring this information to a statistical consult and/or share it with your research team.

Biological Units	
Experimental Units	
Observational Units	

Outcome Variables	Explanatory Variables	Design features	Blocking Factors	Randomisation & blinding	Analysis Method

Step 1: Finalise your research question and choose your study type

Formulating your research question

- Your research question should clearly detail the problem that is being addressed in your study. It should be a focused and answerable question that guides the content of your study.

- Start by identifying a gap in existing knowledge, and then refine your research question to be specific, measurable and relevant to your field.

- Depending on your research field, there may be a specific framework to follow when developing your research question. E.g. the PICO framework which helps structure clinical research questions by breaking them down into: Population, Intervention, Comparison and Outcome.

Write a research project plan

Back to the basics: guidance for formulating good research questions

What's Your Problem? Writing Effective Research Questions for Quality Publications

Terminology of study design

Smoking

Predictor
Explanatory variable
~~Independent variable~~

Lung disease

Response
Outcome
Dependent variable

- We introduced the different types of variables in Research Essentials.
- In the language of experimental design, explanatory variables are often referred to as **factors** (e.g. confounding factors).
- The value of the factor for an individual (e.g. smoker, non-smokers) are referred to as levels of the factor.
- When we manipulate a factor to test its effect, it becomes a **treatment** variable.
- Other explanatory variables are referred to as design variables (often including blocking variables).

Step 1: Finalise your research question(s)

What are your research questions? Having a very clear idea of what your research questions is vital for achieving good experimental design.

Avoid the urge to rush into collecting data and just “worry about the analysis later.”

You can only get meaningful results from a well-controlled and adequately-powered experiment. The type of analyses possible are also dictated by your experimental design.

Think of this paradox: often the sooner you start your experiments, the longer your project will take. Invest the time in good design.

Step 1: Finalise your research question(s)

- If you are going to use regression analysis, how will you frame your research question? As descriptive, predictive or causal? Think about what you would like to learn from the data generated and what the aim of your model is.

- Need to get the most precise estimates.

- Need to consider the level of subject matter knowledge/theory:
    - **Descriptive:** Fitting different models to create hypotheses about associations (weak theory).
    - **Predictive:** Getting most accurate predictions of future observations.
        - Interpretable – test hypotheses about predictors that may be modifiable;
        - Non-interpretable - if interested in prediction only, may use some Machine Learning approaches without making assumptions/interpretations.
    - **Causal:** Fitting an *a priori* defined theoretical model – needs strong theory e.g., controlled experiment – aim is hypothesis testing/causal inference.

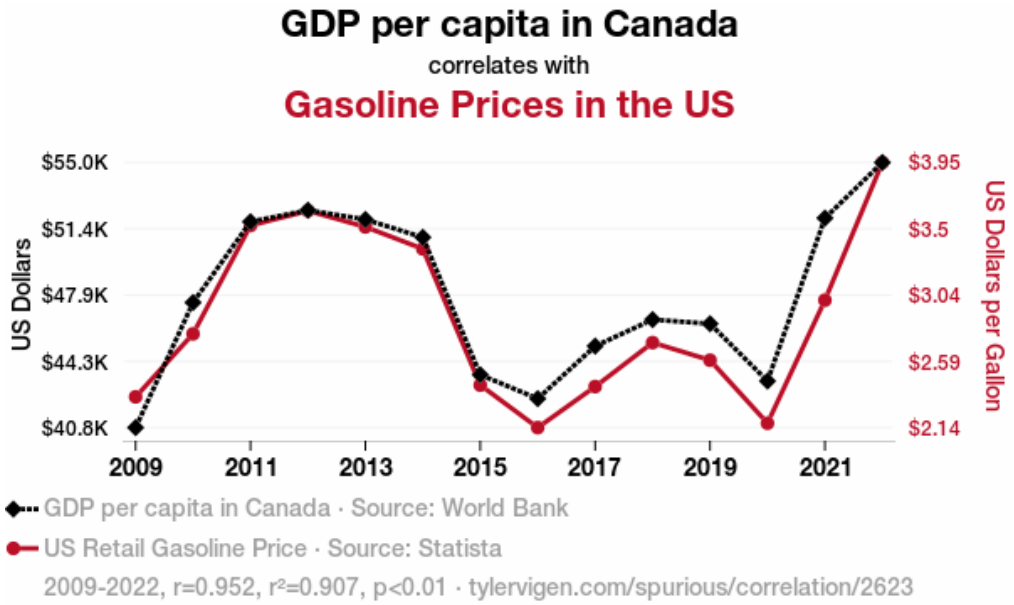
- Common pitfalls: Using association models to imply causality without justification.

Step 1: Finalise your research question(s)

- On the Uses and Abuses of Regression Models: A Call for Reform of Statistical Practice and Teaching.

Type	Goal	Example	Key considerations
Descriptive	Summarise or characterise data or describe participants with no intention to draw inference about a broader population.	Studies that examine the prevalence or incidence of a condition in a population or across subgroups of a population.	Often the easiest to perform, analysis plan should address issues of sampling and measurement bias. Likely uses univariable and multivariable regression. Data visualisation of outcome distribution can guide the summary measure to use.
Predictive	Develop an algorithm to forecast an outcome.	Developing an algorithm to predict the risk of dying in a population of patients with a particular disease, given a set of available covariates. Often concern binary outcomes.	Modern approaches include resampling methods and shrinkage estimation for internal validation. Should consider external validation to assess generalisability. May involve machine learning (ML) techniques though these may not be superior to regression methods in clinical prediction settings. See the TRIPOD guidelines.
Causal	Understand the causal effect of one variable on another. “What if..” type research questions.	Investigating the extent to which health outcomes in a population would be different if a particular intervention were made.	Ideally answered by randomised experiments, but observational studies can be used with limitations, e.g. 'target trial’ framework could be adopted. Some key considerations include identifiability, effect heterogeneity, collapsibility across strata and the impact of sparse data. Can incorporate ML techniques.

Correlation ≠ Causation



Tyler Vigen: Spurious Correlations

AI interpretation: *As Canadians got richer, they couldn't resist the lure of competitive lumberjack sports. American demand for maple syrup spiked, leading to higher transportation costs for syrup imports, which put pressure on US gasoline prices due to increased tanker usage and syrup-related trade disputes. In the end, it was a sticky situation for everyone except the pancake industry.*

A strong correlation, a significant p-value, and a fancy graph are not enough to establish cause and effect!

Step 1: Finalise your research question(s)

So, you want to infer causality?

- In many studies, we are interested in understanding cause-and-effect relationships. E.g. In health research we may aim to find modifiable factors that can improve health outcomes.

- A **causal relationship** exists when a change in one variable leads to a change in another variable. E.g., a high protein diet causing weight gain.

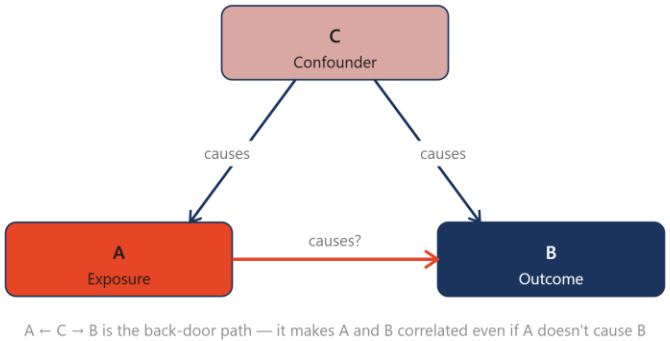
- You should clearly specific the factor you think may influence the outcome (explanatory variable) and the outcome you want to measure (outcome variable).

Explanatory Variable:
Amount of protein supplement provided in a diet intervention

→

Outcome variable:
Weight change 3 months after intervention

Draw a Directed Acyclic Graph (DAG) to examine causality



A ← C → B is the back-door path — it makes A and B correlated even if A doesn't cause B

***Confounding:** A third factor that influences both the exposure and the outcome, potentially leading to a misleading conclusion about cause and effect.

- The best place to start with a potential causal research question is to draw a DAG.
- DAGs show the flow of causal relationships between your variables, including hypothesised outcomes, explanatory variables and potential confounders*.
- They depict your expertise and prior knowledge in the field in a concise, clear way.
- Arrows show the direction of causation. A → B means A causes B, not the other way around. Nodes are connected via the arrows.
- Acyclic – there are no feedback loops, so a variable cannot cause itself, and the graph flows in one direction only.
- Know which variables to include in your model and which to leave out. For more information see our Statistical Model Building workshop.

Causality in Observational vs. Designed experiments

Observational

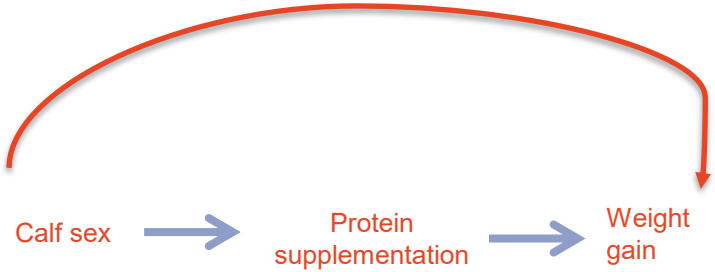
- Difficult to demonstrate causality off a single study.
- Observational studies should be prospectively designed if you want to demonstrate causal inference.
- Causal inference is possible, but heavy reliance on modelling and study design choices, e.g. adjustment (regression, weighting, matching) attempting to recreate randomisation.
- Strong assumptions. Potential outcomes theory – Neyman Rubin model.
- Must adjust for confounding, but there may be potential bias from unmeasured confounders.

Frank Harrell: Causal by Design

Experimental

- In designed experiments, one or more explanatory variables are manipulated by the experimenter. This allows you to evaluate a causal hypothesis (e.g. “diet affects weight gain”).
- Randomisation ensures that treatment assignment is independent of all confounders, observed and unobserved, so the simple difference in means is a valid estimate of the average causal effect.
- Costly and not always ethically possible. Limited generalizability but stronger causal power.

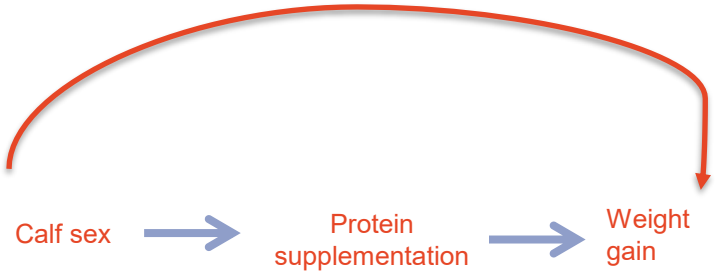
Observational vs. Designed experiments: Confounding variables



* Terminology can be confused by referring to additional explanatory variables that which change the apparent effect of the explanatory variable of interest as “confounders”, without reference to causality (through e.g. drawing a causal DAG). Incorrect adjustment for ‘confounders*’ in our analysis can actually **increase** bias. See our model building workshop for further discussion.

- **Internal validity** is the validity of the experiment for the sample chosen, including proper control of confounding variables.
- To properly consider other variables to reduce bias from confounding, we must consider causal DAGs.
- One example of confounding is where a variable that is not of interest in the study (calf sex) affects both the treatment variable (protein supplementation) and the outcome of interest (weight gain). In our study, we would need to consider the sex of the calves carefully even though we may not be interested in that effect (red arrow).
- When confounding is so strong that the effects cannot be disentangled (e.g. a poorly designed experiment in which *all* male cows receive protein supplement, and all females do not) the study is referred to as **confounded.**

Observational vs. Designed experiments: Confounding variables



Simple randomisation: Assign calves randomly to protein.

Blocking: Use sex as a blocking variable and randomise within blocks.

- We can control for confounders in **designed experiments** by using randomisation. This effectively removes the arrow (blocks the path) between protein supplementation and calf sex.
- We will discuss how to implement randomisation later on in this workshop.

Observational vs. Designed experiments: Confounding variables

This example would typically be examined using a designed experiment but for completion, let's say this was an observational study of farming practices.

Matched pairs: Find pairs of calves with similar characteristics (sex, weight, age) who received or didn't receive protein supplements.

Adjust for differences: Include the sex of the calf as a binary factor in a regression model with protein supplementation as explanatory and weight gain as outcome (note that we may need an interaction term here if we think the protein supplement affects the weight gain of males and females differently).

Subset: Only choose calves within a weight range where males and females overlap and look at the average difference of those with and without protein supplementation.

- In an **observational study,** we don't have an experimental intervention, so we must use another strategy to adjust for confounding:
  - Design our study to sample participants with similar characteristics (e.g. matched pairs).
  - Attempt to adjust for differences in characteristics in our analysis (e.g. a regression model).
  - Find some way to remove the effect of our confounding factor on our outcome by considering subsets of our population.

