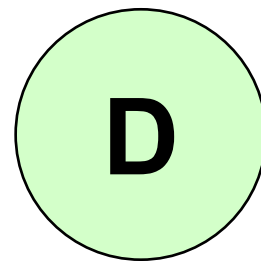
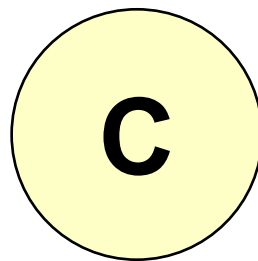
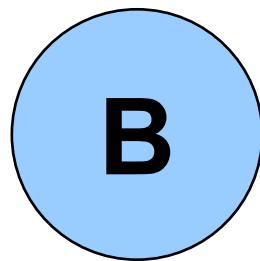
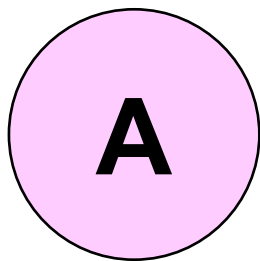


# Research Integrity

## *or How to be a Good Scientist*

Jo Montgomery, Sarah Inglesfield,  
Anne Segonds-Pichon, Laura Biggins and Simon Andrews

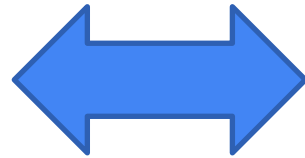
v2025-10



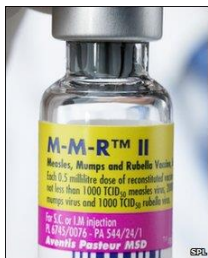
# An illustration

# Research Integrity

MMR = Measles, Mumps and Rubella



# Autism



1998 **Link** MMR autism

Wakefield *et al.*

**n=12** children

<https://pubmed.ncbi.nlm.nih.gov/9500320/>

2010

**Retraction** of the Wakefield *et al.* paper

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2831678/>

2010

Andrew Wakefield found guilty of serious professional **misconduct**

<https://www.bmj.com/content/340/bmj.c593>

2002-2005: **No link** MMR autism

Many more studies

<https://pubmed.ncbi.nlm.nih.gov/15366972/>

<https://pubmed.ncbi.nlm.nih.gov/15877763/>

1999 **No link** MMR autism

Taylor *et al.*

**n=498** children

<https://pubmed.ncbi.nlm.nih.gov/10376617>

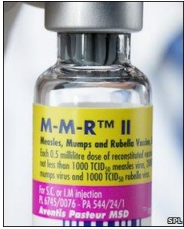
2001: **No link** MMR autism

Dales *et al.*

**n=600-1900** children each year over 14 years

<https://pubmed.ncbi.nlm.nih.gov/11231748/>

# Research Integrity



Retracted article

2010

See the [retraction notice](#)

12 years

> *Lancet*. 1998 Feb 28;351(9103):637-41. doi: 10.1016/s0140-6736(97)11096-0.

## Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children

A J Wakefield<sup>1</sup>, S H Murch, A Anthony, J Linnell, D M Casson, M Malik, M Berelowitz, A P Dhillon, M A Thomson, P Harvey, A Valentine, S E Davies, J A Walker-Smith

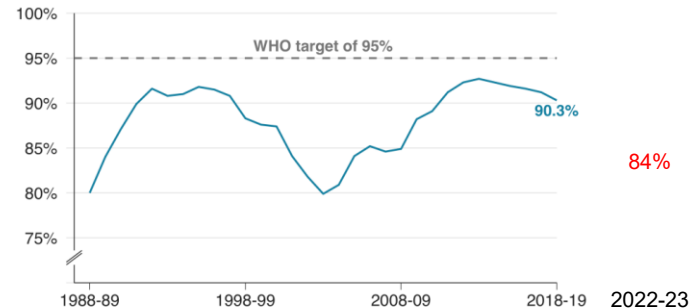
Affiliations + expand

PMID: 9500320 DOI: 10.1016/s0140-6736(97)11096-0

- Scientists and organisations across the world **spent a great deal of time and money** refuting the results of a minor paper in the *Lancet*.
- MMR vaccination: 1995: 95% to 2002: 81 %**
- Measles outbreaks in the UK in 2008 and 2009
- 2020:** Uptake of MMR vaccine: **91%** (still below herd immunity)

### MMR first dose coverage is falling in England

% of children who completed first dose of vaccine at 24 months



Source: Cover of Vaccination Evaluated Rapidly, Public Health England

BBC

# Research Integrity



Rise in measles cases prompts vaccination campaign in England

UKHSA declares national incident as figures suggest more than 3.4 million children have not had MMR jab



Measles cases in the West Midlands have been at their highest level since the mid-1990s.  
Photograph: MedStockPhotos/Alamy

- **2016:** UK declared measles free  
Now lost this status
- Increases in number of measles cases
- **2022-23:** : **84% children in England**
- **(74% in London, vs 90% South West)**
- Need vaccination rate of **95%**
- **Current decline due to:**
  - Vaccine misinformation
  - Not serious?
  - Difficulty accessing appointments
  - Impact of covid pandemic

Obvious examples are obvious, but...





# Outline of the course

# Research Integrity

```
graph TD; RI[Research Integrity] --> Q[Questioning]; RI --> D[Definitions]; RI --> P[Publications]; RI --> M[Misconduct]; RI --> R[Responsibility]; RI --> E[Ethics]; RI --> K[Keeping Track]; RI --> B[Bottom line]; RI --> GS[Good Science]
```

What does it mean?

Questioning

Definitions

Publications

Misconduct

When is it tested?

Responsibility

Keeping Track

Ethics

In Practice

Bottom line

How can we apply it?

Good Science

# What is Research Integrity?

# Research Integrity

## a.k.a. Scientific Integrity

- **Scientific integrity** (From Wikipedia):  
Scientific integrity deals with ‘**best practices**’ or rules of professional practice of researchers.
- Organisation for Economic Co-operation and Development (OECD) report, 2007 **replication (or reproducibility) crisis** and the **fight against scientific misconduct**.

# Research Integrity

## a.k.a. Scientific Integrity

- The **replication crisis** - scientific studies are **difficult or impossible to replicate** or reproduce.

### Psychology:

Open Science Collaboration  
(100 papers from 2012)

36% of the replications yielded significant findings vs 97% in the original studies.

### Cancer Research:

Reproducibility Project: Cancer Biology  
(53 papers from 2010 to 2012)

25% experiments could be reproduced.  
Replication effect sizes were 85% smaller on average than the original findings.

<https://osf.io/82fth/>

<https://www.sciencenews.org/article/cancer-biology-studies-research-replication-reproducibility>

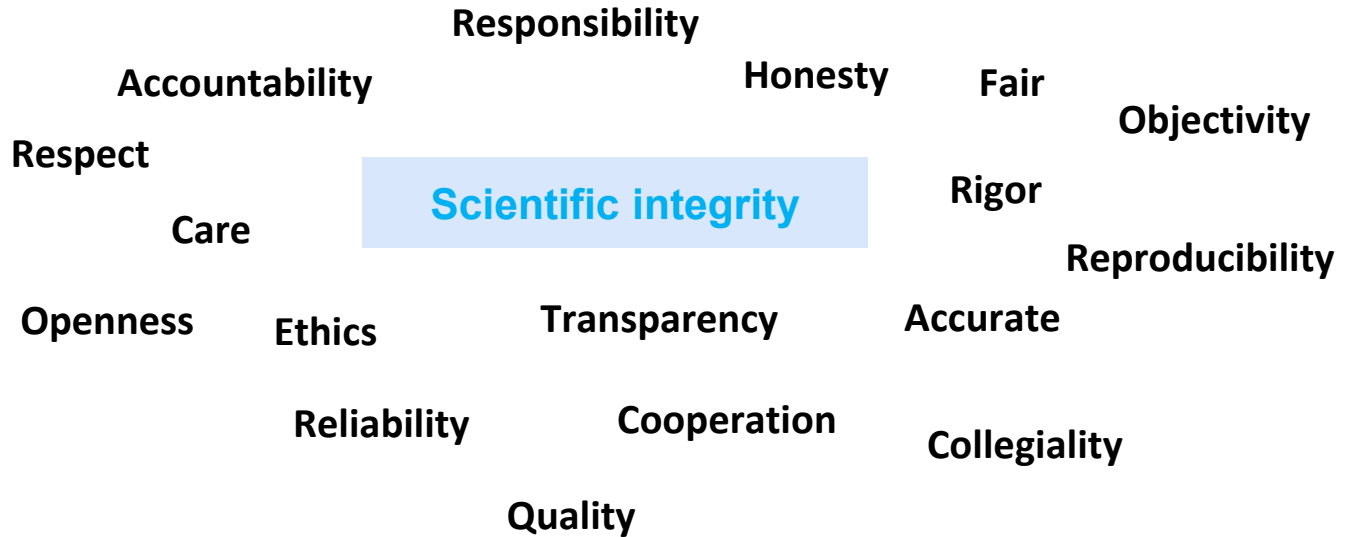
# Which words are most important/synonymous with integrity?

## Exercise



# Research Integrity

## Many words



# Research Integrity

## More than words

Accountability  
Accurate  
Care  
Collegiality  
Cooperation  
Ethics  
Fair  
Honesty  
Objectivity  
Openness  
Quality  
Reliability  
Reproducibility  
Respect  
Responsibility  
Rigor  
Transparency

**Scientific integrity**



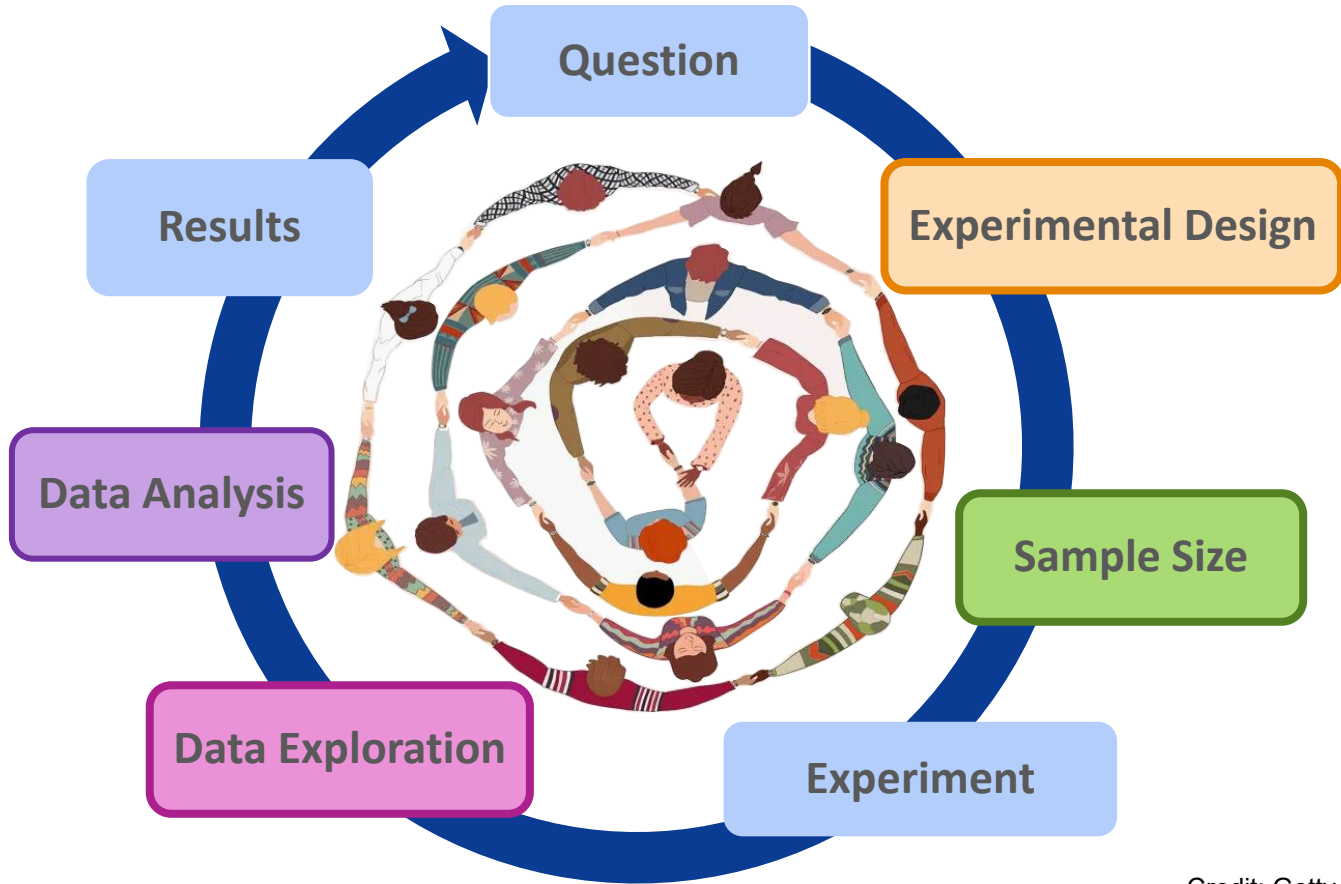
# Research Integrity

Research Integrity means conducting, reviewing and reporting research in a responsible way that allows ourselves and others to have confidence that our research is trustworthy and of the highest quality. It also sets the culture of the Institute as an inclusive, respectful, open and outstanding place to work.