Assessing causality in epidemiology: revisiting Bradford Hill to incorporate developments in causal thinking

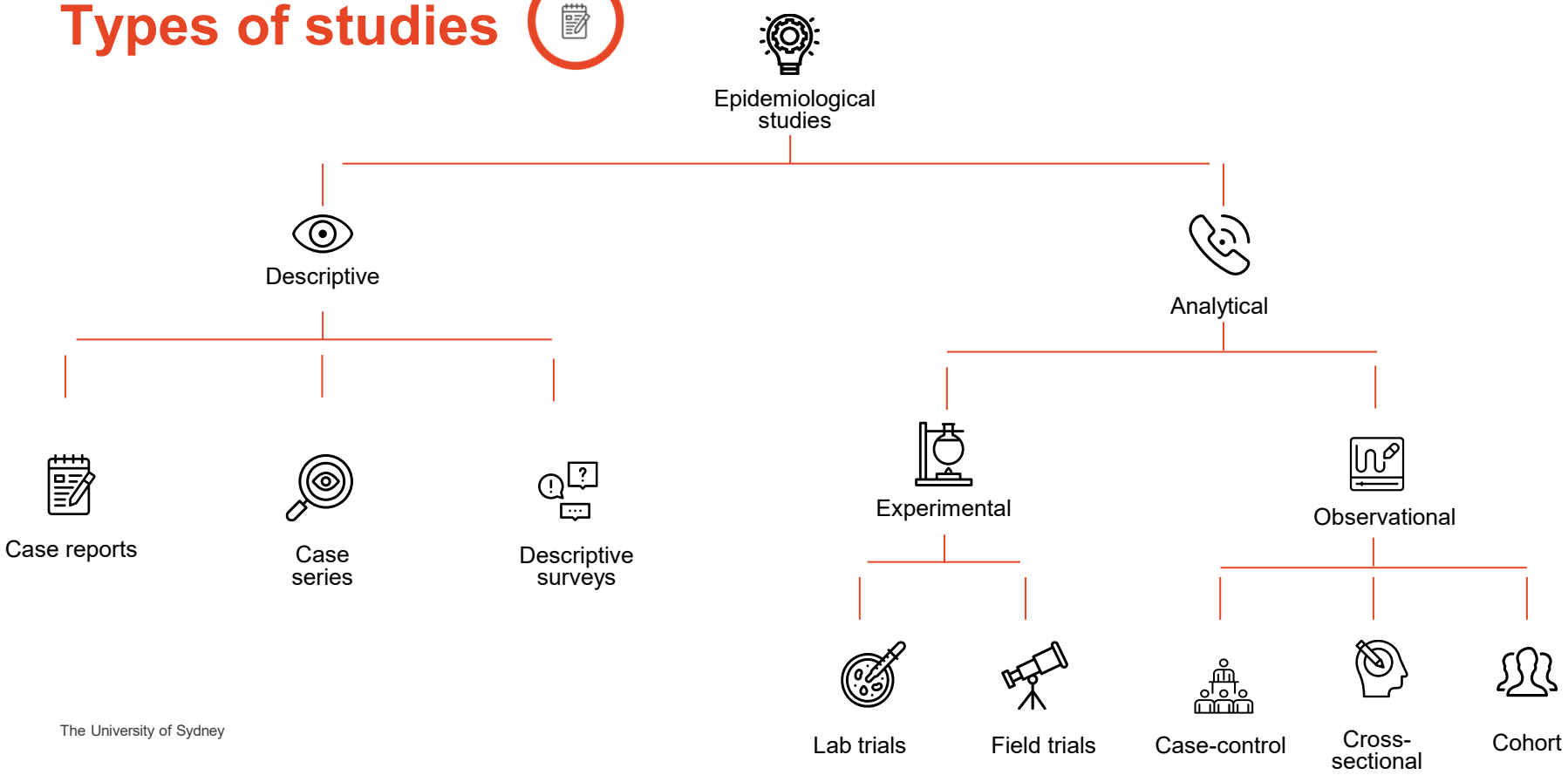
Other lines of evidence to examine causality

Bradford Hill Criteria

1. **Strength (effect size):** A small association does not mean that there is not a causal effect, though the larger the association, the more likely that it is causal.
2. **Consistency (reproducibility):** Consistent findings observed by different persons in different places with different samples strengthens the likelihood of an effect.
3. **Specificity:** Causation is likely if there is a very specific population at a specific site and disease with no other likely explanation. The more specific an association between a factor and an effect is, the bigger the probability of a causal relationship.
4. **Temporality:** The effect has to occur after the cause (and if there is an expected delay between the cause and expected effect, then the effect must occur after that delay).
5. **Biological gradient:** Greater exposure should generally lead to greater incidence of the effect. However, in some cases, the mere presence of the factor can trigger the effect. In other cases, an inverse proportion is observed: greater exposure leads to lower incidence.
6. **Plausibility:** A plausible mechanism between cause and effect is helpful (but Hill noted that knowledge of the mechanism is limited by current knowledge).
7. **Coherence:** Coherence between epidemiological and laboratory findings increases the likelihood of an effect. However, Hill noted that "... lack of such [laboratory] evidence cannot nullify the epidemiological effect on associations".
8. **Experiment:** "Occasionally it is possible to appeal to experimental evidence".
9. **Analogy:** The effect of similar factors may be considered.

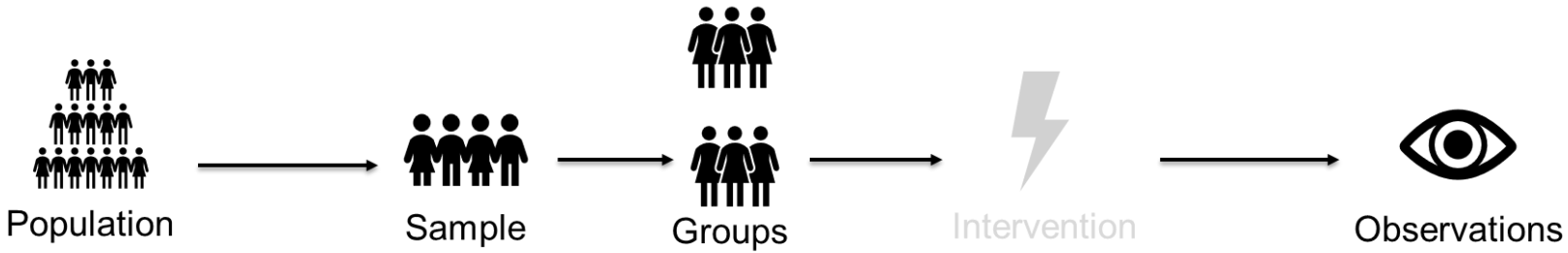
Exposure-wide epidemiology: revisiting Bradford Hill

Types of studies



Observational studies

- **Observational studies** do not involve any experimental intervention.

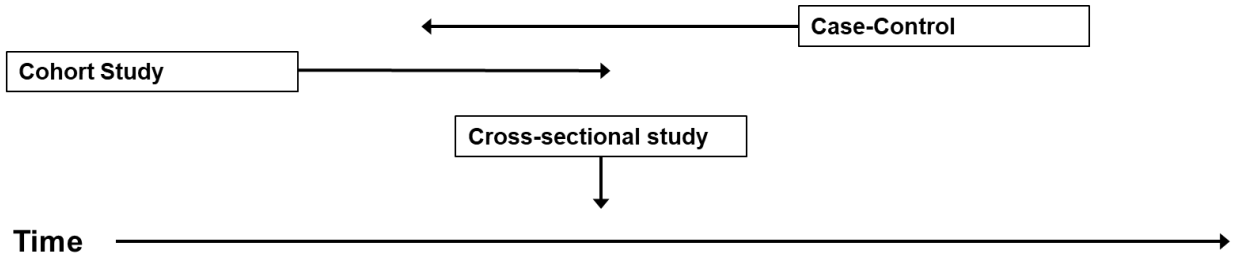


Adapted from: Psych DB: Critical appraisal and statistics

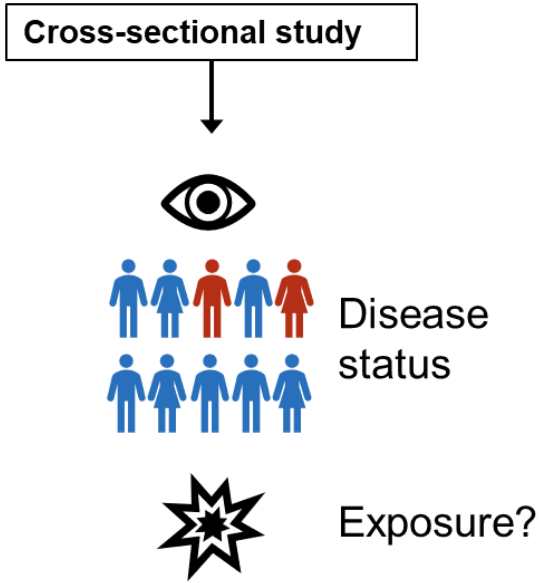
Observational studies

- The type of observational study you will perform depends on when you are collecting your observations, and how you will be sampling from the population.

- Let's go through these and examine how they attempt to find an association between 'exposure' to a risk factor, and the development of disease. These classic epidemiology study designs arose from the human health context, but are applicable to any field.



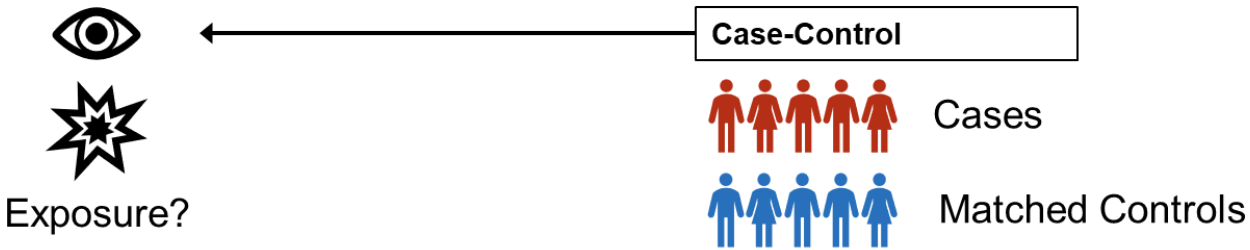
Observational studies: Cross-sectional



- Look at a single time-point.
- Relatively quick and easy to run.
- Cannot establish any temporal order of exposure and disease status, only the prevalence of disease in the population at a particular point in time.

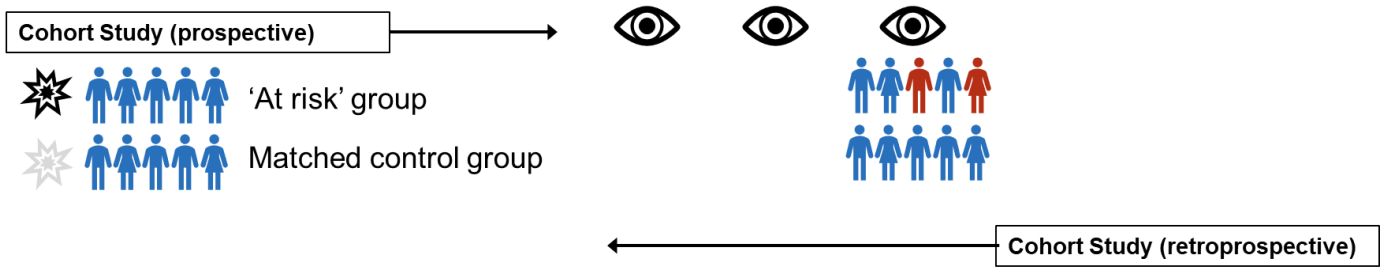
Observational studies: Case-control

- Assemble cases and controls matched on several factors, similar except for disease status.
- Look at exposure to some potential risk factor to examine association with disease.
- Good for studying rare disease with relatively common exposures.
- Sampling: Enrich for cases.



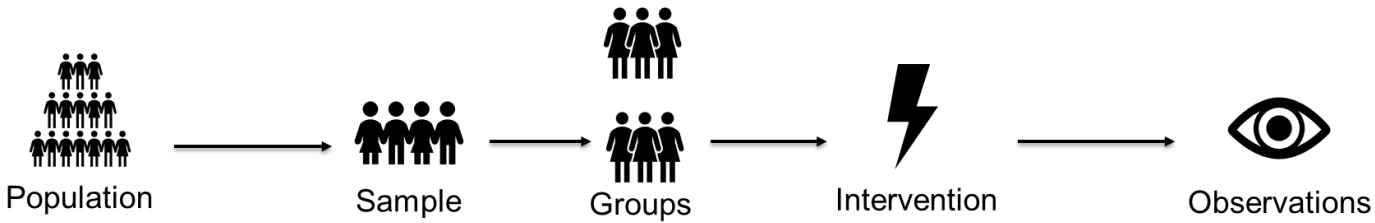
Observational studies: Cohort

- Assemble groups often based on exposure to some risk factor for developing disease and follow up over time to measure development of disease in the cohort and often other outcomes.
- Can establish temporal relationship between exposure and development of disease.
- Often more expensive and logistically challenging (loss to follow up, longitudinal observation).
- Good for studying rare exposures, and relatively common diseases.
- Can be done prospectively (enrolment prior to any disease development) or retrospectively (using existing cohorts that were followed through time).
- Sampling: Enrich for ‘at risk’ subjects.