BI-RES-005 RESEARCH INTEGRITY POLICY

# Research Integrity in practice

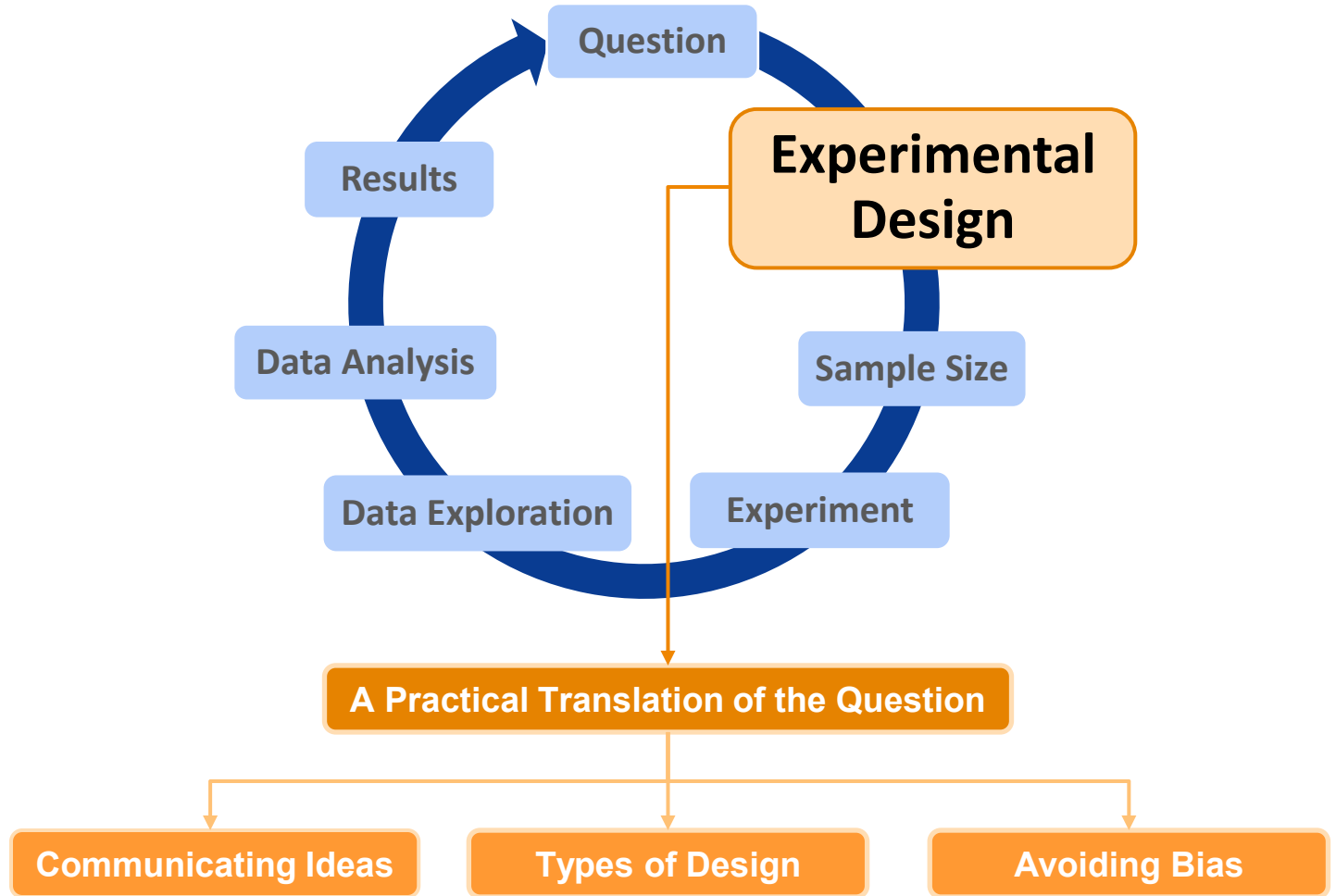
# Research Integrity: In Practice



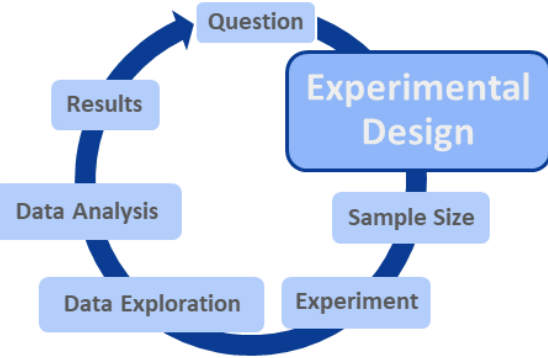
# Research Integrity in practice

## Experimental design

- Accountability
- Accurate
- Care
- Collegiality**
- Cooperation**
- Ethics
- Fair
- Honesty
- Objectivity**
- Openness
- Quality**
- Reliability**
- Reproducibility**
- Respect
- Responsibility
- Rigor**
- Transparency



# Clear Communication

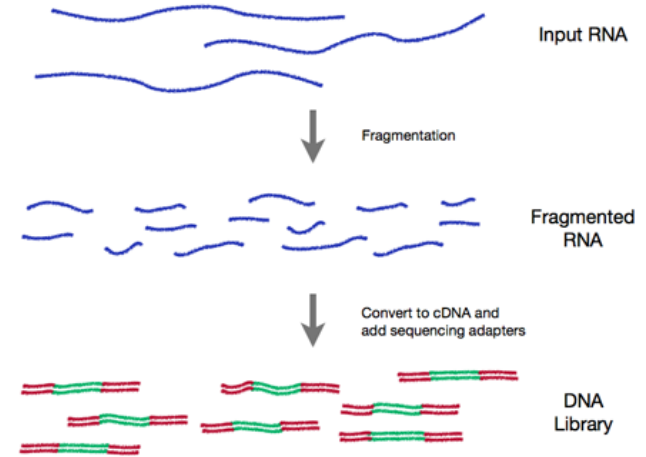


Accountability  
Accurate  
Care  
**Collegiality**  
**Cooperation**  
Ethics  
Fair  
Honesty  
Objectivity  
Openness  
Quality  
Reliability  
Reproducibility  
Respect  
Responsibility  
Rigor  
Transparency



**Library**, Cambridge

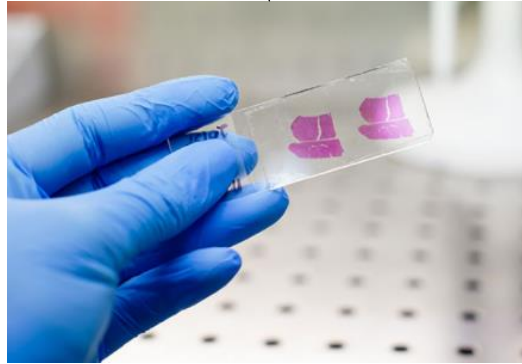
## **Library**



RNA seq **Library**

# Clear Communication

## Sample

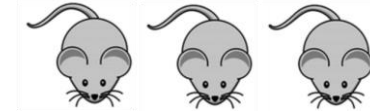


Tissue **Sample**

Control



Treated



**Sample** of mice

- Different words to describe the same data/graphs
- Different traditions in different labs, areas of science

Question

Results

Experimental  
Design

Sample Size

Data Analysis

Data Exploration

Experiment

Accountability

Accurate

Care

**Collegiality**

**Cooperation**

Ethics

Fair

Honesty

Objectivity

Openness

Quality

Reliability

Reproducibility

Respect

Responsibility

Rigor

Transparency

# Appropriate Type of Design

Distinguish real differences from experimental artefacts...

...But experiments can be big and complicated

Seminal Wellcome Trust GWAS Study:

- 14000 cases of 7 diseases & 3000 shared controls
- Each processed at different sites and genotyped on distinct series of plates

Condition 1

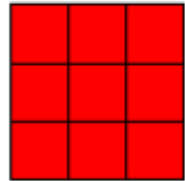


Plate 1

Condition 2

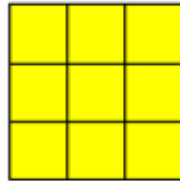


Plate 2

Condition 3 etc...

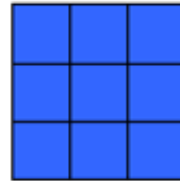
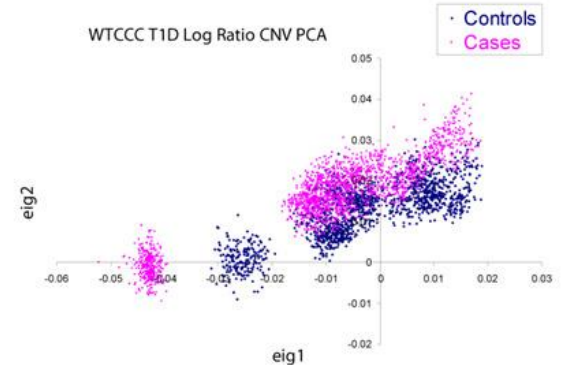
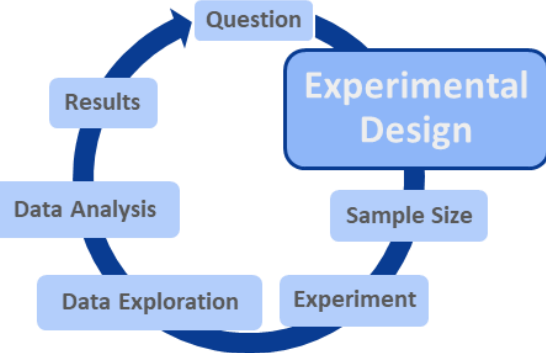


Plate 3



Differences between control and cases are confounded by plate

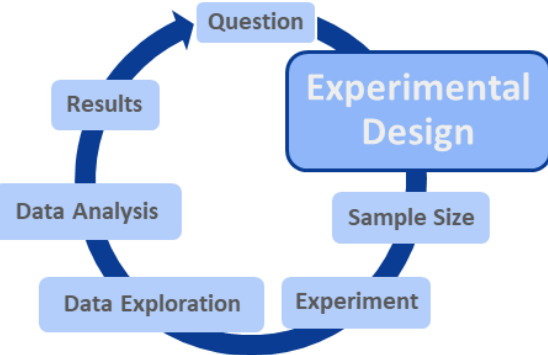


Accountability  
Accurate  
Care  
**Collegiality**  
**Cooperation**  
Ethics  
Fair  
Honesty  
Objectivity  
Openness  
**Quality**  
**Reliability**  
Reproducibility  
Respect  
Responsibility  
**Rigor**  
Transparency





# Appropriate Type of Design



Accountability  
Accurate  
Care  
**Collegiality**  
**Cooperation**  
Ethics  
Fair  
Honesty  
Objectivity  
Openness  
**Quality**  
**Reliability**  
Reproducibility  
Respect  
Responsibility  
**Rigor**  
Transparency

## GenADA multi-site collaborative study:

- 875 Alzheimers patients, 850 controls, 9 sites
- Randomised Block Design

Plate 1

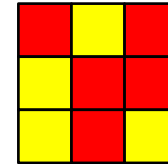
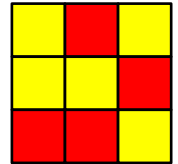
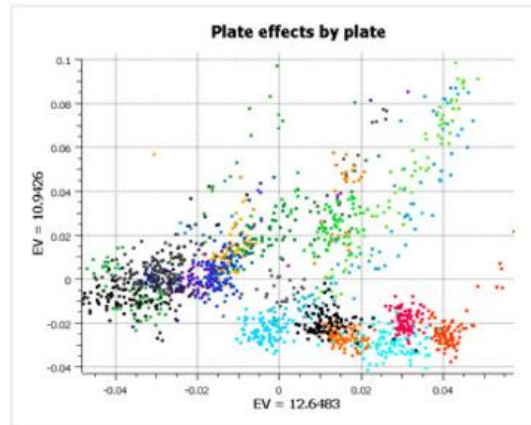


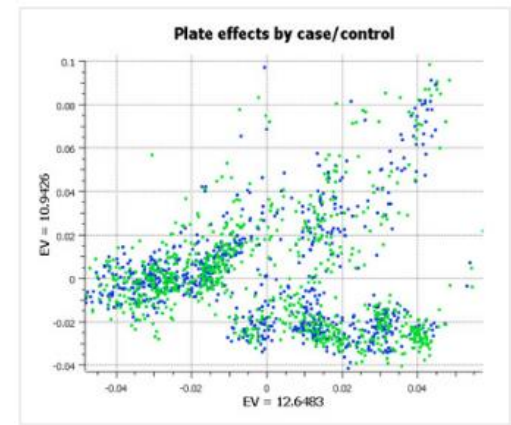
Plate 2 etc...



Still have differences between plates



Doesn't confound the experiment





# Avoiding Bias

Humans are not always good at remaining objective...



**Cognitive Bias = Pareidolia**

**Biases can also impact our experiments**

# Confirmation bias

## Exercise



# A Quick Exercise on Confirmation Bias and Hypothesis Testing

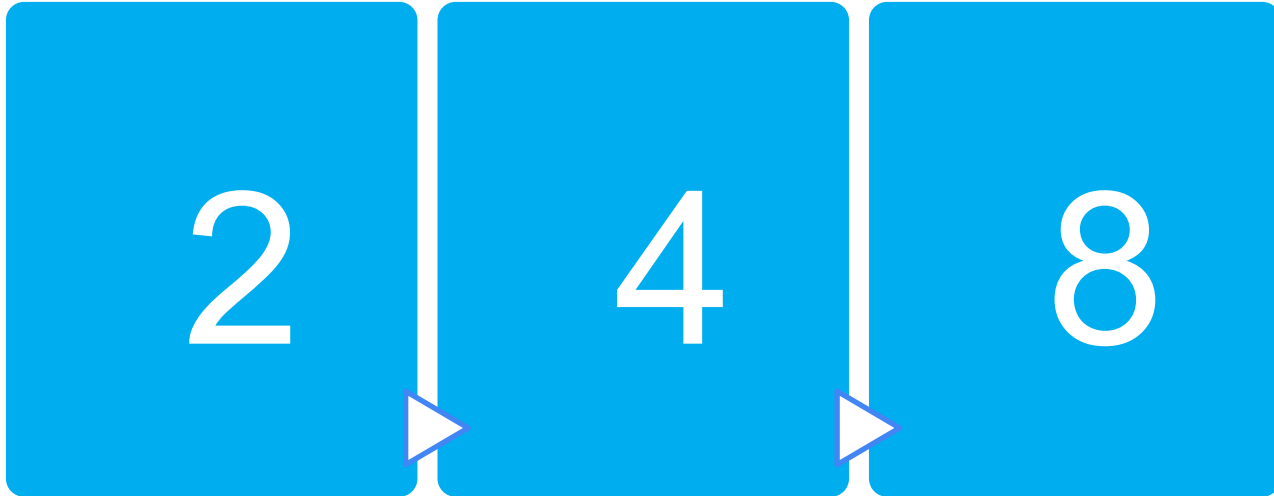
You will be presented with 3 numbers in a sequence.

You need to guess the **rule** that governs the sequence.

You can suggest any 3 numbers you like, and we will tell you whether or not your sequence follows the rule.



# A Quick Exercise on Confirmation Bias and Hypothesis Testing

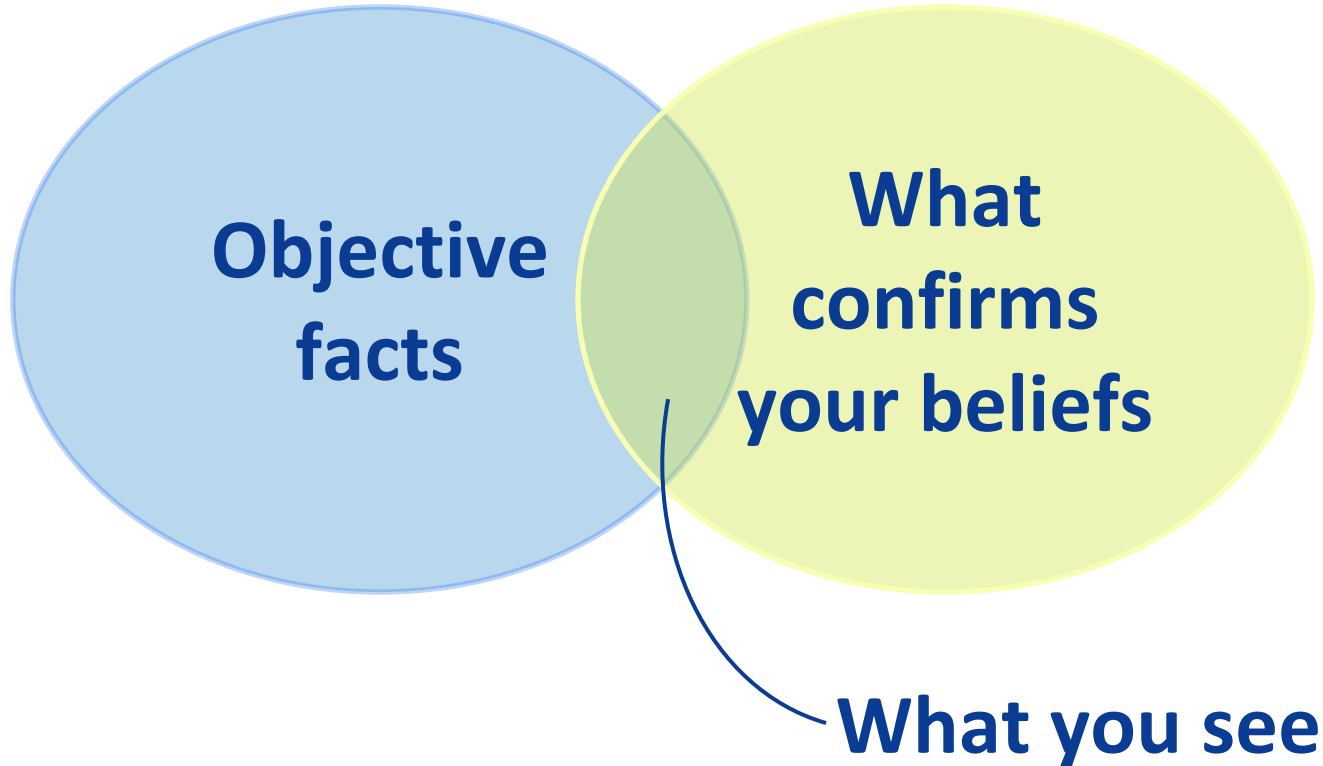


# A Quick Exercise on Confirmation Bias and Hypothesis Testing

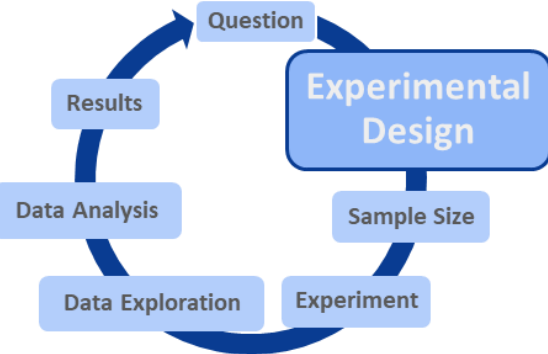
What's the rule?



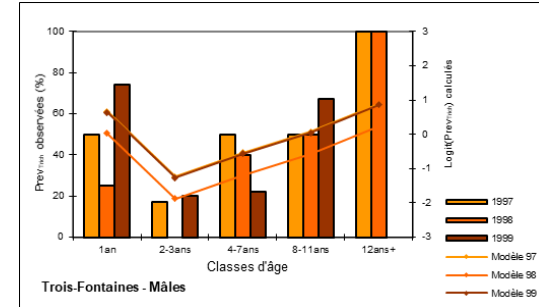
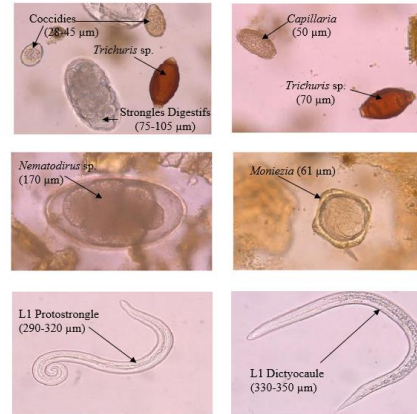
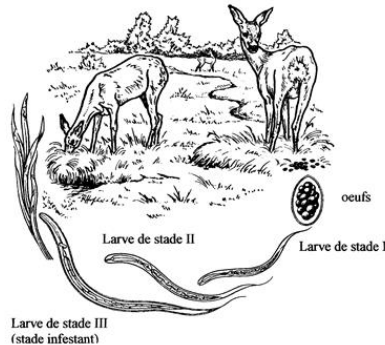
# Confirmation bias



# Experimenter Bias

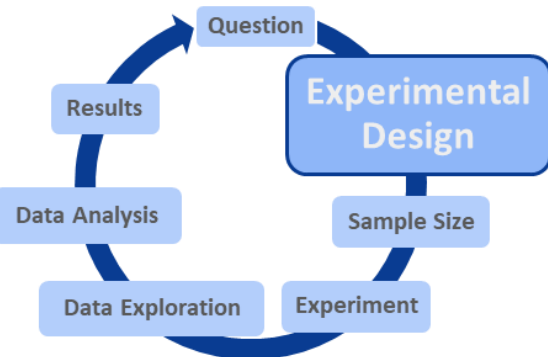


Accountability  
 Accurate  
 Care  
 Collegiality  
 Cooperation  
 Ethics  
 Fair  
 Honesty  
**Objectivity**  
 Openness  
 Quality  
 Reliability  
**Reproducibility**  
 Respect  
 Responsibility  
 Rigor  
 Transparency



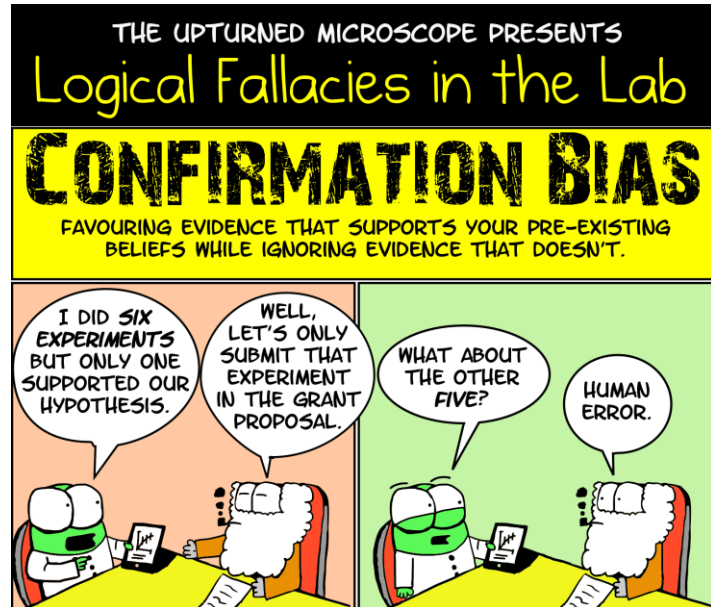
Scientists should be **blind** ... if they want to be **objective**!

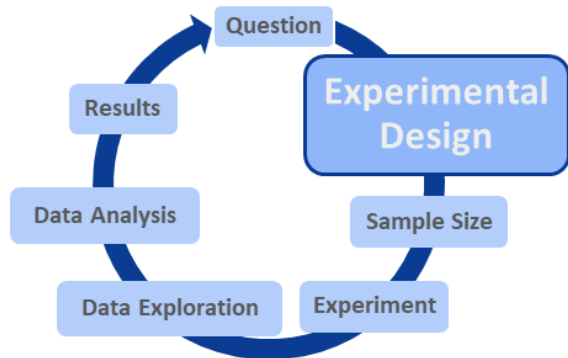




- Accountability
- Accurate
- Care
- Collegiality
- Cooperation
- Ethics
- Fair
- Honesty
- Objectivity**
- Openness
- Quality
- Reliability
- Reproducibility**
- Respect
- Responsibility
- Rigor
- Transparency

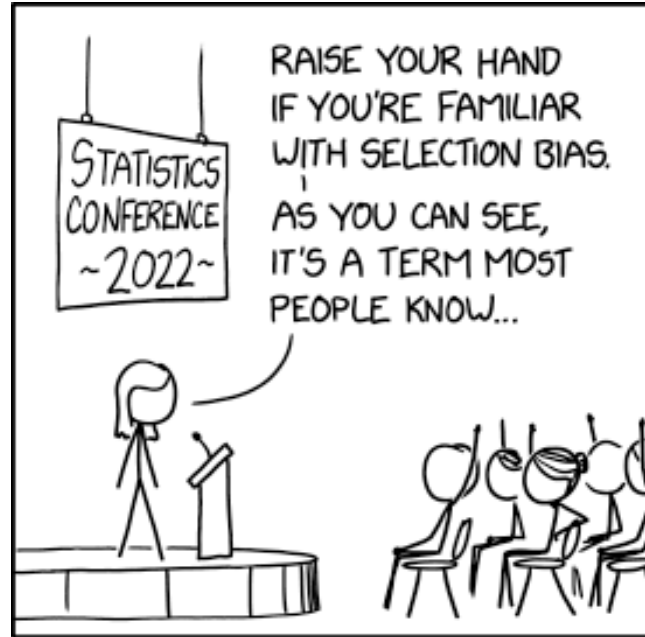
# Blinding is important to avoid Confirmation bias



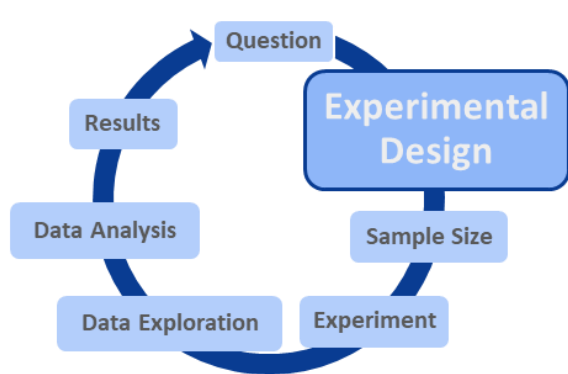


Accountability  
Accurate  
Care  
Collegiality  
Cooperation  
Ethics  
**Fair**  
**Honesty**  
**Objectivity**  
Openness  
Quality  
**Reliability**  
**Reproducibility**  
Respect  
Responsibility  
**Rigor**  
Transparency