Designed experiments

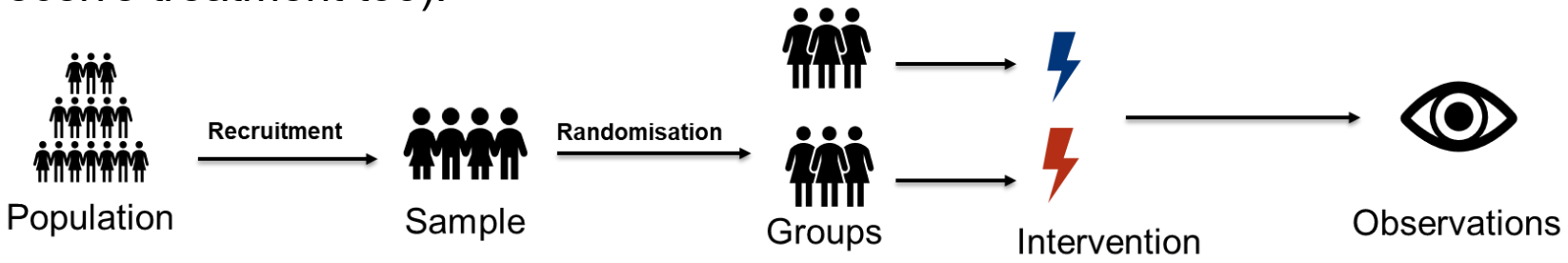
- **Designed experiments** do involve an experimental intervention. Other factors in the experiment are more likely to be designed and controlled (e.g. taking place in controlled conditions inside a lab or a clinic).



Adapted from: Psych DB: Critical appraisal and statistics

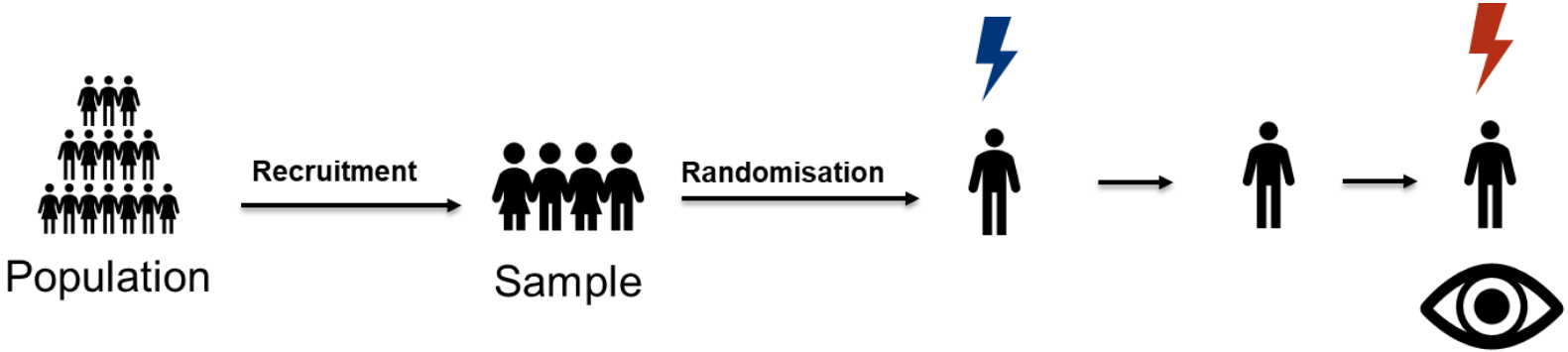
Designed experiments: Randomised Control Trial (RCT)

- Considered by many to be the gold standard for a real-world intervention (drug trials, educational interventions, etc).
- Not always possible or ethical to randomise to different treatments and hence perform an RCT.
- Sometimes interim analysis performed to determine whether trial should be stopped because of clear demonstration of benefit (and placebo group should receive treatment too).



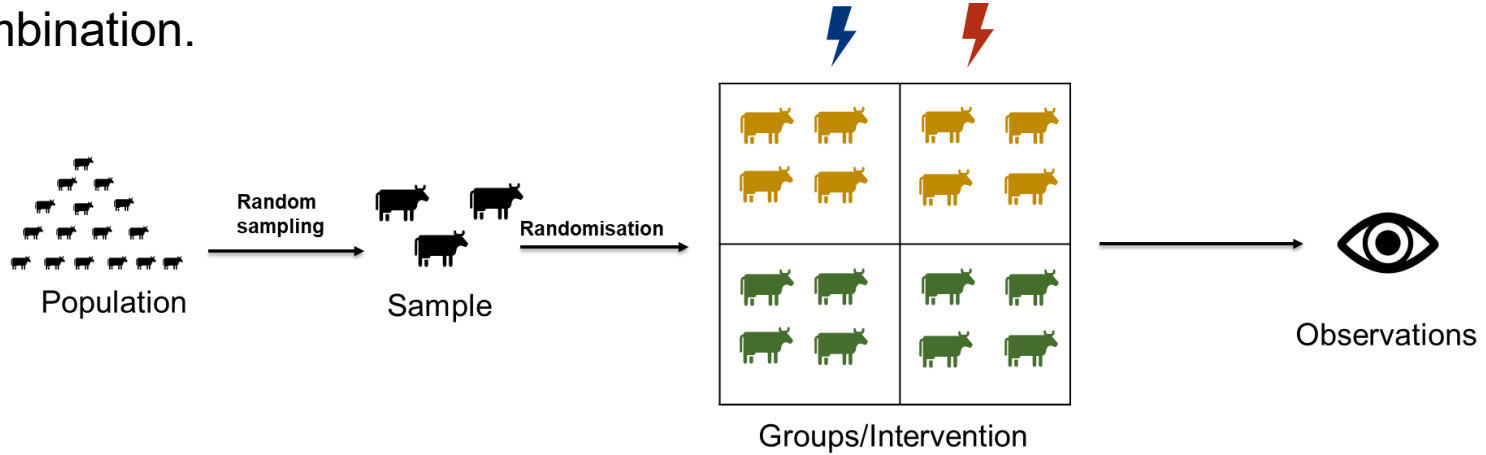
Designed experiments: Crossover trial

- Crossover trials are a ‘within-subject’ version of a randomised control trial.
- Can be used in contexts where the treatment effect can be ‘washed out’ in the subjects between different treatments.
- More efficient than an RCT (more power for the same number of subjects).
- Not always possible e.g. if the treatment effect is not transient.



Designed experiments: Factorial trial

- Common in agricultural/engineering disciplines where subject factors can be easily controlled (e.g. easy to source a male calf of a particular breed).
- Also used extensively in psychology in ‘within subject’ designs.
- Most efficient way to estimate the effects of multiple factors, alone or in combination.



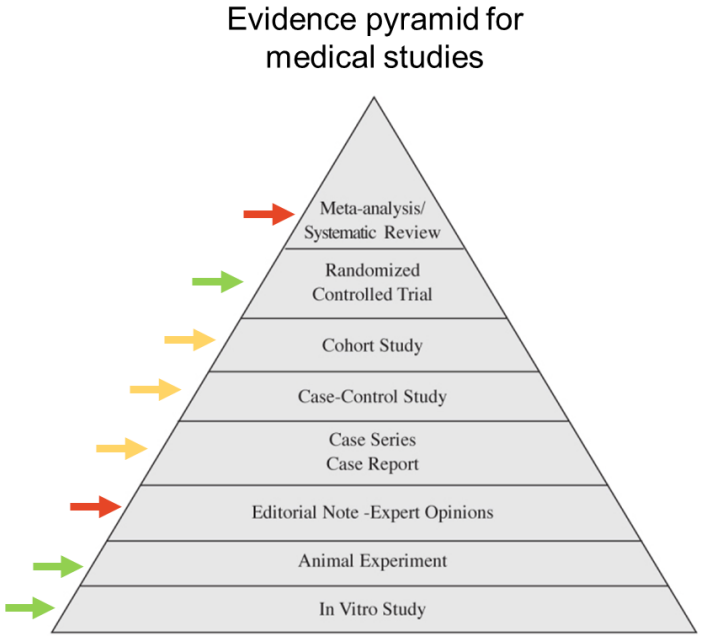
Before-After Control-Intervention (BACI) designs: Common in ecology

- ANZ Guidelines for fresh and marine water quality: Study design type
- BACI's allow for assessment of change.
    - 2 site types, monitored before and after potential effect.
        - Control – unaffected by potential effect.
        - Impact – subjected to potential effect.
- If the impact site changes relative to the control site, the effect is likely the cause.
- Inference based on Time x Site interaction, e.g. in ANOVA.
- Multiple BACI designs allow you to use multiple sampling events before and after potential effect. Modelled through linear mixed models. See our Statistical Modelling workshop series.

Choose your study type

Study Designs in Medicine

The evidence pyramid and introduction to randomized controlled trials. *Pandis NAm J Orthod Dentofacial Orthop. 2011 Sep; 140(3):446-7.*

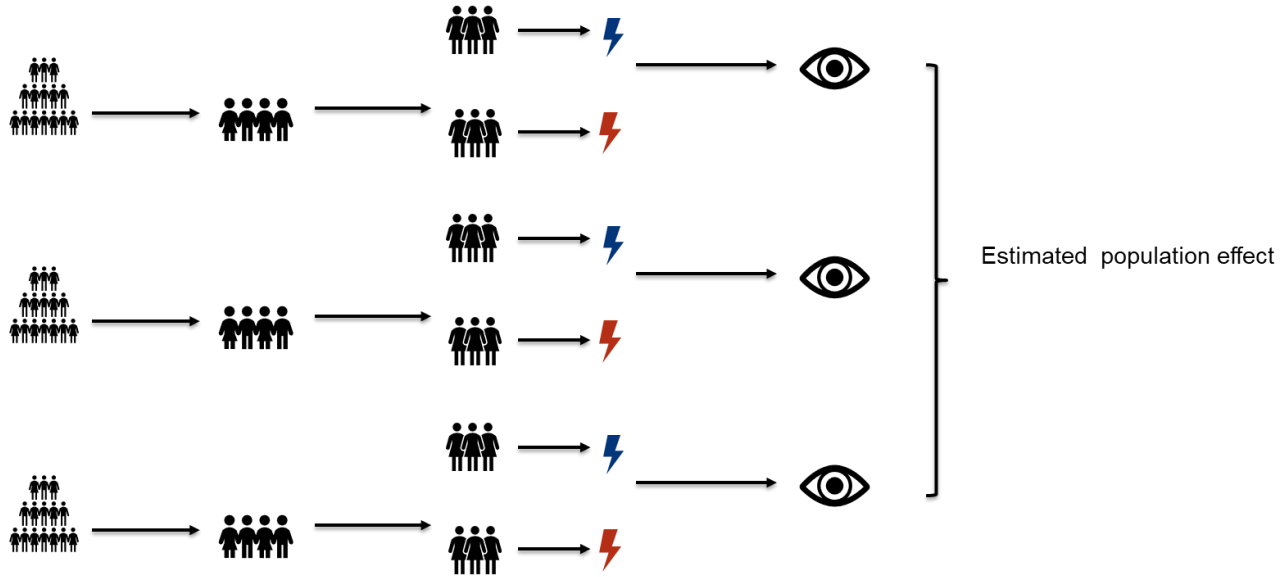


Is there a best study type?

- Consider the evidence pyramid on the left.
- Some will involve be designed experiments and others are observational, while others involve synthesis of primary literature.
- The optimal study type will depend on what is known and the feasibility of each study type in your chosen field.

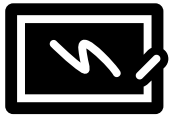
Synthesis of primary literature: Systematic review or meta-analysis

- Meta-analysis combines the estimates from multiple primary studies.
  - See our meta-analysis workshop for more detail.



Target trial emulation

Design and Implementation of Observational Studies Emulating a Target Trial



- A Framework for Causal Inference From Observational Data.
- Mimics a hypothetical RCT with observational data, aiming to reduce bias and strengthen causal credibility.
- Need to define the research question, eligibility criteria, treatment, time zero, assignment mechanism, follow-up period, outcome, estimand* and analysis plan.
- It forces you to explicitly encode the causal question and design assumptions. i.e. You define the causal estimand *before* you analyse, as you would in an RCT.
- Can address research objectives that for ethical or logical reasons, cannot be examined in RCTs.
- Garbage in, garbage out: although widely used, TTE is often difficult to implement well and is frequently done poorly, prompting the development of the TARGET Statement to improve methodological rigour.
- In a review of TTE, only 57% developed a prespecified protocol, 54% did not review existing RCTs, 44% did not report all methodologic components, 15% used post baseline information inappropriately for eligibility, only 17% provided follow-up diagrams to define time zero and only 31% of trials addressed unmeasured confounding.

Reporting of Observational Studies Explicitly Aiming to Emulate Randomized Trials

*For definition of estimand, see next slide.

Estimands, Estimators, and Estimates

- **An estimand is the true effect of the intervention**.
- i.e. A target quantity of what the study aims to measure and a summary of patient outcomes.
- Should summarise the causal effects of treatments in the target population. E.g. differences in mean outcomes, or differences in mortality rates in the population.
- You may have more than 1 estimand in a study to capture both positives and adverse effects.
    - •Dictated by your research question.

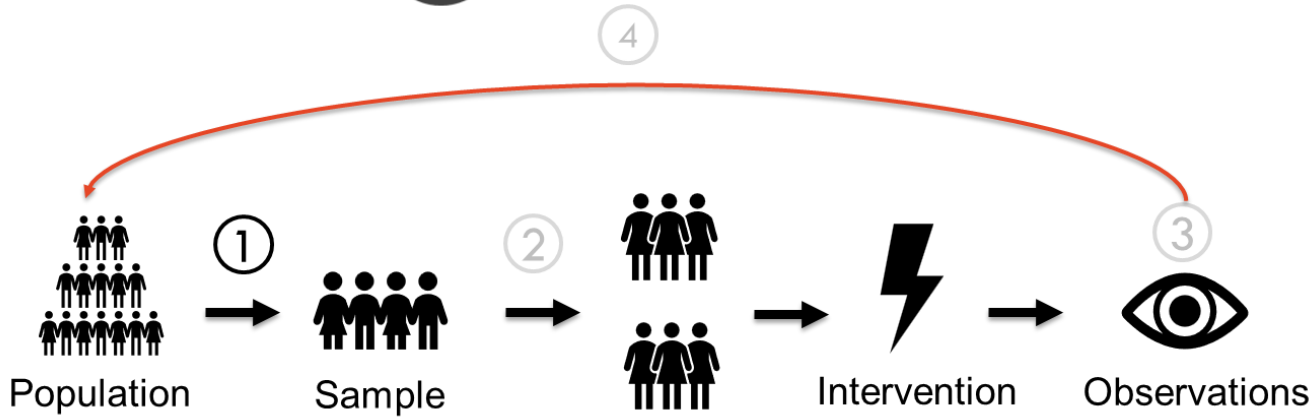
- **An estimator is a formula or algorithm used to estimate the target quantity from clinical trial data.** E.g. difference in sample means between 2 groups.
- Should summarise the causal effects of treatments in the sample of individuals in the study.
- Should provide a valid and unbiased estimate of the study estimand. Must have good internal and external validity.

- **Trial data provide only estimates of the trial estimands**, as participants are sampled from the population and outcomes are not always observed for all of the randomised participants.
    - As such, there are limitations like non-adherence to treatments or attrition.
- Numeric value obtained when the estimator is applied to the actual trial data.

JAMA: Estimands, Estimators, and Estimates

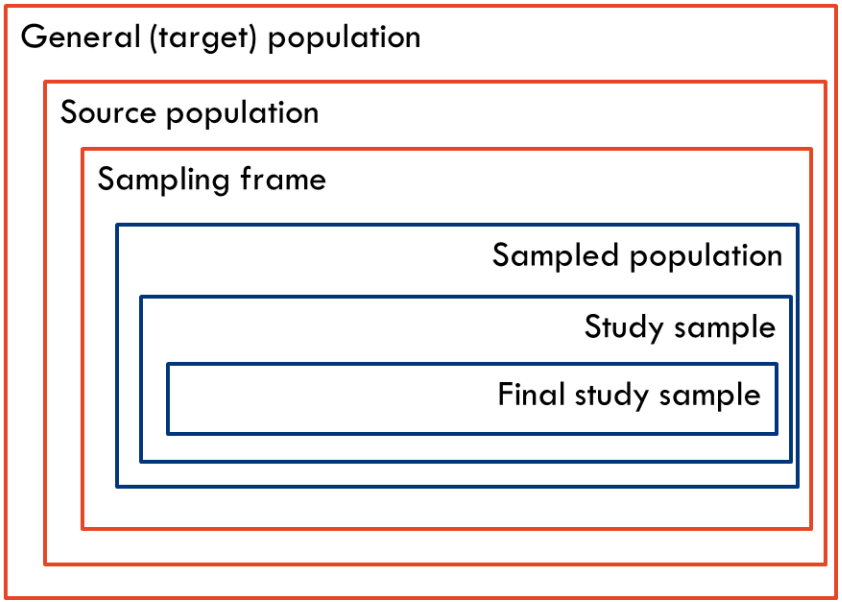
Step 1: Sampling

Step 1: Sampling



- In estimating some population property, we need to consider ‘external validity’.
- External validity is how generalisable the study is to a wider population and depends on the size and representativeness of the sample used.
- It is the [final] sample on which we take observations that we are using to infer something about our population.

Generalisability to your target population



Would the estimates obtained from each of these sets of individuals be the same (or similar enough) as the estimates for the larger box? Is the smaller box representative of the target population?

- Your **target population** is the population about which you would like to make inferences.
- From a subset of the target population called the **source population**, you define your sampling frame.
- The **sampling frame** is a list of individuals from which you draw your sample, you may exclude some individuals if you find they aren't part of your target population.
- You conduct your study and after excluding some participants to obtain your **study sample**, you run your study. You end up with complete data for all individuals in a **final study sample**.

Strata

- We may be able to group subjects who are similar to each other into strata.
- An example of strata is the age group of your subjects.
- Defining strata allows us to adjust our sampling to ensure representativeness of our sample to the target population (e.g. we could sample to match the distribution of the population into age groups).
- In a designed experiment, the allocation of treatments should be balanced in the different strata.

Source: Hilda statistical reports

Sampling methods: HILDA example

- The sampling method used for the Household, Income and Labour Dynamics in Australia (HILDA) survey. This is a panel survey, meaning that data is collected on the same panel of individuals, in this case in annual ‘waves’.

- Target population: All people living in households in Australia.

- Sampling frame: “The **reference population for Wave 1 of the HILDA Survey was all members of private dwellings in Australia**, with the main exception being the exclusion of people living in remote and sparsely populated areas.”

- Sampling method: “Households were selected using a multi-staged approach designed to ensure representativeness of the reference population. First, **a stratified random sample of 488 1996 Census Collection Districts (CDs)**, each of which contains approximately 200 to 250 households, **was selected from across Australia**. Within each of these areas, depending on the expected response and occupancy rates of the area**, a random sample of 22 to 34 dwellings was selected**…. Nonetheless, despite the region-based stratification, Wave 1 of the HILDA Survey was an equal-probability sample; in particular, the smaller states and territories were not over-sampled. This reflects the focus of the HILDA Survey on producing nationwide population estimates.”

Sampling methods: HILDA example

Sample (initial):

- “Of the 11,693 households selected for inclusion in the sample in 2001, 7,682 households agreed to participate, resulting in a household response rate of 66%. The 19,914 residents of those households form the basis of the ‘main sample’ that is interviewed in each subsequent year (or survey wave), but with interviews only conducted with people aged 15 years or older”.

Sample (subsequent):

- “Table A1 and Figure A1 summarise key aspects of the HILDA sample for the period examined in this report (Waves 1 to 22). Table A1 presents the number of households, respondents and children under 15 years of age in each wave. In Wave 22, interviews were obtained with a total of 15,954 people, of which 12,581 were from the original sample and 3,373 were from the top-up sample. Of the original 13,969 respondents in 2001, 6,180 or 53.7% of those still in scope (i.e., alive and in Australia), were still participating at Wave 22”.

The Household, Income and Labour Dynamics in Australia Survey: Selected Findings from Waves 1 to 22

Sampling methods: Some other examples

A study to measure blood glucose concentrations in people with diabetes

- **Your target population:** People with diabetes in Australia.
- **Your source population:** People presenting to a hospital in NSW with diabetes.
- **Your sampling frame:** The list of 523 people who presented to the hospital in a 3-month period.
- **Your sample:** The first 200 people from your sampling frame who consented to participating in your study.
- **Your final sample:** 184 people with complete data.

A study to evaluate the effectiveness

- **Your target population:** Gamers in Australia who play more than 10 hours a week of video games.
- **Your source population:** 50 000 gamers in Australia who were presented with a Facebook recruitment ad.
- **Your sampling frame:** The first 5000 people who volunteered for the study and met the inclusion criteria.
- **Your sample:** 200 randomly selected from the sampling frame who were selected.
- **Your final sample:** 40 people who showed up at your clinic and completed a gaming addiction treatment program.

External validity: An example

Study to evaluate the effect of a feed supplement on the growth of calves

Study Design:

- 2 groups: 1: Standard feed and 2: Standard feed with supplement
- All calves are the same breed - Charolais
- All based on one cattle farm (sampling frame: a list of all cattle IDs)

- What larger source population does this sample represent?

- What general (target) population might we wish to make inferences about?

External validity: An example

Study to evaluate the effect of a feed supplement on the growth of calves

An expanded study could now include:

- Sex: male, female
- Feed type: grass and grain
- Breed: Charolais, Hereford
- Climate: temperate, arid

- This will expand the external validity of the study to cover a much wider population but make the study potentially much more difficult to carry out.

- Compromise is often necessary. The tension in experimental design is often between what you would *like* to find out and what you feasibly *can* find out in a single experiment.

Generalisability vs. bias

- You may argue that Charolais calves are representative of all calves in a wider population of interest for your outcome of interest (serves as an appropriate model for all cattle breeds).

- The danger is when you wish to make inferences within a wider population, without adequately sampling that population or being reasonably sure that the measured individuals in your population are representative.

- These issues are where your domain expertise can help you optimise your experimental design.

Generalisability and controlled conditions

- You may plan to perform an experiment under laboratory conditions, perhaps using a model organism or in cells grown in vitro.

- The question of generalisability always arises when experiments are not performed under ‘real world’ conditions. Is bias introduced by performing the experiments in a lab compared to the real world?

- The various forms of validity to assess model organisms for human disease.

- On the other hand, controlling conditions often results in far less variability than in the real world (e.g. inbred mouse strains vs. outbred mice). This may be necessary to study subtle effects. We effectively trade off less error for potentially more bias.

- In practice, a combination of lab-based and real-world experiments are often necessary to fully characterise some biological phenomenon.

The units in your experiment

Lazic, Stanley E. *Experimental Design for Laboratory Biologists : Maximising Information and Improving Reproducibility* . Cambridge, United Kingdom: Cambridge University Press, 2016. Print.

- So, what are the units in your experiment? Understanding the types of units will help you recognise design and analysis considerations.

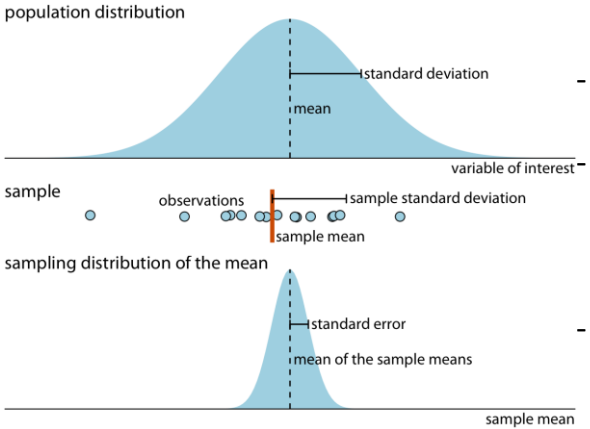
Types of Units (adapted from Lazic, 2016)

- **Sampling Unit/biological unit** – is the entity (animal/plant/thing) about which inferences are made.

- **Experimental Unit** – is the entity that is randomly and independently assigned to experimental conditions or treatments.

- **Observational Unit** – is the entity on which measurements are taken.

The units in your experiment



Recall that the variability in the population affects the variability of our estimates – so it matters that different units have different variability. We need methods that take this into account so that we can accurately infer the variability of our estimates (accurate confidence intervals and p-values).

- Why do these different units matter so much?
- One of the most important assumptions in many statistical methods is statistical independence: simply stated, making one measurement gives you no information about another measurement.
- If this assumption is violated by taking repeated measurements on the same individual, or measurements on individuals that are clustered together (e.g. in a family, or a classroom, or a hospital), we need to choose an appropriate method.
- These methods partition the variability between measurements to different sources: e.g. within-subject and between subject, so that accurate estimates and valid inference can be performed.

Lazic, Stanley E. *Experimental Design for Laboratory Biologists : Maximising Information and Improving Reproducibility* . Cambridge, United Kingdom: Cambridge University Press, 2016. Print.

Replication

The amount of replication needed to reach a desired level of statistical power is the focus of our **Power and Sample Size Calculation workshop**.