# Selection Bias

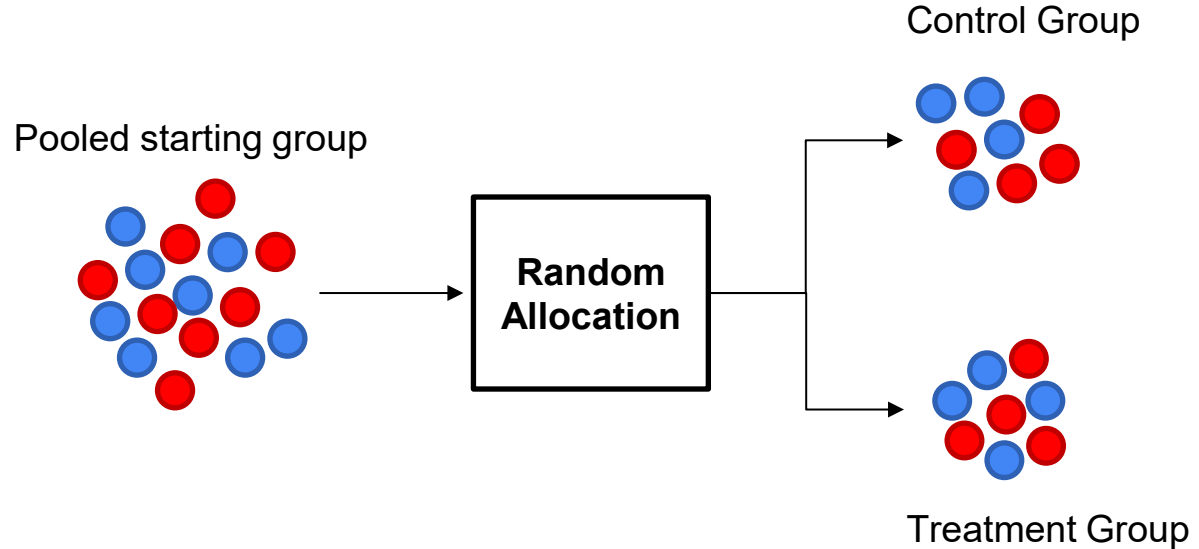


Biases in our sample populations can impact on our conclusions too



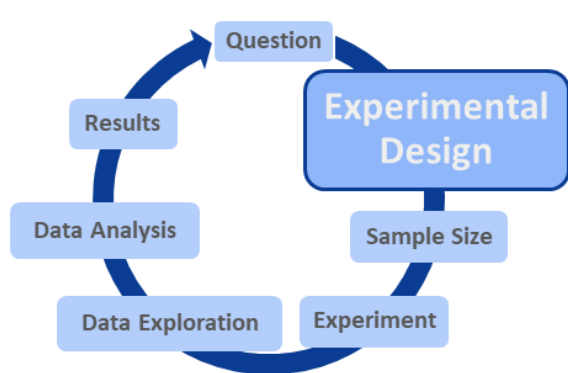
Accountability  
Accurate  
Care  
Collegiality  
Cooperation  
Ethics  
**Fair**  
**Honesty**  
**Objectivity**  
Openness  
Quality  
**Reliability**  
**Reproducibility**  
Respect  
Responsibility  
**Rigor**  
Transparency

# Randomisation



Each experimental unit has equal probability of receiving a treatment

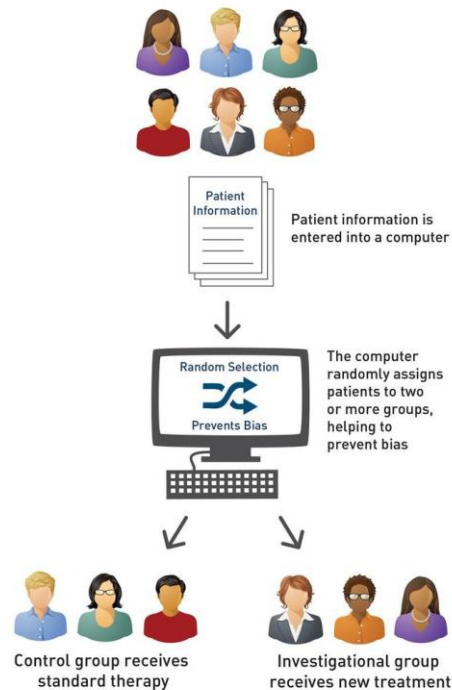
- Minimises Selection Bias
- Reduces systematic differences between groups



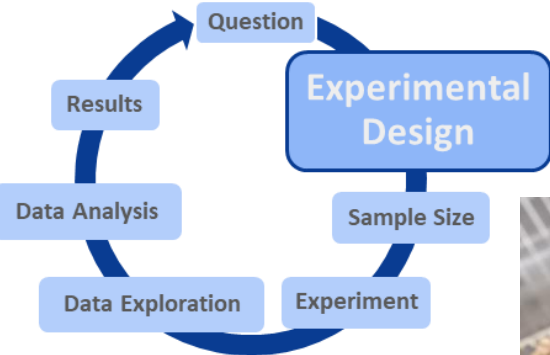
Accountability  
Accurate  
Care  
Collegiality  
Cooperation  
Ethics  
**Fair**  
**Honesty**  
**Objectivity**  
Openness  
Quality  
**Reliability**  
**Reproducibility**  
Respect  
Responsibility  
**Rigor**  
Transparency

# Random Allocation

## CLINICAL TRIALS RANDOMIZATION



# Random Allocation



Accountability  
Accurate  
Care  
Collegiality  
Cooperation  
Ethics  
**Fair**  
**Honesty**  
**Objectivity**  
Openness  
Quality  
**Reliability**  
**Reproducibility**  
Respect  
Responsibility  
**Rigor**  
Transparency



Pick out a mouse at random, first 3 get the treatment

Is this random?

Does it have the potential to introduce bias?

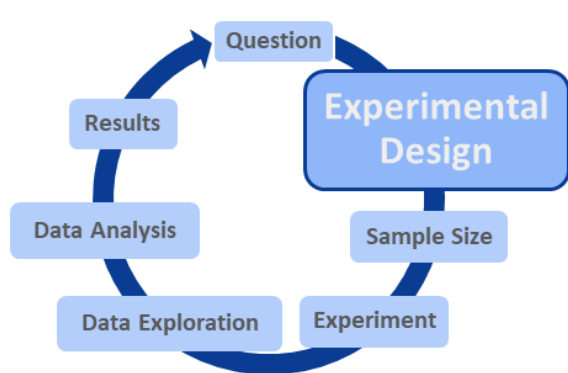
**A**

Biased

**B**

Not Biased

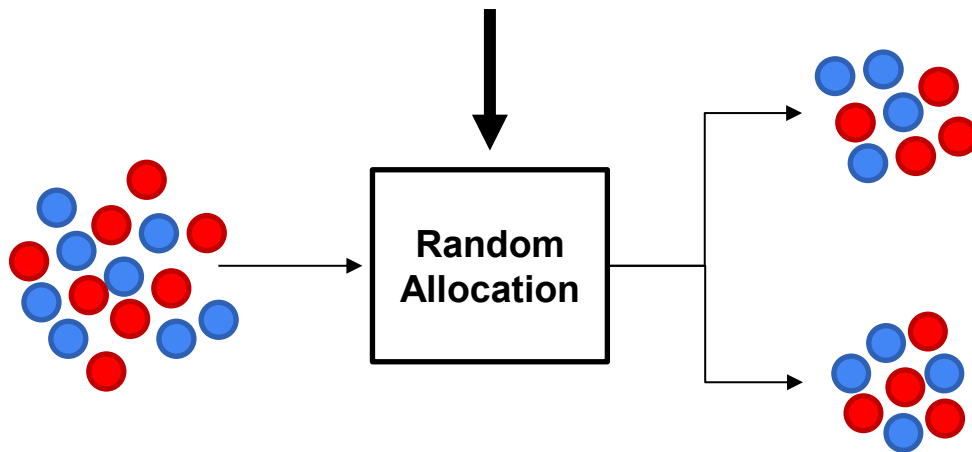




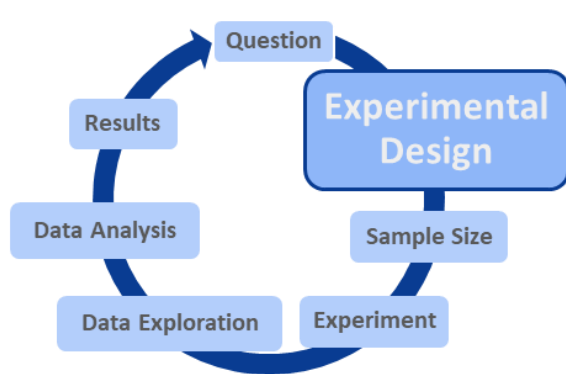
- Accountability
- Accurate
- Care
- Collegiality
- Cooperation
- Ethics
- Fair
- Honesty
- Objectivity
- Openness
- Quality
- Reliability
- Reproducibility
- Respect
- Responsibility
- Rigor
- Transparency

## Randomisation More Broadly

Consider Nuisance Factors



Also consider randomisation **throughout** the experiment



## In a nutshell

### Good experimental design...

- Translates the scientific question into lab work
- Prevents subjectivity
- Reduces effects of nuisance variables

...Is a fair way to do science

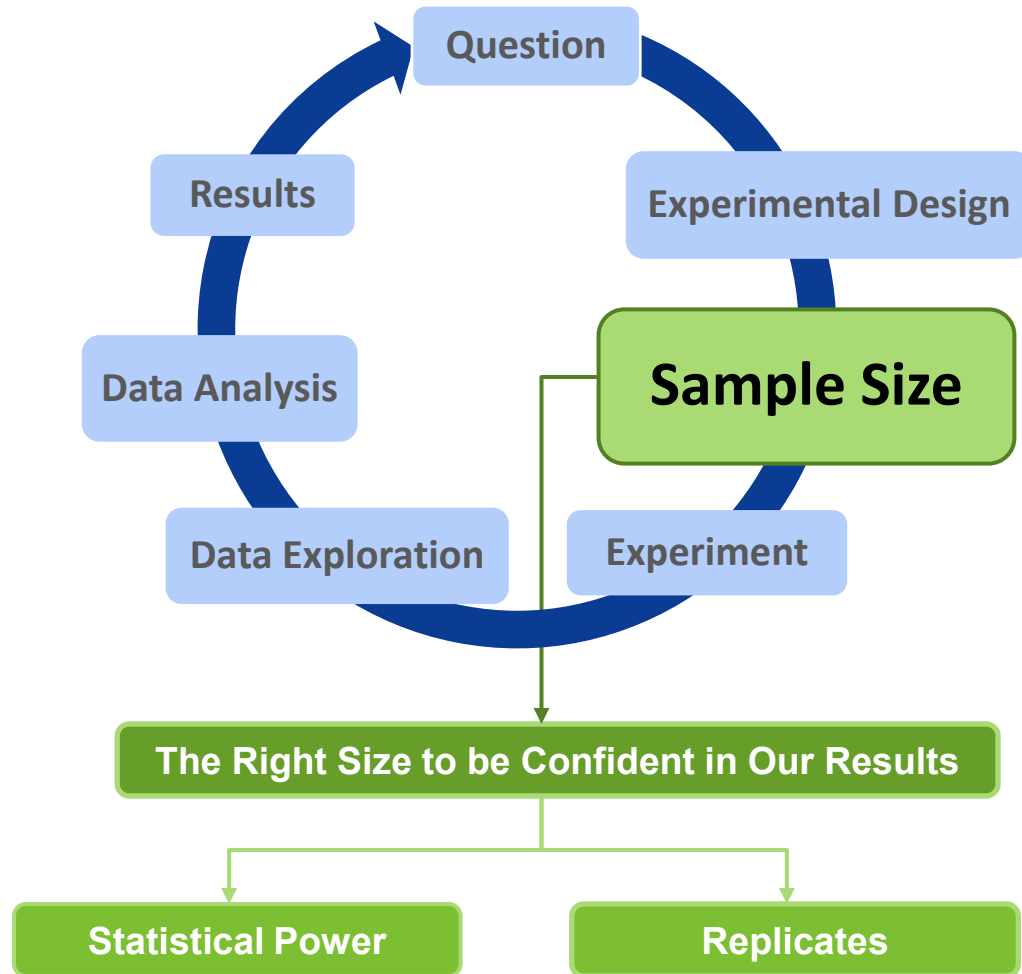
Accountability  
Accurate  
Care  
Collegiality  
Cooperation  
Ethics  
**Fair**  
**Honesty**  
**Objectivity**  
Openness  
Quality  
**Reliability**  
**Reproducibility**  
Respect  
Responsibility  
**Rigor**  
Transparency

# Research Integrity in practice

## Sample size



Accountability  
Accurate  
**Care**  
Collegiality  
Cooperation  
**Ethics**  
Fair  
Honesty  
Objectivity  
Openness  
**Quality**  
**Reliability**  
**Reproducibility**  
**Respect**  
**Responsibility**  
**Rigor**  
Transparency



# Statistical power: an analogy

You send your child into the basement to find a tool. The child comes back and says "it isn't there".

**What do you conclude?**

**In the house**

"If the tool really is in the basement, what are the chances that your child would have found it?"

How long did the child spend looking?



How big is the tool?



or



How messy is the basement?



or



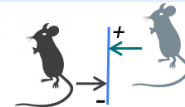
**In the lab**

"If there is a difference between 2 conditions, what are the chances that your experiment will pick it up ( $p < 0.05$ )?"

How many mice do you look at? **Sample size**



How big is the difference? **The absolute effect**



How messy the data are? **Variability**



Accountability

Accurate

Care

Collegiality

Cooperation

Ethics

Fair

Honesty

Objectivity

Openness

Quality

Reliability

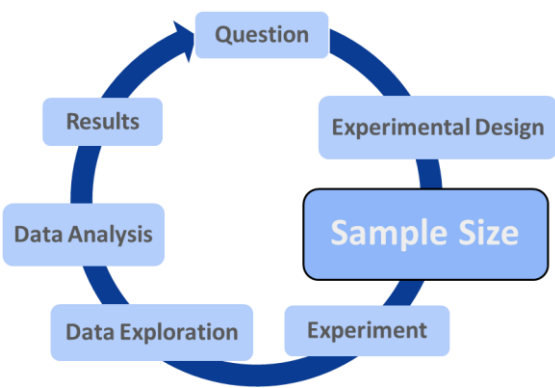
Reproducibility

Respect

Responsibility

Rigor

Transparency



Accountability  
 Accurate  
**Care**  
 Collegiality  
 Cooperation  
**Ethics**  
 Fair  
 Honesty  
 Objectivity  
 Openness  
**Quality**  
**Reliability**  
 Reproducibility  
**Respect**  
**Responsibility**  
**Rigor**  
 Transparency

# The Importance of Statistical Power

“If there is a difference between 2 conditions, what are the chances that your experiment will pick it up ( $p < 0.05$ )?”