- **Replication is required for statistical inference.**
    - To make probabilistic statements about differences between groups, you must be able to estimate biological or experimental variability. As a rule of thumb, at least 3 independent replicates per group are needed for a simple comparison, although more may be required to achieve adequate statistical power.
- **True replication** occurs when you have multiple independent measurements at the level of each experimental unit within each group.
    - E.g. Measuring gene expression in 5 different mice per treatment group.
- **Technical replication** involves taking repeated measurements of the same experimental unit to improve measurement precision or assess measurement error. Technical replicates do not increase sample size and are often averaged before analysis.
    - E.g. Running the same RNA sample on a sequencing instrument multiple times.
- **Pseudo-replication** occurs when repeated measurements on a unit are not truly independent of each other. This artificially inflates the effective sample size. In some cases, repeated measurements can be appropriately analysed using repeated-measures or mixed-effects models that account for the dependence structure.
    - E.g. measuring blood pressure on the same patient 5 times and treating each as a separate patient.
- **Batch effects** are systematic differences introduced by when, where, or how samples are processed, rather than by the biological factors of interest.
    - E.g. Different sequencing runs, reagent lots, lab personnel, plates or processing dates.

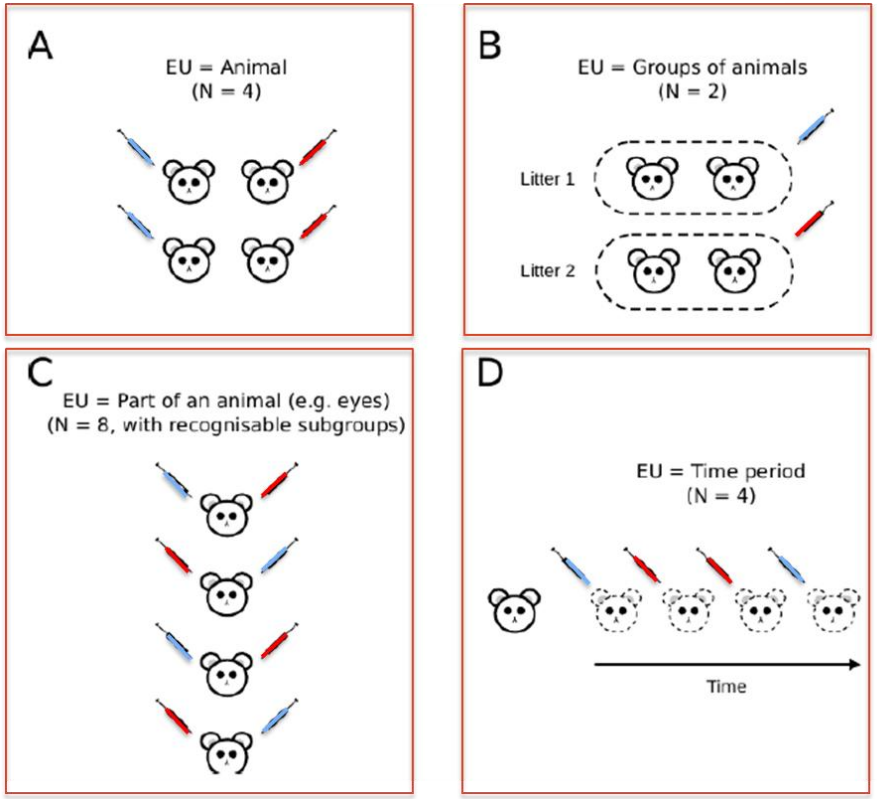
Experimental unit leads to N

- In a simple experiment, sampling unit = experimental unit = observational unit. And your sample size is the number of subjects that you have.

- Things can get complicated when your experimental unit is not the same as your sampling/biological unit. When considering replication, you may need to consider two levels: the number of individuals per cluster, and the number of clusters in your study.

- Animal experiments can be used to illustrate different forms of clustering. In a simple experiment where an experimental drug is administered to mice, the drug may be administered to each mouse. But what about an animal experiment when the treatment is applied to the mother (mare, sow, ewe, etc) and the measurements are carried out on babies in the litter?

Experimental unit leads to N



A quick challenge question

Can you make conclusions regarding causality in the following studies?

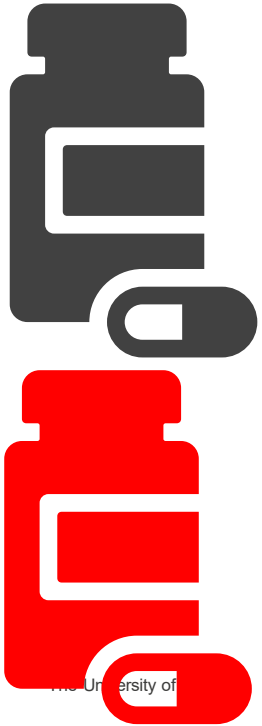
- A randomised control trial where subjects received either a novel drug, or a placebo and their symptoms were examined after 3 weeks.

- An observational (cross-sectional) study examining at the rate of disease in subjects of age 60 who routinely took aspirin (for any reason), vs. those that didn't.

- An observational cohort study examining developmental delay in children. Recruited children of the same age and following them over a 10-year period. Comparing outcomes in those who attended pre-school and those that didn't.

Step 2: Grouping

Experimental controls



- Treatment **variables**/factors usually have at least two **levels**:
  - A new/experimental treatment.
  - Control: i.e. no new treatment.

- Having at least one control level/control group is usually essential to compare the treatment level to:
  - Placebo is often used for a drug treatment.
  - Sham surgery, a form of placebo for surgical interventions.
  - Where a drug is made up in a solution for delivery to animals or cells, ‘Vehicle’ contains everything other than the drug.
  - Wild-type cell line, as opposed to mutant.
  - Standard of care may be used in a clinical experiment where a new clinical intervention is being tested.

Randomisation

What is randomisation?

- Random allocation of treatments to experimental units.

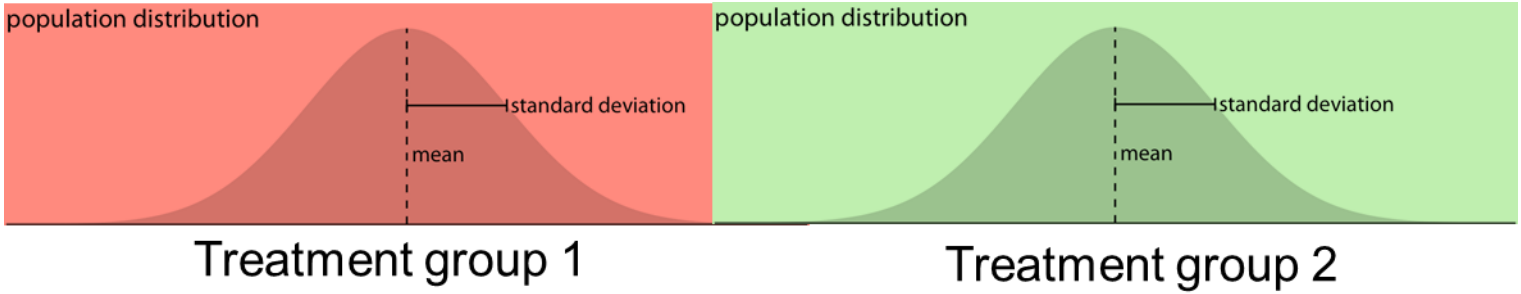
Why randomise? So we can avoid:

- Systematic bias – e.g. allocating all the drug treatments first, then the placebos.
- Selection bias – e.g. subconsciously (or consciously!) choosing healthy patients for the treatment.
- Unknown unknowns – potential confounding factors we don't even know exist.

- Randomisation greatly strengthens our causal inference by minimising the effect of confounding variables.

A philosophical point: randomisation is *not* about perfect balance

Unknown-unknowns cannot be accounted for in an observational study. Some have theorised that all factors including unknowns-unknowns could be perfectly balanced between groups in a randomised experiment. In practice, there are too many unknown-unknowns to achieve perfect balance of all factors across treatment groups with randomisation. Randomisation does not rely on this perfect balance (which in theory would result in no variability between trials!), instead it allows us to balance the major sources of variability and end up with treatment groups with potentially different means, and individuals that vary to the same extent as individuals in our target population would (after any blocking). Thus, we can make as precise estimates as possible in the presence of inevitable randomness.



S. Senn: Randomisation is not about balance, nor about homogeneity but about randomness (Guest Post)

Experimental Design – Completely Randomised Design

CRD (unrestricted design)

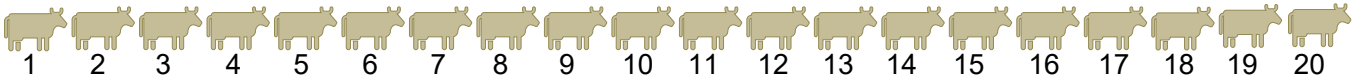
Example: Evaluate the effect of a feed supplement on the growth of calves

- Suppose that we have no information about the calves (subjects) that we might otherwise use.
- In this case we treat all subjects the same and use randomisation to eliminate allocation biases (e.g. so the biggest, pushiest cows don't end up in one group).

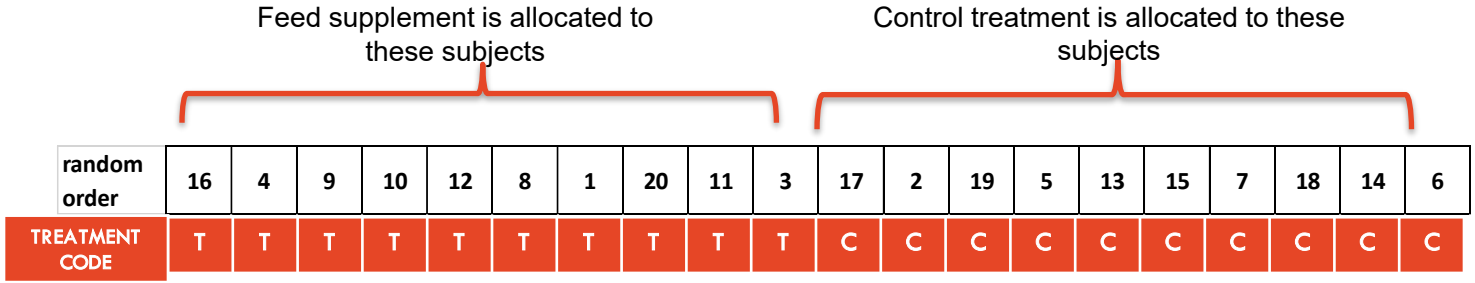
Experimental Design – CRD

A1	B1
16	0.031141412
4	0.041331209
9	0.044095909
10	0.132242434
…	…

- Suppose we have 20 subjects and 2 treatments (T and C)
- Assign an ID number to each subject from 1 to 20



- Generate a randomly ordered sequence of numbers 1 to 20 (eg from Excel)
- In Excel use formulae: with IDs in A1; B1=rand(); copy down 20 rows; sort on B1



Study designs incorporating restricted randomisation

The completely randomised design is the simplest design and corresponds to unrestricted randomisation.

Most other designs introduce restrictions on the randomisation process to control known sources of variation.

We will now discuss these approaches…

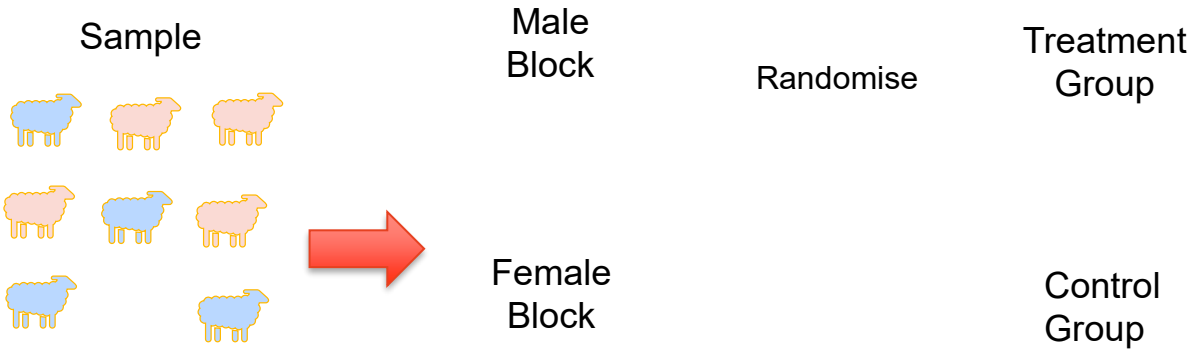
Blocking

- In practice, we may be able to group subjects who are more similar to each other into **blocks.** This is the designed experiment equivalent of **strata** in observational studies.
- A blocking variable is a variable that is thought to affect the outcome but is typically not of interest to the experimenter.
- Balancing your blocking variables within treatment groups will minimise error in your estimation of the treatment effects.
- You may have more than one blocking variable, but it is often only feasible to have a few, so choose those which have the greatest potential effect on your outcome.
- Rule of thumb: “Block what you can, randomise [for] what you cannot” – George Box.

Experimental Design – Randomised Block Design

Randomised Block Design (RBD)

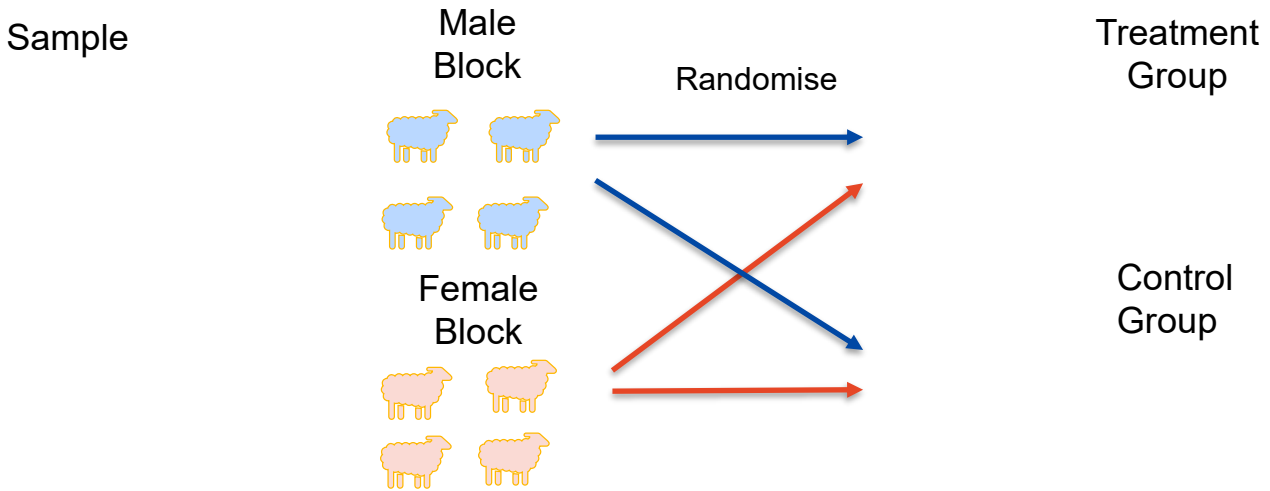
- Example: Evaluate the effect of a feed supplement on the growth of sheep.
- Suppose now that we are able to source equal numbers of males and females.
- Use sex as a block variable and randomise within blocks.



Experimental Design – Randomised Block Design

Randomised Block Design (RBD)

- Example: Evaluate the effect of a feed supplement on the growth of sheep.
- Suppose now that we are able to source equal numbers of males and females.
- Use sex as a block variable and randomise within blocks.



Randomised Block Design (RBD)

- Sex will be a blocking variable.
- Male Block

Treatment Code	T	T	T	T	T	C	C	C	C	C
Random order	10	8	2	1	4	9	3	7	6	5

- Female Block

Treatment Code	T	T	T	T	T	C	C	C	C	C
Random order	14	20	19	11	15	18	12	17	16	13

- The allocation is randomised within each block.
  Codes for M 1~10, codes for F 11~20.
- What would be the disadvantage of not blocking for sex in this case?
- How will this experiment be analysed?

Randomised Block Design (RBD): Analysis

- If we do not think males and females have potentially different responses to treatment, then we could use an ANOVA model that adjusts for block and tests for the effect of supplement.

  $Y = b_o + (\text{supplement})b_1 + (\text{male})b_2 + e$

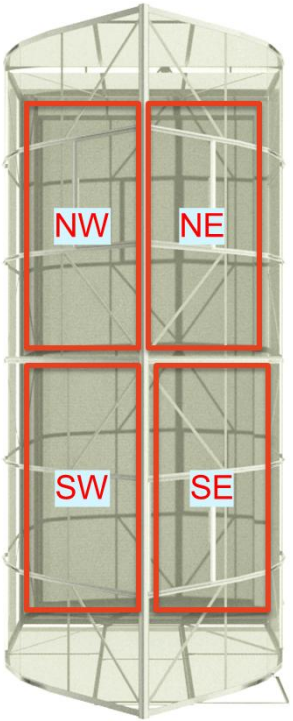
- If we hypothesise difference in male and female response to protein supplement we use an interaction model.

  $Y = b_o + (\text{supplement})b_1 + (\text{male})b_2 + (\text{supplement*male})b_3 + e$

Main effect of interest: b_1
Blocking variable effect: b_2
Interaction is the additional effect of supplement for male (interaction): b_3

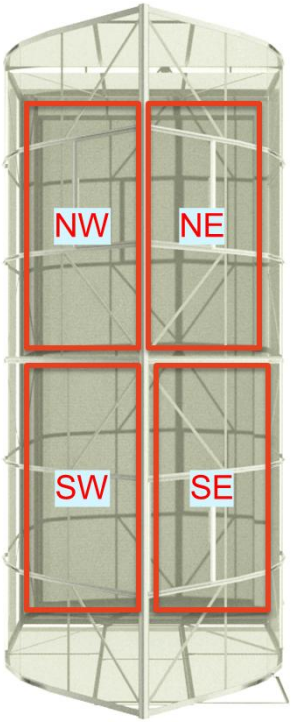
- See our Linear Models series of workshops for more details on analysis using ANOVA.

Latin Square design



- Used to create a balanced design with more than one blocking factor.
- **Example:** Growing plants in a greenhouse using different fertilisers.
- **Treatments:** Fertiliser A, B, C & D
- **Row block:** Shelf position 1,2,3,4
- **Column block:** Corner position NE, NW, SE, SW
- We can create a **4 x 4 Latin Square design**.

Latin Square design



- **Example:** Growing plants in a greenhouse using different fertilisers.
- Each treatment occurs once per shelf position and once per corner position.

Row = Greenhouse corner	Column = Shelf position: 1	2	3	4
NW	A	B	C	D
NE	D	A	B	C
SW	C	D	A	B
SE	B	C	D	A

Incomplete block design

- The previous examples worked out well because the size of the block matched the number of treatments (four shelves, four corner positions and four fertilisers).

- The advantage of a complete block design is because all blocks contain all treatments, there is no confounding of treatment effect with the effect of block membership, analysis is simpler and more efficient (less replicates required for same accuracy).

- In reality, our experiments may not work out so nicely. We may want to use blocks that do not have as large a block size as the number of treatments.

- In this case we can use a form of incomplete block design, we still aim for balance across the experiment, but cannot achieve it within a single block.

Balanced incomplete block design

- Every pair of treatments occurs together in a block the same number of times.
- Example: 4 treatments A, B, C & D.
- Block size is only 3 (e.g. 3 shelves, not 4).

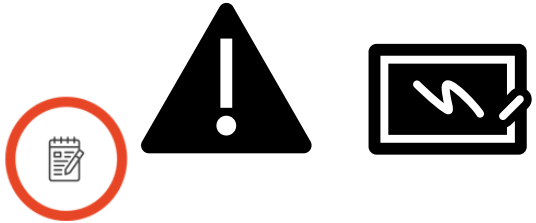
	Column = Shelf position		
	1	2	3
Block 1	A	B	C
Block 2	A	B	D
Block 3	A	C	D
Block 4	B	C	D

- Don’t forget, blocks could be batches, days, cycles, fields, etc.

Split-plot design

- Split-organ (within-subject) designs, where different treatments are randomised to sub-units within the same participant.
  - E.g. split-mouth (dentistry), paired eye (ophthalmology), left/right limb studies (rehabilitation or sports science), split-face (dermatology).
  - The participant acts as a ‘block’ and treatments are randomised to the body region within the participant.

Allocation in randomised control trials



- In RCTs you may have heard the term ‘block randomisation’ to refer to allocation of patients to treatments. This is very different to our Randomised Block Design.

- In RCTs we are randomly allocating treatments to participants who enter the study at different times. Before the trial we come up with a randomisation sequence. We use shorter blocks of sequence to avoid imbalance in the number in each treatment group at any point during trial.

- E.g. for a two-arm trial with (T)reatment and (C)ontrol:
- There are six possible blocks of size 4: TCTC, CTCT, TCCT, CTTC, TTCC, CCTT
- If we were recruiting a maximum of sixteen patients, we could randomly choose four of these blocks for our random allocation sequence: TCTC, CTCT, CTCT, CCTT

- To preserve blinding (concealment of allocation), you can use the above technique with random block sizes.

- If we have defined strata in an RCT, we can come up with a sequence for each strata, which is referred to as a ‘stratified randomisation’.

Common types of randomisation

Choosing and evaluating randomisation methods in clinical trials: a qualitative study

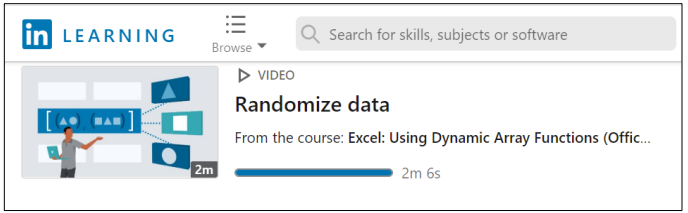
Choice of randomisation method should depend on the study context, objectives, sample size and resources available.

Randomisation method	Description	Pros	Cons
Simple randomisation	Participants are assigned to treatment groups purely at random, often using random number generators.	Easy to implement; ensures unpredictability. Good for large groups.	Can lead to imbalances in group sizes, especially in small trials, reducing statistical power.
Block randomisation	Ensures that treatment groups remain balanced at predefined intervals by dividing participants into blocks and randomly assigning treatments within each block.	Maintains balance in group sizes; reduces selection bias.	Can be complex to implement; block size may be predictable; groups rarely comparable in terms of covariates.
Stratified randomisation	Participants are divided into strata based on specific characteristics (e.g., age, gender) and then randomly assigned to treatment groups within each stratum.	Ensures balance of important covariates across treatment groups.	Requires knowledge of stratification factors; more complex to implement when there are many covariates; requires identification of participants before group assignment.
Minimisation	Assigns participants to treatment groups in a way that minimises imbalance in predefined covariates.	Highly effective in achieving balance even in small samples; adaptable to changing trial conditions.	Can be complex to implement; can have issues with unbalanced allocation ratios; may introduce predictability if not properly managed.

Methods for randomising a sequence

We saw how random sequences are useful for:

- Random sampling of individuals from a sampling frame.
- Random allocation of treatments (randomisation).



Other resources for Randomisation:

- Another nifty random number sequence generator is at: www.random.org/sequences

- If you use R, the {blockrand} package is useful for permuted block randomisations. For other randomisation r packages, see: Randomization for Clinical Trials with R.

- Have a look at this video on LinkedIn Learning (through your USyd account, check instructions on Services Portal).

Step 3: Data collection

Data collection best practices

- See our Research Essentials workshop.

  - Data storage
    - Back up EVERYTHING including original data collection forms or raw data (images, electrical signals, DNA sequences, whatever)
  - Data entry - will you be using manual data entry?
    - Ideally double-data entry followed by comparison
    - Be wary of spreadsheets – especially entering, editing analysing in the same sheet
    - Statistical software generally doesn't allow easy editing once you have entered your data

- Paired data is always more useful than unpaired data – so consider if there is a way to link your measurements to individual subjects.

Blinding

Blinding: Who, what, when, why, how?

- Blinding/masking is the concealment of group allocation from one or more individuals included in a study.

- Blinding is another method to avoid bias. It can reduce or eliminate confounding bias due to conscious or unconscious preferences or expectations.

- If possible, blinding should occur at all stages of the research workflow, and be applied to:
    - Participants, clinicians, data collectors, outcome adjudicators and data analysts.

- You should disclose in your methodology if it is not possible to blind, so that readers can assess how this may affect the reliability of your results.

Blinding

Blinding: Who, what, when, why, how?

In RCTs:
- Blind trials (or single blind) – the subject does not know if they are in the treatment or the placebo group.
- Double-Blind trials – Both the subject and the technician are not aware of the assigned treatment.
- Open trial – All the treatment information is known to the subject and technician/experimenter.

Laboratory experiments can also benefit from blinding to prevent bias:
Example 1: Histology cell counts
- Counting cells requires judgement (e.g. location sampling, recognising cell types).
- The technician should not know the identity of the specimens.
- Use an ID code to anonymise the samples. Randomisation of processing order will also help.

Example 2: Animal behaviour
- Many animals respond to the way they are handled.
- The technician should (ideally) not know the identity of the animal's treatment group.

Can you use blinding in your research to guard against unconscious bias?

Step 4: Data analysis

Why consider analysis of your data before you have even collected it?

- First and foremost, you should have some idea of the power of your experiment to know whether it is worth doing as designed.
- In order to calculate the power of an experiment, you need a statistical analysis plan.
- More generally, thinking ‘ahead’ to the analysis of your data may reveal aspects of your experiment that you hadn’t thought about (controls, number of groups, group size), which will in turn affect what data you collect… “an important way to ensure that you collect all the data you need and that you use all the data you collect”.
- Important for research integrity and quality, guarding against data-driven results and ensuring reproducibility.
- Our Research Essentials workshop discusses possible analysis methods for different variable types, and our other workshops will help you develop a detailed analysis plan once you have chosen your statistical methods.