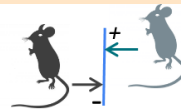
Increase **Sample size** to increase **Power** →

How many mice do you look at? **Sample size**



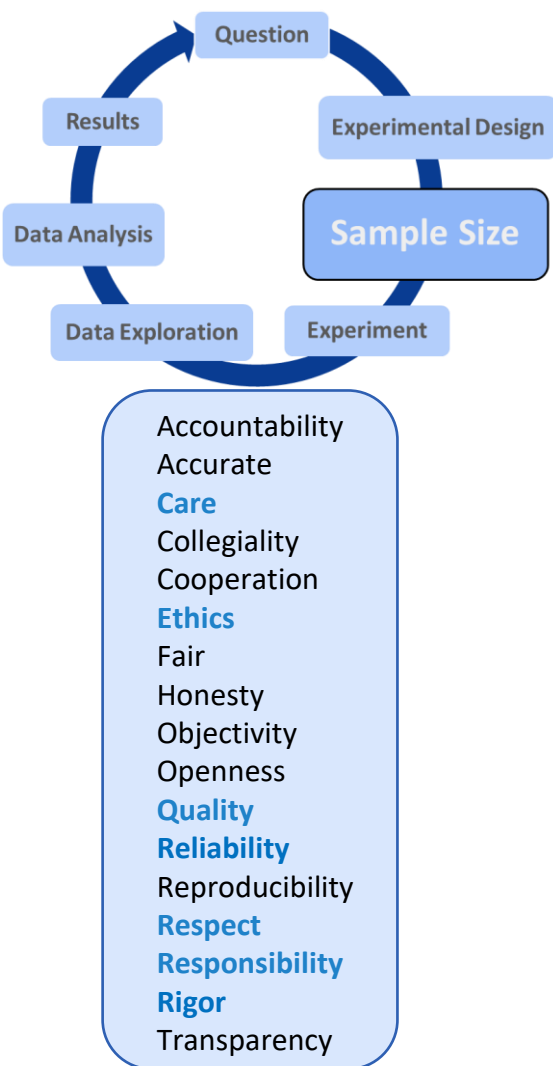
**Little to no control**

How big is the difference? **The absolute effect**



How messy the data are? **Variability**





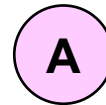
# The Importance of Statistical Power

Low Powered Studies have a greater chance of failing to detect a real effect

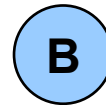
BUT that's all probability...

So some underpowered studies will detect a real effect

Are these results trustworthy?

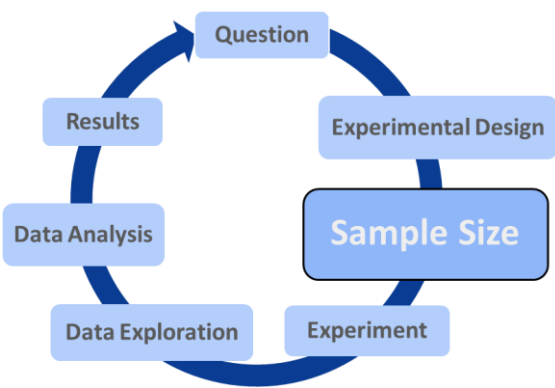


Yes



No



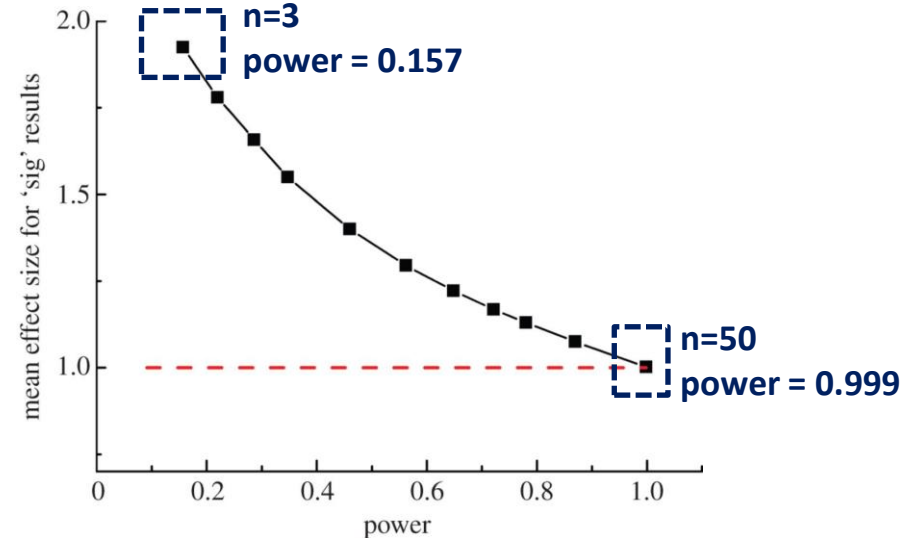
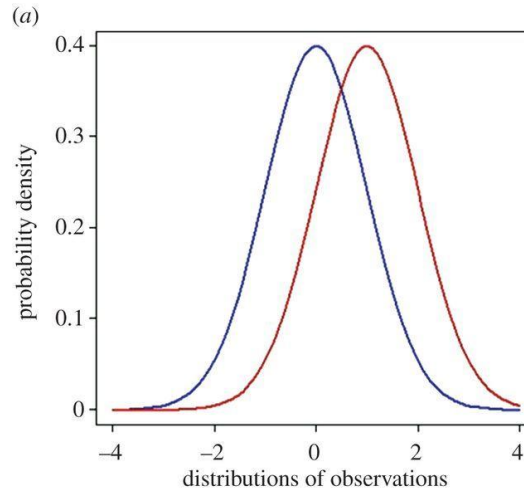


# The Importance of Statistical Power

## The Problem of the “Inflation Effect”...

Accountability  
 Accurate  
**Care**  
 Collegiality  
 Cooperation  
**Ethics**  
 Fair  
 Honesty  
 Objectivity  
 Openness  
**Quality**  
**Reliability**  
 Reproducibility  
**Respect**  
**Responsibility**  
**Rigor**  
 Transparency

### The TRUE effect size = 1

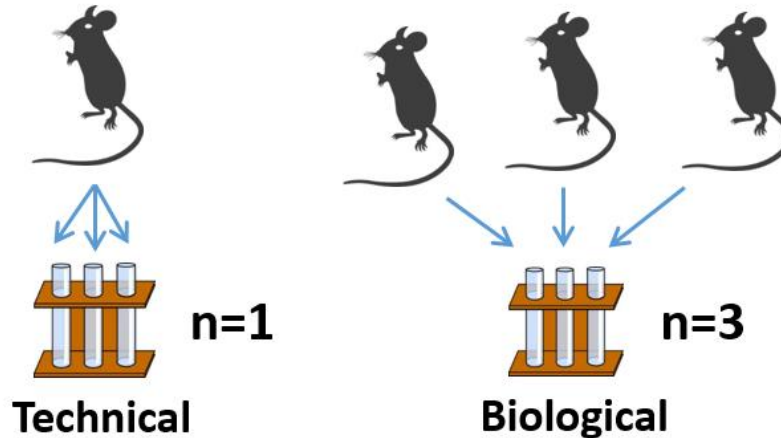


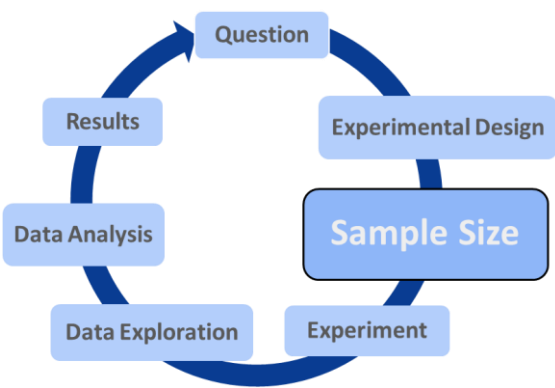


# Defining Replicates

By the way: *replicates* = repeat = sample = library

## Technical versus biological replicates





Accountability  
 Accurate  
**Care**  
 Collegiality  
 Cooperation  
**Ethics**  
 Fair  
 Honesty  
 Objectivity  
 Openness  
**Quality**  
**Reliability**  
 Reproducibility  
**Respect**  
**Responsibility**  
**Rigor**  
 Transparency

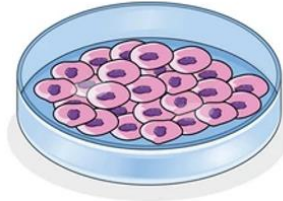
# What is a Biological Replicate in Practice?



1 biological replicate:

**A** 1 mouse

**B** 1 cage



1 biological replicate:

**A** 1 cell

**B** 1 petri dish

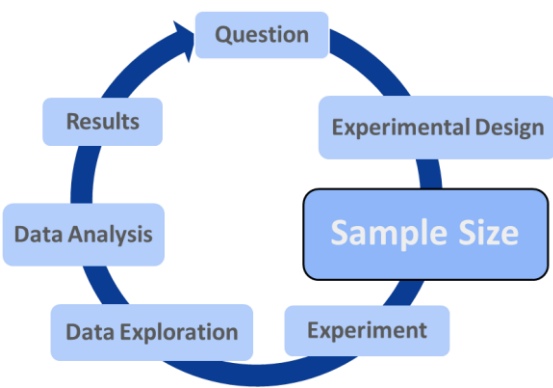


1 biological replicate:

**A** 1 worm

**B** 1 petri dish





## In a nutshell

Sample Size & Power are key to confident results

Underpowered Studies are more likely to:

- Fail to detect real effects
- Overestimate the effect size of detected effects

More biological replicates increase our evidence

How Many?

Formalise with power calculations....

Accountability  
Accurate  
**Care**  
Collegiality  
Cooperation  
**Ethics**  
Fair  
Honesty  
Objectivity  
Openness  
**Quality**  
**Reliability**  
Reproducibility  
**Respect**  
**Responsibility**  
**Rigor**  
Transparency

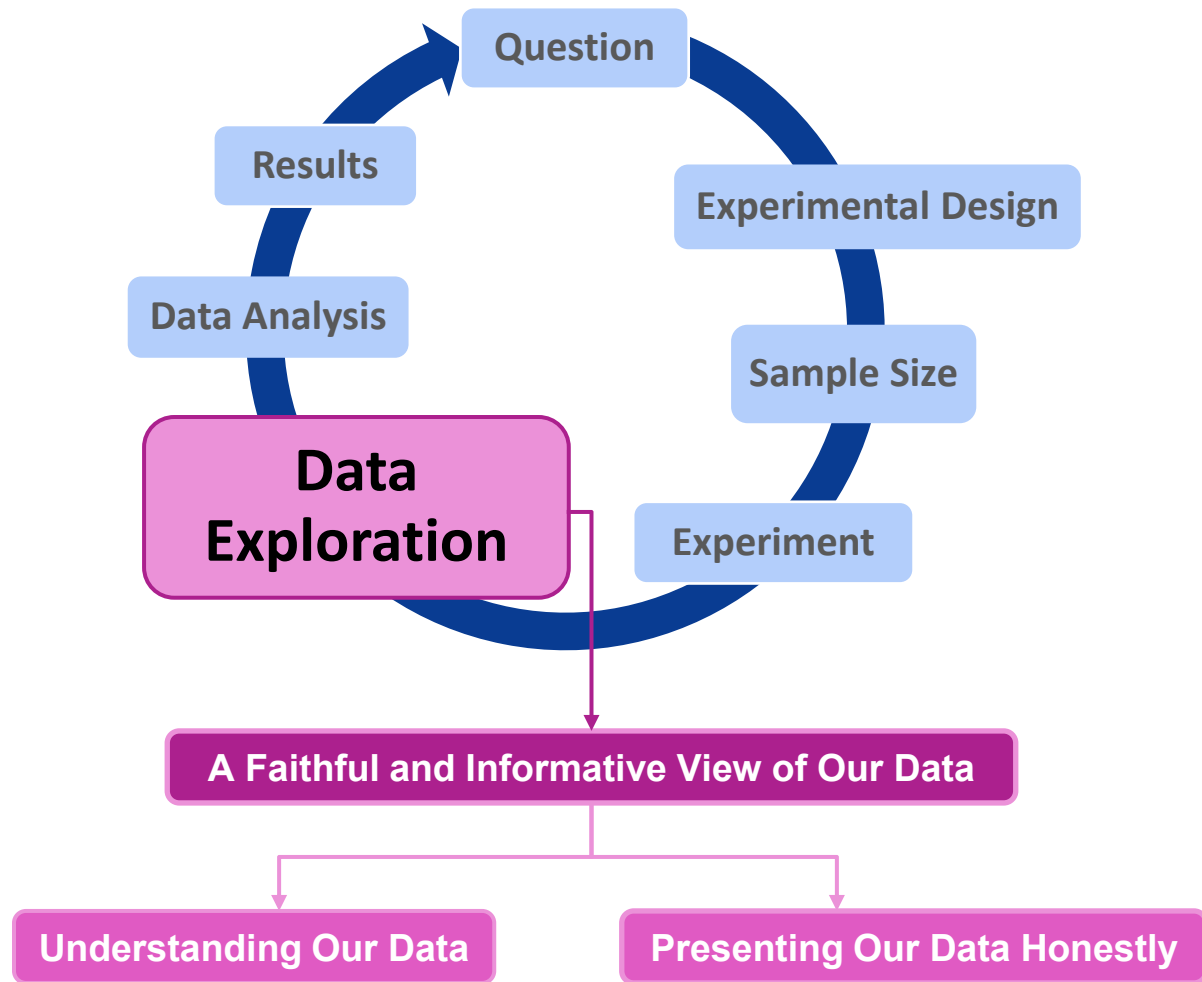




# Research Integrity in practice

## Data Exploration & Data Analysis

- Accountability
- Accurate**
- Care
- Collegiality**
- Cooperation**
- Ethics
- Fair**
- Honesty**
- Objectivity
- Openness**
- Quality**
- Reliability**
- Reproducibility
- Respect
- Responsibility**
- Rigor
- Transparency**



Accountability  
Accurate  
Care  
Collegiality  
Cooperation  
Ethics  
Fair  
Honesty  
Objectivity  
Openness  
Quality  
Reliability  
Reproducibility  
Respect  
Responsibility  
Rigor  
Transparency

# Data Exploration

Understanding Our Data:

- The Biology
- The Quality

**Builds Our Understanding and Confidence**



# Data Exploration: Understanding Our Data

Accountability  
Accurate  
Care  
Collegiality  
Cooperation  
**Ethics**  
**Fair**  
**Honesty**  
**Objectivity**  
**Openness**  
**Quality**  
**Reliability**  
**Reproducibility**  
Respect  
**Responsibility**  
**Rigor**  
**Transparency**

Often we summarise our data to key values

N	182
X Mean	54.26
Y Mean	47.83
X SD	16.76
Y SD	26.93
Correlation	-0.06

Can be really useful...

...And also really not!

# Data Exploration: Anscombe's Quartet

Accountability  
Accurate  
Care  
Collegiality  
Cooperation  
**Ethics**  
**Fair**  
**Honesty**  
**Objectivity**  
**Openness**  
**Quality**  
**Reliability**  
**Reproducibility**  
Respect  
**Responsibility**  
**Rigor**  
**Transparency**

"A computer should make both calculations *and* graphs"

Dataset 1		Dataset 2		Dataset 3		Dataset 4	
X1	Y1	X2	Y2	X3	Y3	X4	Y4
10.00	8.04	10.00	9.14	10.00	7.46	8.00	6.58
8.00	6.95	8.00	8.14	8.00	6.77	8.00	5.76
13.00	7.58	13.00	8.74	13.00	12.74	8.00	7.71
9.00	8.81	9.00	8.77	9.00	7.11	8.00	8.84
11.00	8.33	11.00	9.26	11.00	7.81	8.00	8.47
14.00	9.96	14.00	8.10	14.00	8.84	8.00	7.04
6.00	7.24	6.00	6.13	6.00	6.08	8.00	5.25
4.00	4.26	4.00	3.10	4.00	5.39	19.00	12.50
12.00	10.84	12.00	9.13	12.00	8.15	8.00	5.56
7.00	4.82	7.00	7.26	7.00	6.42	8.00	7.91
5.00	5.68	5.00	4.74	5.00	5.73	8.00	6.89

4 datasets

Each consisting of X and Y variable

All 4 datasets have the same summary statistics...

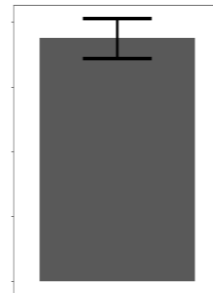


**Francis  
Anscombe**

# Data Exploration: Anscombe's Quartet

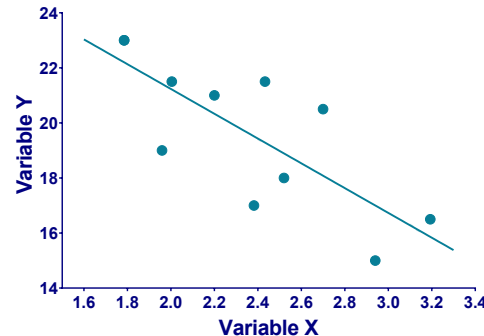
The average and spread of the conditions appears the same

	Dataset 1		Dataset 2		Dataset 3		Dataset 4	
	X1	Y1	X2	Y2	X3	Y3	X4	Y4
N	11	11	11	11	11	11	11	11
Mean	9.00	7.50	9.00	7.50	9.00	7.50	9.00	7.50
STD	3.31	2.03	3.31	2.03	3.31	2.03	3.31	2.03
SEM	1.00	0.612	1.00	0.612	1.00	0.612	1.00	0.612



The relationship between X & Y can be described the same

	Correlation (r)	Line of Best Fit
x1 vs. y1	$r = 0.8164$	$Y = 0.5001 * X + 3.000$
x2 vs. y2	$r = 0.8162$	$Y = 0.5000 * X + 3.001$
x3 vs. y3	$r = 0.8163$	$Y = 0.4997 * X + 3.002$
x4 vs. y4	$r = 0.8165$	$Y = 0.4999 * X + 3.002$



Accountability

Accurate

Care

Collegiality

Cooperation

Ethics

Fair

Honesty

Objectivity

Openness

Quality

Reliability

Reproducibility

Respect

Responsibility

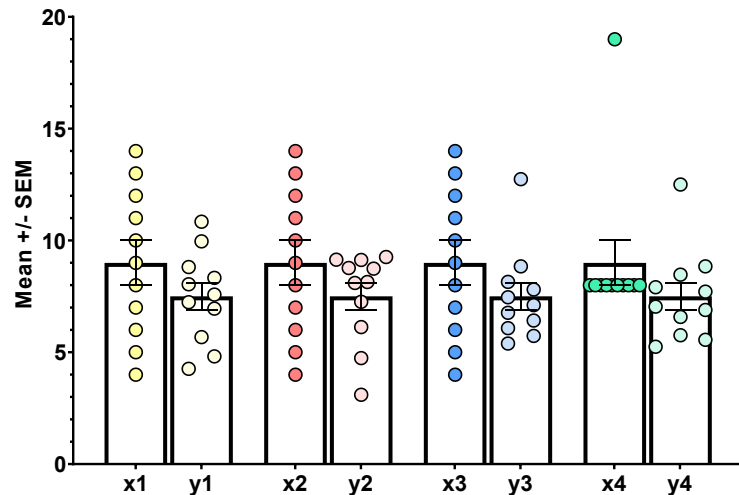
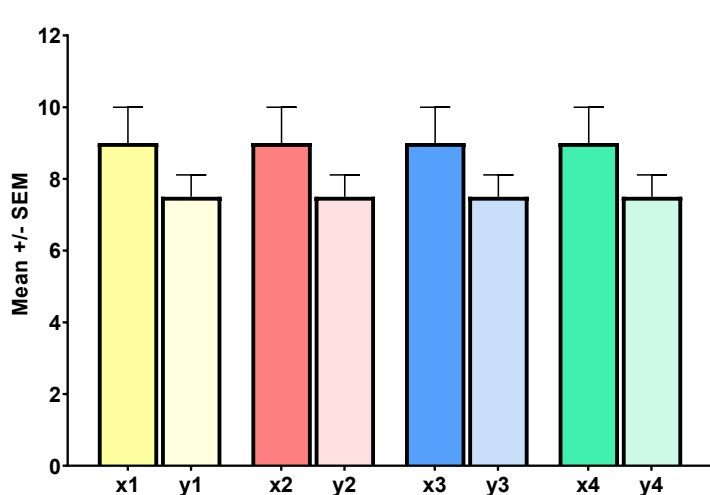
Rigor

Transparency

# Data Exploration: Anscombe's Quartet

Accountability  
Accurate  
Care  
Collegiality  
Cooperation  
Ethics  
Fair  
Honesty  
Objectivity  
Openness  
Quality  
Reliability  
Reproducibility  
Respect  
Responsibility  
Rigor  
Transparency

How does the data behave within the groups?



We're still not getting a complete view of the data  
What about the relationship between X & Y?



# Data Exploration: Anscombe's Quartet

Accountability

Accurate

Care

Collegiality

Cooperation

Ethics

Fair

Honesty

Objectivity

Openness

Quality

Reliability

Reproducibility

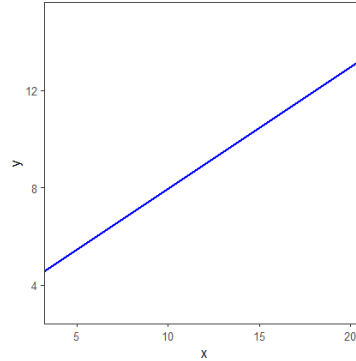
Respect

Responsibility

Rigor

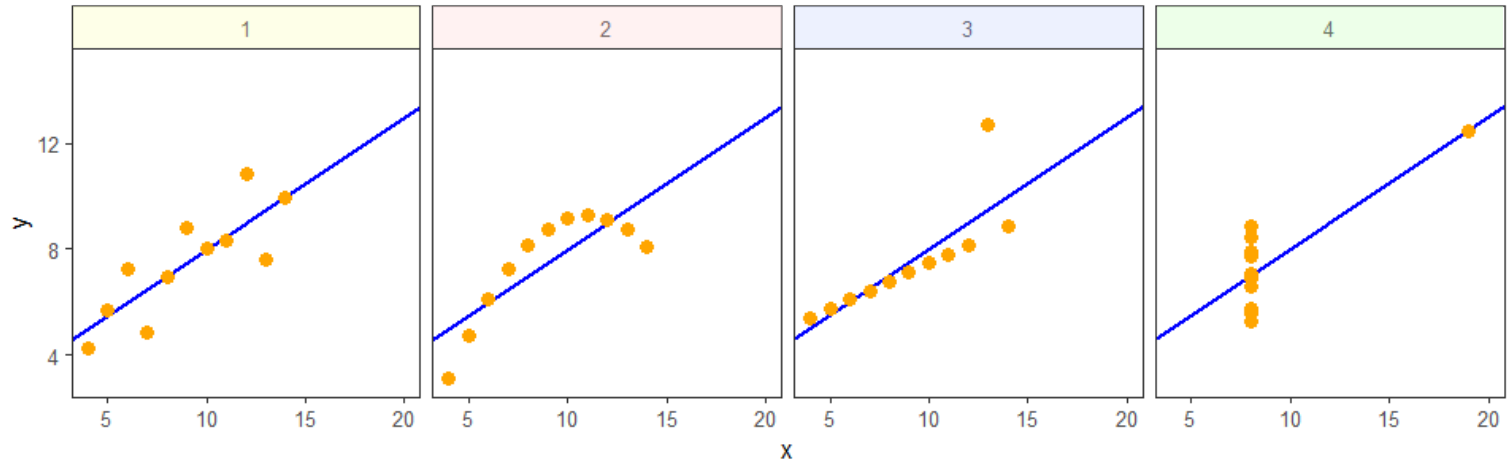
Transparency

What about the relationship between X & Y?



$r = 0.816$

$$Y = 0.500 * X + 3.00$$

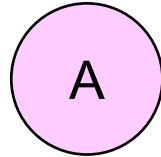


# Can You Predict The Data Structure?

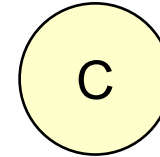
We already know what the data will look like...

Accountability  
Accurate  
Care  
Collegiality  
Cooperation  
**Ethics**  
**Fair**  
**Honesty**  
**Objectivity**  
**Openness**  
**Quality**  
**Reliability**  
**Reproducibility**  
Respect  
**Responsibility**  
**Rigor**  
**Transparency**

Line

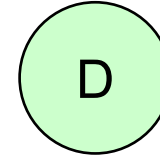
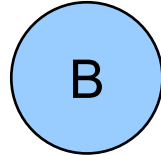


N	182
X Mean	54.26
Y Mean	47.83
X SD	16.76
Y SD	26.93
Correlation	-0.06



Unstructured

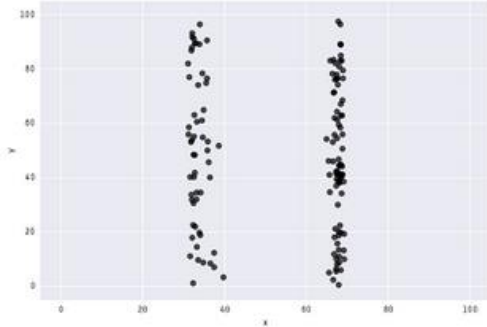
Circle



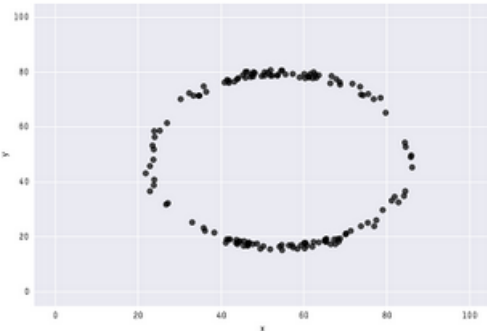
I Don't Know...  
A Dinosaur!



# The Datasaurus Dozen



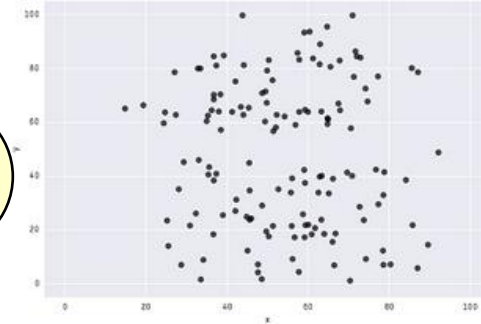
A



B

N	182
X Mean	54.26
Y Mean	47.83
X SD	16.76
Y SD	26.93
Correlation	-0.06

C



D

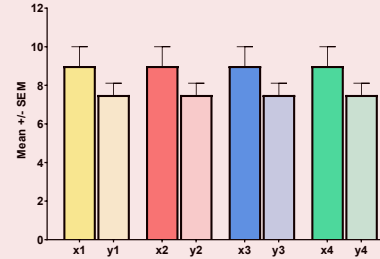


And More!