Some statistical rules of thumb for modelling…

- No rationale for 1 variable per 10 events criterion for binary logistic regression analysis
- Adequate sample size for developing prediction models is not simply related to events per variable
- A simulation study of sample size demonstrated the importance of the number of events per variable to develop prediction models in clustered data
- Sample size for binary logistic prediction models: Beyond events per variable criteria
- Relaxing the Rule of Ten Events per Variable in Logistic and Cox Regression
- One in ten rule
- The number of subjects per variable required in linear regression analyses

Consider a linear model for your data analysis

- To analyse your data appropriately you may need to think about what kind of linear model you need. Even if your outcome is categorical, even if you plan to do a simple group comparison (such as a t-test or ANOVA) you should consider which model is appropriate.

- A linear model allows you to adjust for confounding factors that may affect your outcome.

One or more fixed factors

- These are usually the explanatory variables chosen by the experimenter. They have defined levels or categories and for factors of interest, we want to quantify the difference between them.

Your model may also include random factors

- These are usually incidental to the purpose of the experiment (such as blocking variables).
- The levels of the random factor should be chosen from a larger population of possible values of the variable.
- We don't estimate the difference between levels of a random effect, rather we use the effect to partition variance and thereby reduce within group variance.
- You may use multiple levels of random effects (e.g. individuals within households within districts).

- Discussed in further detail in the Statistical Model Building and Linear Models workshops.

Analysis of repeated measures

Repeated Measures Design (or within-subjects design)

- Repeated measures are not technical replicates when they represent another aspect of the same subject/sample, typically observations over time.
- Repeated measures are pseudoreplicates (not independent observations).
- There are specific statistical procedures to deal with RM's, discussed in our Linear Models workshops.
- A paired t-test is a simple repeated measures extension to an 'unpaired' t-test.

- It is recommended to calculate the amount of replication required for your experiment using power calculation. The topic of our Power and Sample Size workshop.

Tips from the consulting room

Study pre-registration

CENTER FOR OPEN SCIENCE
SCIENCE WORKS BEST IN THE OPEN

https://www.cos.io/initiatives/prereg

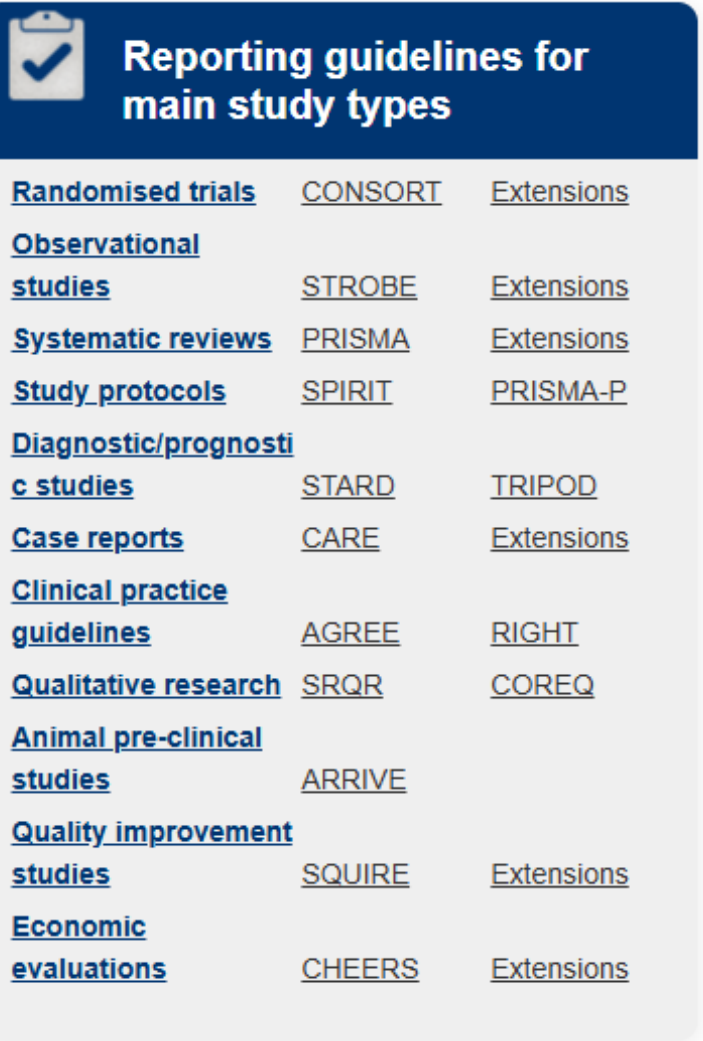
- You can future-proof your research by pre-registering your study.
  - This involves specifying your research plan in advance of your study and submitting it to an open registry.
  - It separates hypothesis-generating exploratory studies from hypothesis-testing confirmatory studies. More suited to confirmatory than exploratory research.

- Preregistration increases the credibility of your findings, allows you to stake your claim to ideas earlier and helps with the study planning process.

Best practices for data management and sharing in experimental biomedical research

Best-practice reporting guidelines

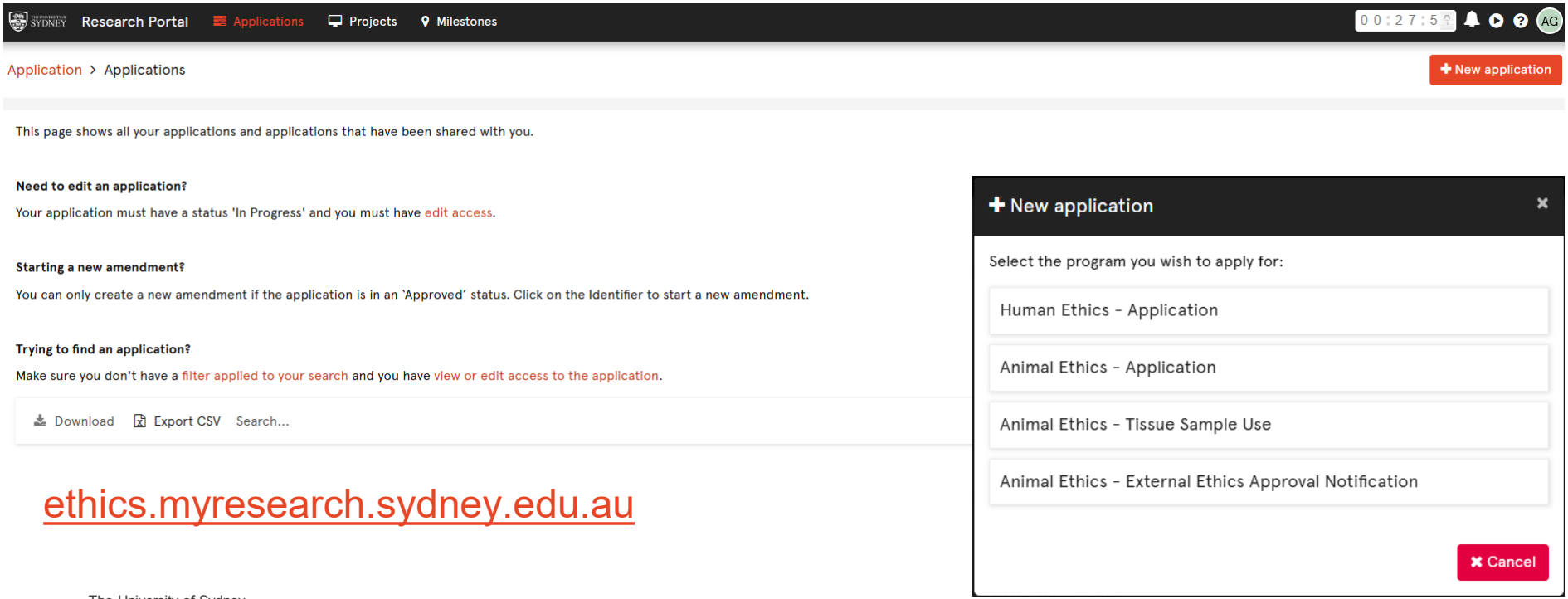
- The Equator Network (Enhancing the QUAlity and Transparency Of health Research) is a centralised platform for researchers to access a wide range of reporting guidelines.
    - Seeks to improve the reliability and value of published health research by promoting transparent, standardised and accurate reporting.
- There are guidelines for different study types to help you with your manuscript writing.
    - Provides you with a check-list of information required so that your manuscript can be easily understood by the reader and reproduced by other researchers.



Best-practice reporting guidelines

Best practices for data management and sharing in experimental biomedical research

Human and animal ethics approval



ethics.myresearch.sydney.edu.au

Further information: How do I manage my research data?

Research Data Management

Research Data Consulting
Research Integrity & Ethics Administration
digital.research@sydney.edu.au
Book a consultation

Supported platforms

- eNotebook
- RedCap
- Research data store
- OneDrive (Office 365)

Do not use

- Google Drive
- Survey Monkey
- Portable Drives e.g. USB

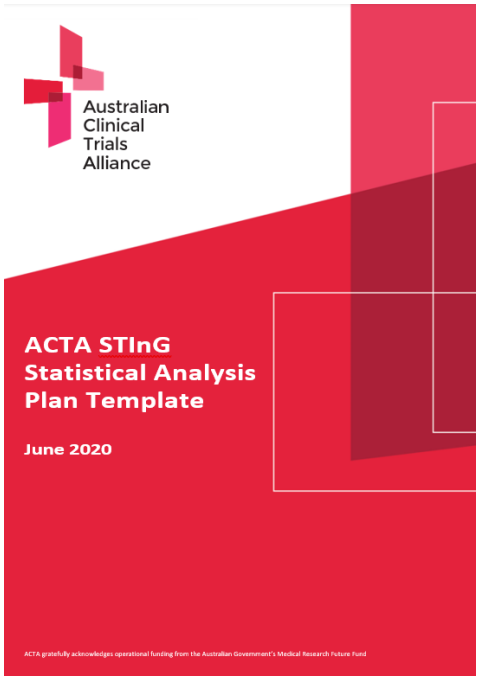
University-supported research data platforms: Features and comparisons

Research data that is managed optimally improves research efficiency and reach, as well as ensuring its integrity and security, and meeting legislative/policy/funding/publishing requirements.

The Research Data Consulting team assists researchers to enhance their research productivity and improve data management practices. They provide:

- Short consultations to integrate digital tools and data management into your research.
- Training and functional support for university supported tools/platforms.

Statistical analysis plans (SAPs)



- SAPs as well as study protocols are a great tool to improve communication and collaboration between you and other researchers.

- They should include information on:
  - Objectives and hypotheses
  - Primary and secondary outcomes
  - Study design
  - Data collection methods
  - Data management
  - Statistical analysis including handling of missing data and sensitivity analyses
  - Reporting of findings

- Australian Clinical Trials Alliance: Statistical analysis plan
- Guidelines for the Content of Statistical Analysis Plans in Clinical Trials
- A template for the authoring of statistical analysis plans
- Developing a Quantitative Data Analysis Plan for Observational Studies

Other tips and tricks

- If you have a target journal in mind, check out their instructions to authors so you know what to include in your manuscript.
    - Also check out recent publications in a relevant or similar topic, to see how they have structured their manuscript, and the level of detail needed for a peer-reviewed publication.

- Sometimes you must take a pragmatic approach and be realistic about the resources and funding you have available.

- Always think back to your research question and the end goal of your research. Think of the ‘story’ you want to tell.

- Put your project aside and revisit it later with fresh eyes – this may provide you with new perspectives which you previously hadn’t considered.

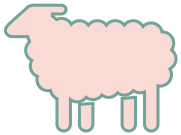
- Pilot your study if you can and have a contingency plan for all stages of the research process.

Conclusions

Challenge question – Sheep vaccine experiment

Research question: Does the use of a new vaccine result in a different incidence of parasite infection compared to the standard treatment?

- I would like 12 sheep in each group (total n = 24)
- I have 12 sheep aged 1yr and 12 sheep aged 2yrs



Q: How should you allocate the treatments to the sheep?

a. vaccinate 6 of the younger sheep and 6 of the older sheep with each treatment.

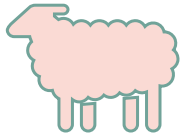
b. vaccinate12 younger sheep with the new vaccine and 12 older sheep with the standard vaccine treatment.

Option b will result in age and treatment being _____

Challenge question – Sheep vaccine experiment

Research question: Does the use of a new vaccine result in a different incidence of parasite infection compared to the standard treatment?

- I would like 12 sheep in each group (total n = 24)
- I have 12 sheep aged 1yr and 12 sheep aged 2yrs



Q: How should you allocate the treatments to the sheep?

a. vaccinate 6 of the younger sheep and 6 of the older sheep with each treatment.

b. vaccinate12 younger sheep with the new vaccine and 12 older sheep with the standard vaccine treatment.