# Data Exploration: Seeing is Believing!

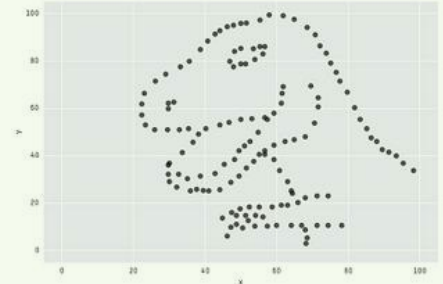
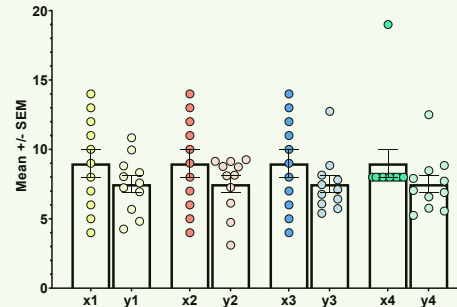
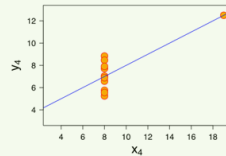
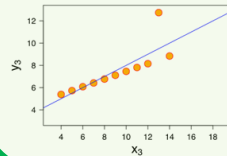
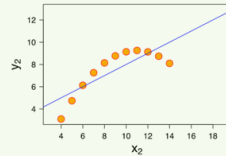
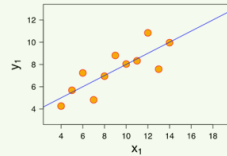
Beware of statistical or graphical summaries...

	Correlation (r)	Line of Best Fit
x1 vs. y1	r = 0.8164	Y = 0.5001*X + 3.000
x2 vs. y2	r = 0.8162	Y = 0.5000*X + 3.001
x3 vs. y3	r = 0.8163	Y = 0.4997*X + 3.002
x4 vs. y4	r = 0.8165	Y = 0.4999*X + 3.002



N	182
X Mean	54.26
Y Mean	47.83
X SD	16.76
Y SD	26.93
Correlation	-0.06

...without proper exploration & visualisation!



Accountability

Accurate

Care

Collegiality

Cooperation

Ethics

Fair

Honesty

Objectivity

Openness

Quality

Reliability

Reproducibility

Respect

Responsibility

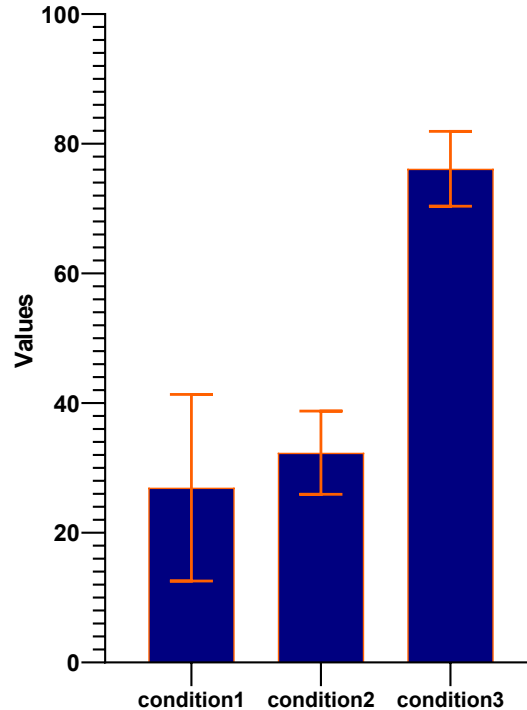
Rigor

Transparency

# Data Exploration Exercise



# What do you think of this graph?

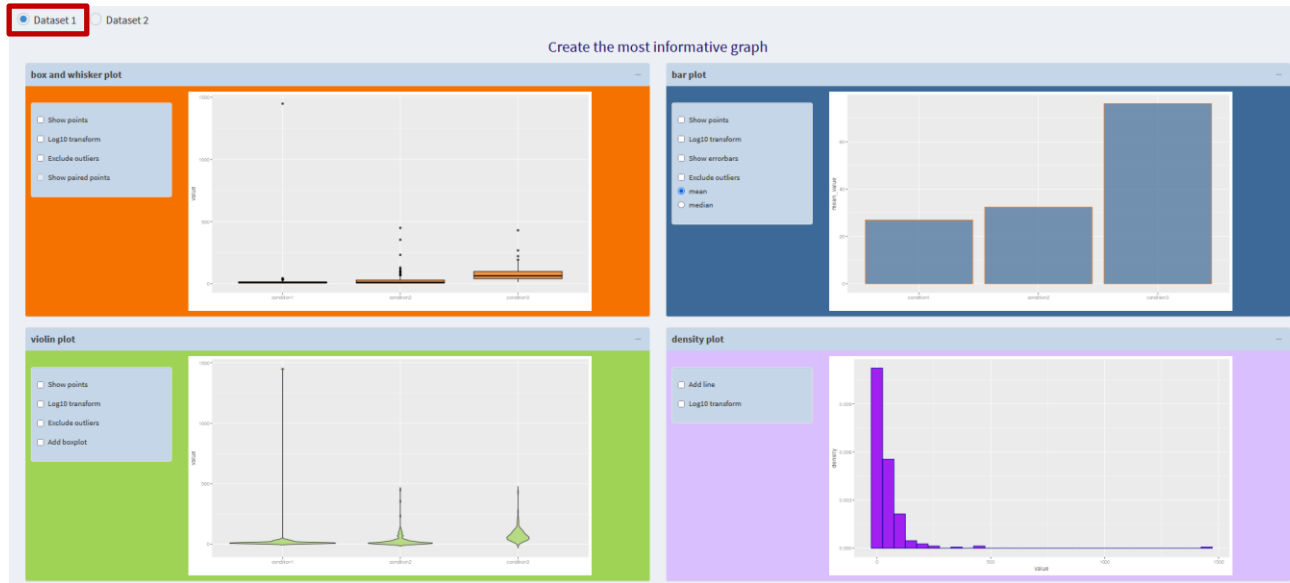


Values from 3 conditions from one experiment



# Representing Dataset 1: Create the most informative graph

<https://tinyurl.com/RldataExp> **Make sure you are looking at Dataset 1**

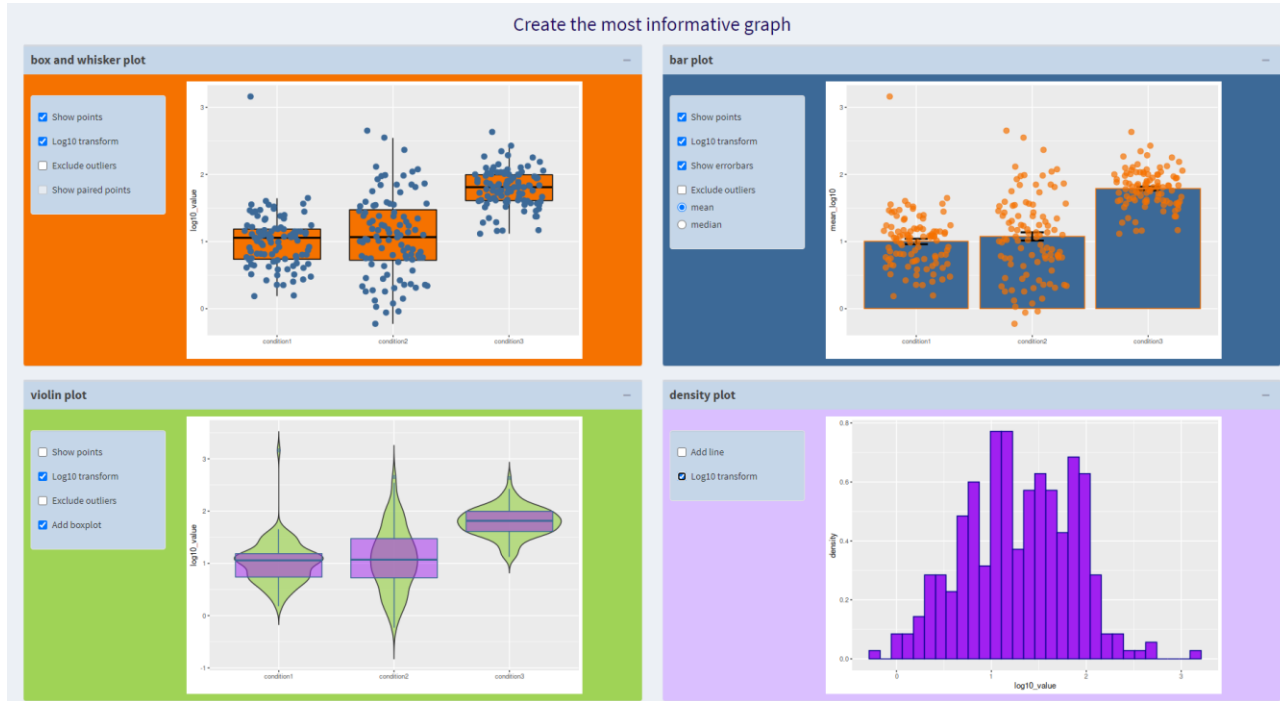


Plot for initial exploration of the data?  
Plot for presentation/ publication?



# Representing Dataset 1:

## Create the most informative graph



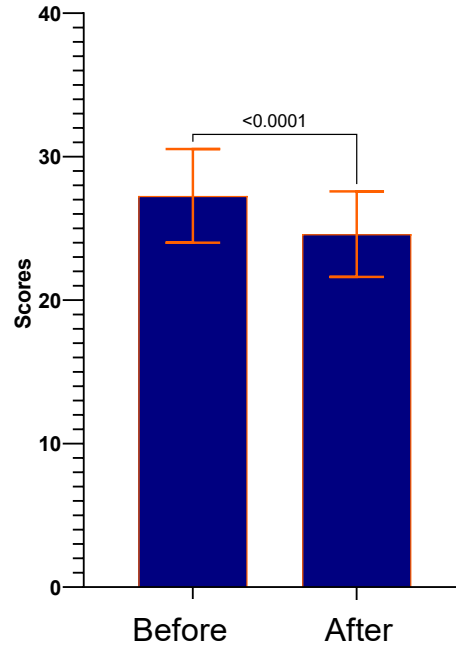
Which plots did we choose and why?

Any differences between initial exploration and presentation?





# What do you think of this graph?



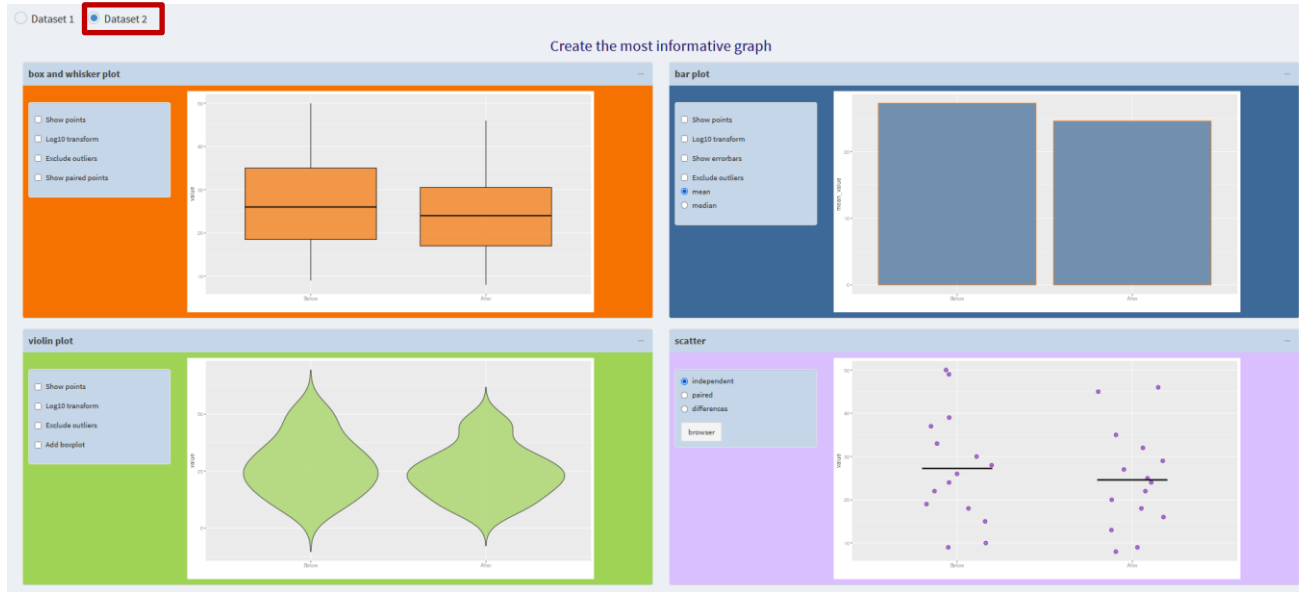
Values for before-after treatment from 4 experiments.