Option b will result in age and treatment being confounded/aliased.

Challenge question – Chicken drug and diet experiment

Research question: Groups of treated and untreated chickens are placed on a high fat diet. What is the effect on weight gain?

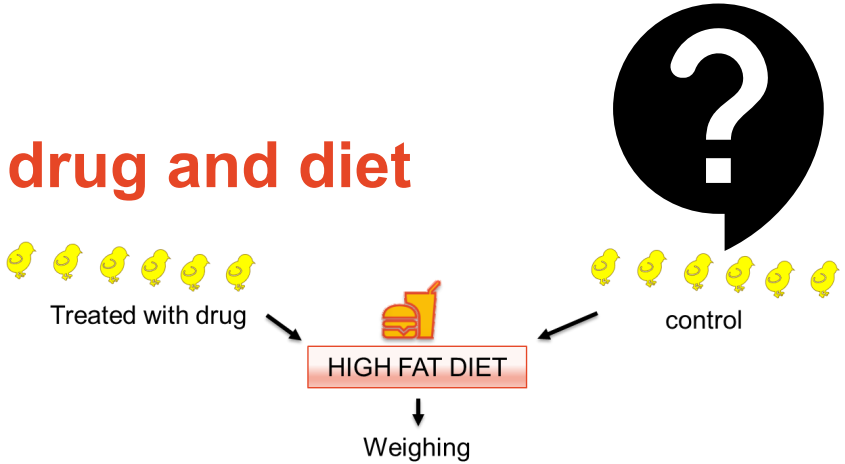
- What are the outcome and explanatory variables?
- What are the treatment variables?
- What are the blocking variables?
- What factors should be fixed in the analysis?

Challenge question – Chicken drug and diet experiment

Biological Units	chicks
Experimental Units	chicks?
Observational Units	chicks

Outcome Variables	Explanatory Variables	Design features	Blocking Factors	Randomisation & blinding

Challenge question – Chicken drug and diet experiment



Outcome variables	Explanatory Variables	Design features	Blocking Factors	Randomisation & blinding
• Weight gain	• Drug treatment (y/n)	• Chick breed • High Fat Diet • Feeding routine • Feeding ad libitum? • Time of day for weighing • Housing – number of chicks per cage?	• Sex? • Chick age, batch, etc	• Drug treatment allocation • Order of handling • Order of weighing

Take home messages

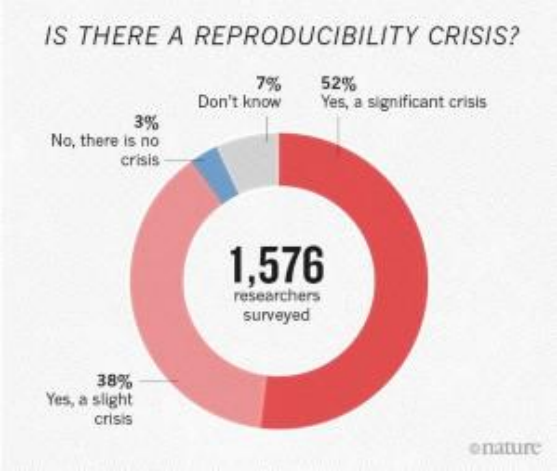
- Wherever possible, reduce or eliminate variation due to factors other than the explanatory variable of interest (often treatment variable).
- Sometimes undesirable variation cannot be avoided due to things beyond your control.
- Use blocking variables in your design to manage factors that are most likely to cause variation.
- Use randomisation to prevent bias due to allocation and increase precision in the face of unknown variation outside your control.
- Use replication to improve precision of your estimated effects.

- Use the general experimental design workflow in this workshop and note the double arrows between stages of the design: there will often be an iterative process of improvement:
    - Point out the problems
    - Discuss the implications
    - Propose a way forward

Be alert, not alarmed: your study doesn't have to be perfect, but it should be well thought out through a process of study design.

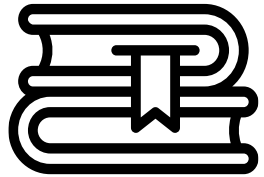
The reproducibility crisis

- 90% of respondents to a *Nature* survey agreed that there is a reproducibility crisis in scientific research fields.
- More than 70% of researchers have been unable to reproduce another scientist's experiments, and more than 50% could not reproduce their own.
- Nearly 90% of respondents agreed that "More robust experimental design", "better statistics" and "better mentorship" would improve reproducibility.



1,500 scientists lift the lid on reproducibility

Other resources



Books on Experimental Design

- “The Design of Experiments” by Fisher, Ronald Aylmer, 1935.
- “Experimental Design for Laboratory Biologists: Maximising Information and Improving Reproducibility” by S.E. Lazic
- “Statistics for Experimenters” by Box, Hunter & Hunter
- Interactive E-book on Experimental Design: Computer-Assisted Statistics Textbooks

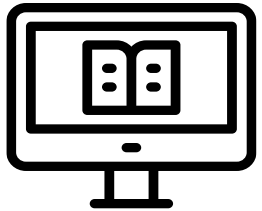
Books on Causality

- “The Book of Why” by Judea Pearl (interesting ideas on causality, confounding, approaches to data)

Books on Bias and Statistical thinking

- “Thinking, Fast and Slow” by Daniel Kahneman
- The Art of Statistics by Sir David Spiegelhalter

Other resources



Podcasts on Experimental Design and stats in general

- The Casual Inference podcast, partnered with the American Journal of Epidemiology
- #143 – John Ioannidis, M.D., D.Sc.: Why most biomedical research is flawed, and how to improve it
- #269 – Good vs. bad science: how to read and understand scientific studies
- JAMAevidence: JAMA Guide to Statistics and Methods
- JAMA: Pragmatic Trials: Practical Answers to “Real-world” Questions With Harold C. Sox, MD
- JAMA: Case-Control Studies: Using “Real-world” Evidence to Assess Association, With Dr Irony
- JAMA: Randomization in Clinical Trials from the JAMA Guide to Statistics and Methods

Software for Experimental Design

- The Grammar of Experimental Designs: Eddible R-package
- The experimental design assistant: Helping researchers worldwide design robust and reliable experiments

Further assistance at The University of Sydney

SIH

- Statistical Resources website: containing our workshop slides and our favourite external resources (including links for learning R and SPSS).
- Hacky Hour: an informal monthly meetup for getting help with coding or using statistics software.
- 1on1 Consults can be requested on our website or here (click on the big red ‘contact us’ link).

SIH Workshops

- Create your own custom programs tailored to your research needs by attending more of our Statistical Consulting workshops. Look for the statistics workshops on our training page or on our Training calendar.
- Sign up to our mailing list to be notified of upcoming training.

Other

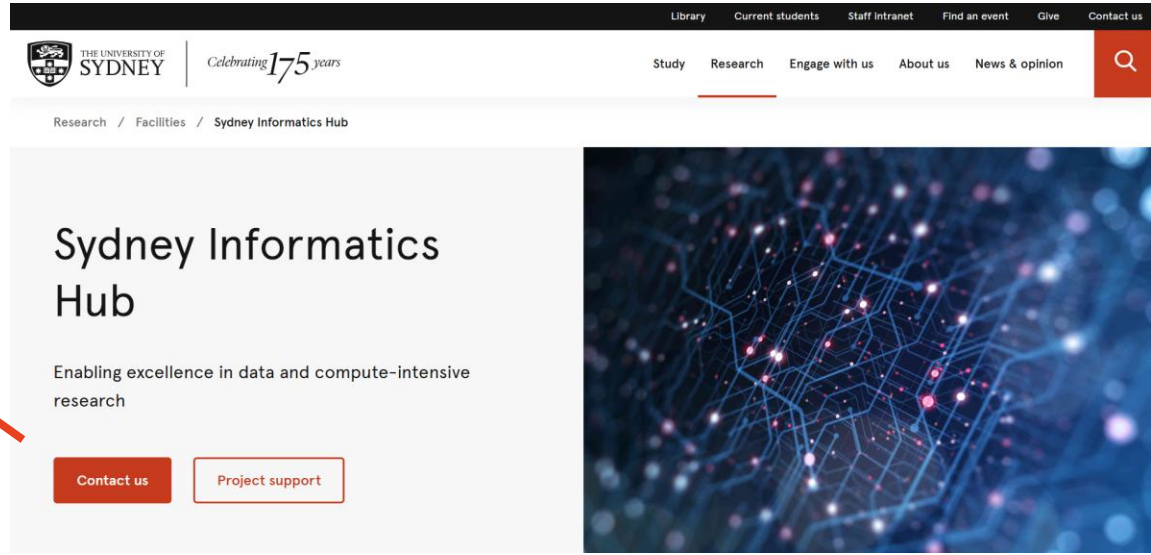
- Open Learning Environment (OLE) courses
- Linkedin Learning

How to engage with us

sydney.edu.au/informatics-hub

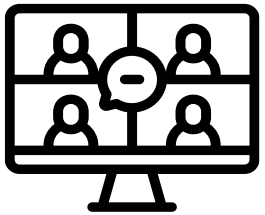
Contact

And fill out our request form



Or send us an email:
sih.info@sydney.edu.au

How to use our workshops

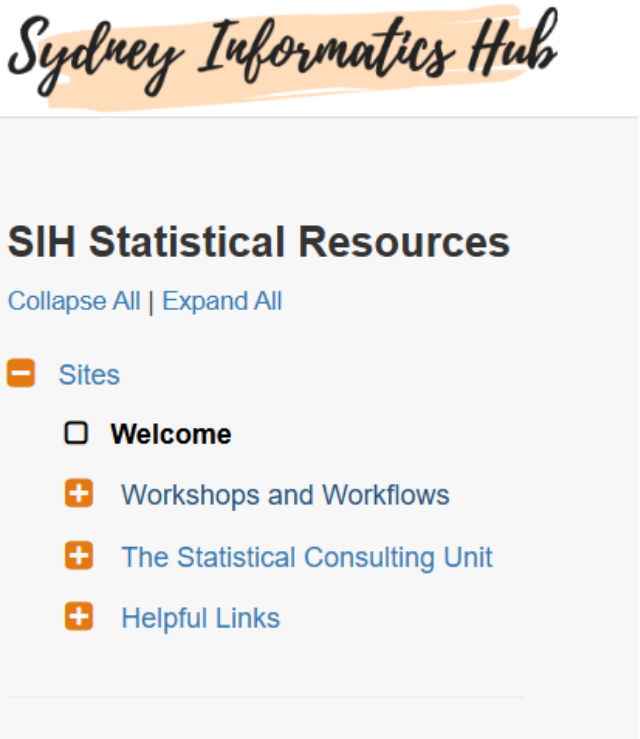


- Workshops developed by the Statistical Consulting Team within the Sydney Informatics Hub form an integrated modular framework. Researchers are encouraged to choose modules to create custom programs tailored to their specific needs. This is achieved through:
    - Short 90-minute workshops, acknowledging researchers rarely have time for long multi day workshops.
    - Providing statistical workflows appliable in any software, that give practical step by step instructions which researchers return to when analysing and interpreting their data or designing their study e.g. workflows for designing studies for strong causal inference, model diagnostics, interpretation and presentation of results.
    - Each one focusing on a specific statistical method while also integrating and referencing the others to give a holistic understanding of how data can be transformed into knowledge from a statistical perspective from hypothesis generation to publication.

For other workshops that fit into this integrated framework, refer to our training link page under statistics, found below:
Workshops and training

Online statistical consulting resources

Sydney Informatics Hub



Sydney Informatics Hub: Statistical Resources

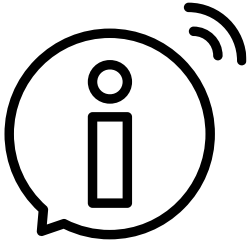
Workshops and Workflows: Download all our workshops with their step-by-step workflows on how to do analysis.

References: 1-3 Curated links on a variety of common statistical concepts and tests e.g.. Basic Theory, Meta Analysis, Software, etc. ***A great place to start looking for help!!***

Statistical Consulting Unit/What to expect in a consult: For tips on how to get the most out of your consult.

(You may need to click “Expand all” on the Sites sidebar to get an overview of the available pages).

A reminder: Acknowledging SIH



- All University of Sydney resources are available to researchers free of charge.

- The use of the SIH services including the Artemis HPC and associated support and training warrants acknowledgement in any publications, conference proceedings or posters describing work facilitated by these services.

- The continued acknowledgment of the use of SIH facilities ensures the sustainability of our services.

Suggested wording for use of workshops and workflows:

- *“The authors acknowledge the Statistical workshops and workflows provided by the Sydney Informatics Hub, a Core Research Facility of the University of Sydney.”*

We value your feedback

- We want to hear about you and whether this workshop has helped you in your research. What worked and what didn't work.

- We actively use the feedback to improve our workshops.

- Completing this survey really does help us and we would appreciate your help! It only takes a few minutes to complete (promise!)

- You will receive a link to the anonymous survey by email.