# Representing Dataset 2: Create the most informative graph

<https://tinyurl.com/RIDataExp>

**Make sure you are looking at Dataset 2**

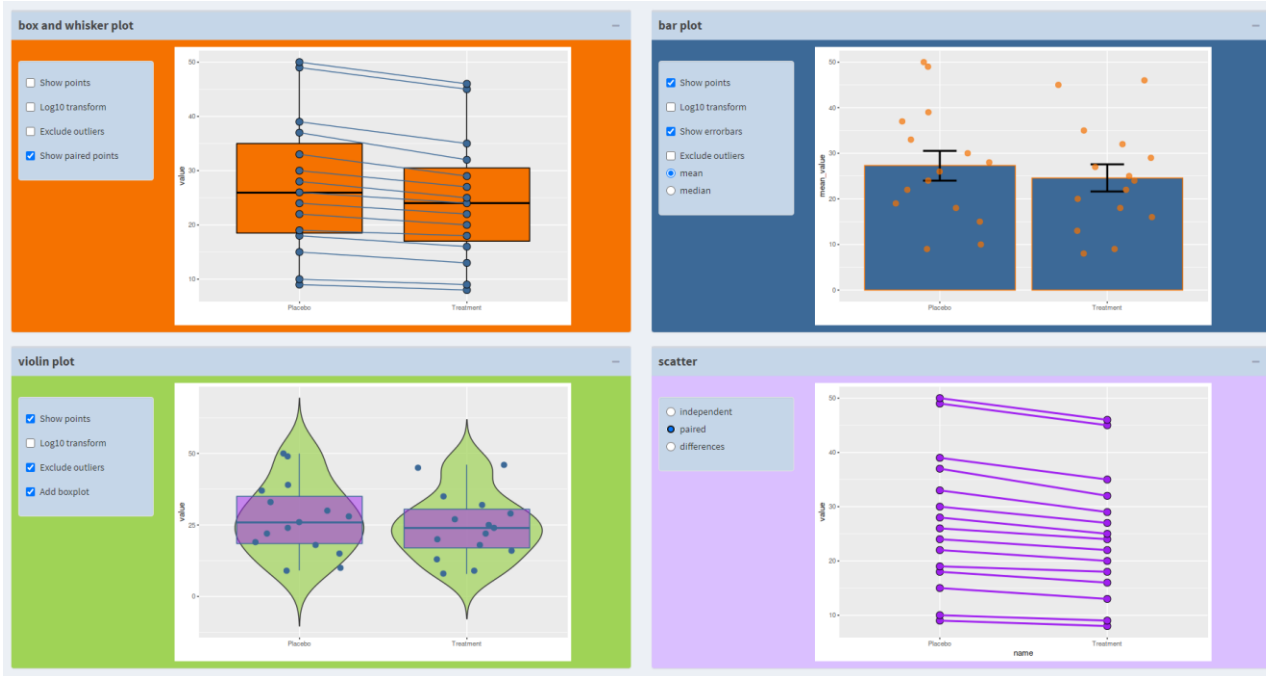


Create a plot to best represent dataset 2  
Think about experimental design and statistics



# Representing Dataset 2:

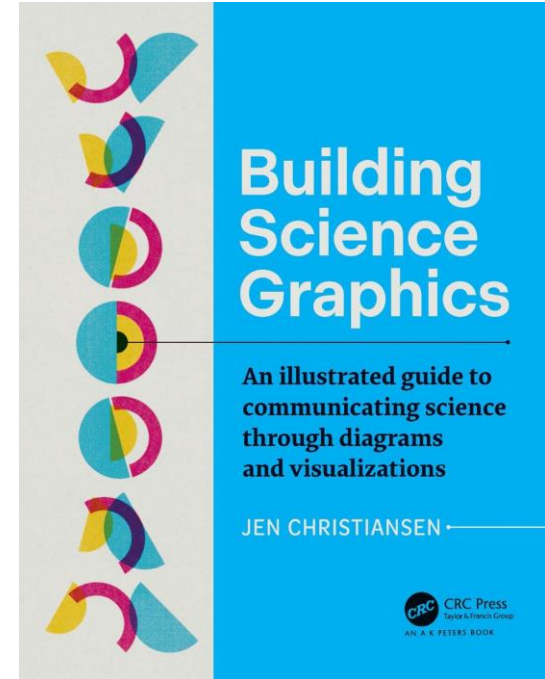
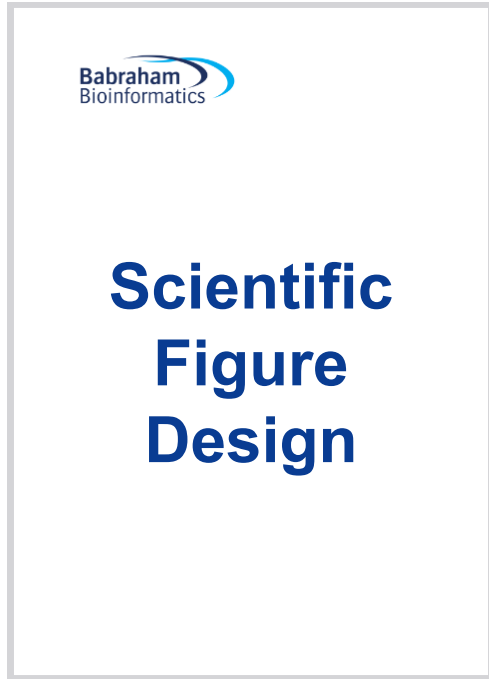
## Create the most informative graph



Which plots did we choose and why?  
What do you think about the stats now?



# An Aside to Help with Better Figure Design...

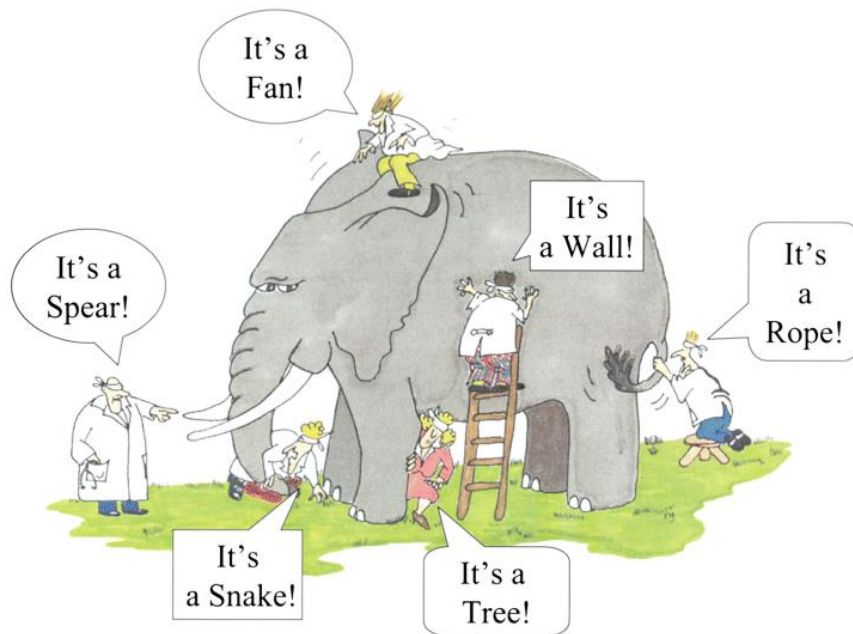


# Data exploration

## What can go wrong if we don't do it?!

# Data exploration

## Less Exploring more Assuming!

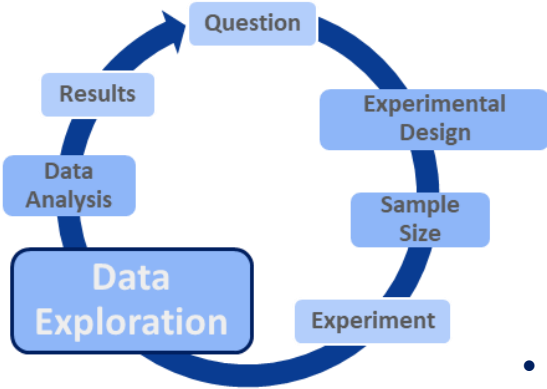


**Risk missing the actual story the data is telling...**

# Data Exploration...



...Suspicious Summaries



### Accountability

Accurate  
Care  
Collegiality  
Cooperation

### Ethics

Fair

Honesty

Objectivity

Openness

Quality

Reliability

Reproducibility

Respect

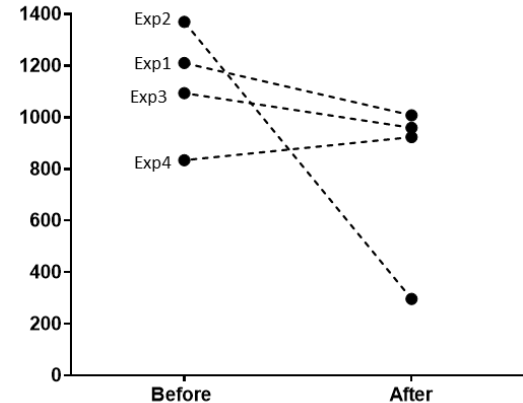
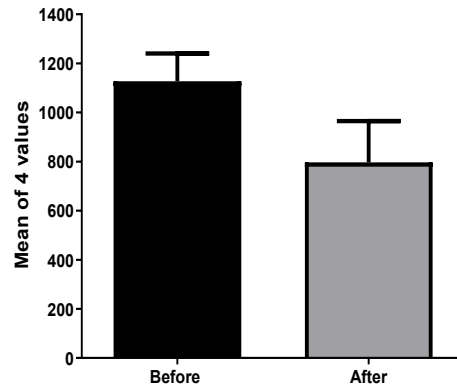
Responsibility

Rigor

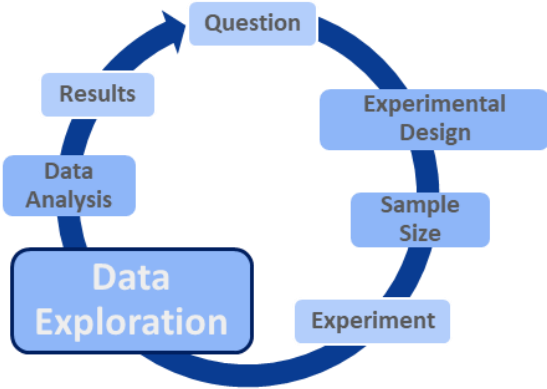
Transparency

# Example 1

- Four experiments: Before-After treatment effect on a variable of interest.
- Hypothesis: Treatment will decrease the levels of the variable of interest







## Example 2

### Accountability

Accurate  
Care  
Collegiality  
Cooperation

### Ethics

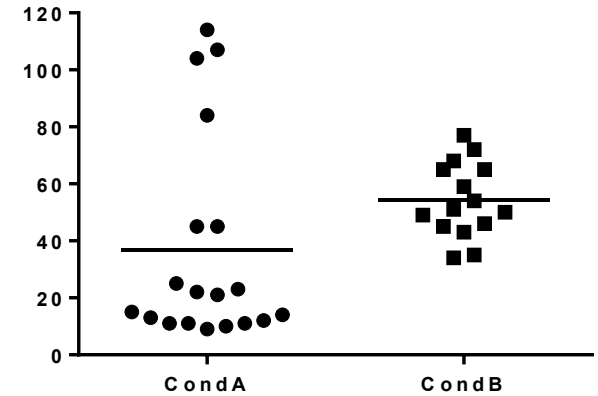
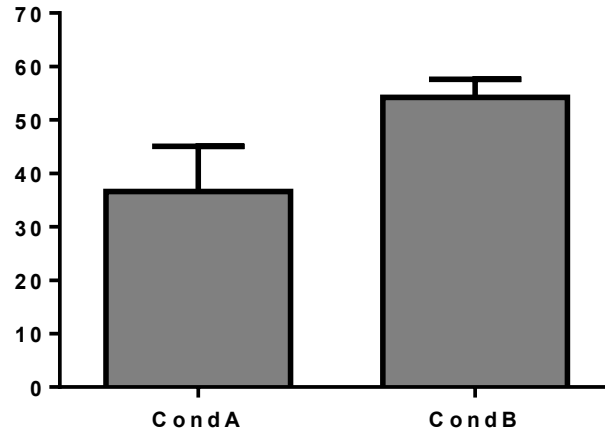
Fair  
Honesty  
Objectivity  
Openness  
Quality

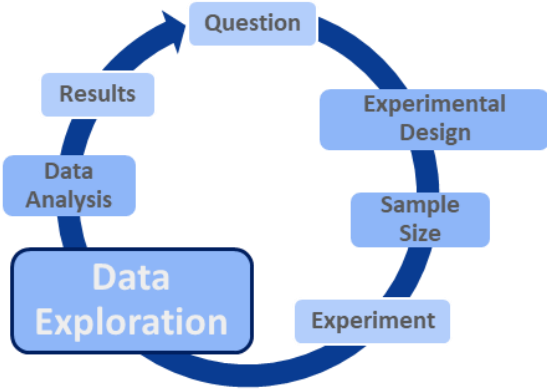
### Reliability

Reproducibility  
Respect  
Responsibility

### Rigor

Transparency





## Example 3

### Accountability

Accurate  
Care  
Collegiality  
Cooperation

### Ethics

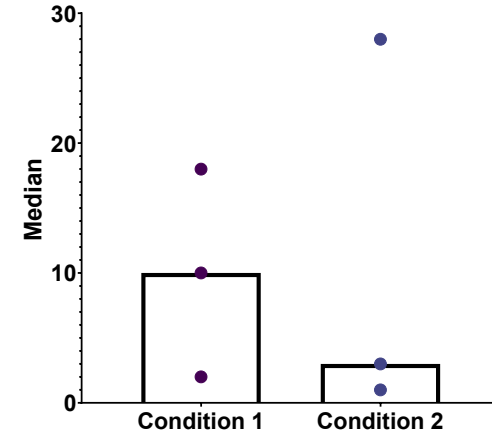
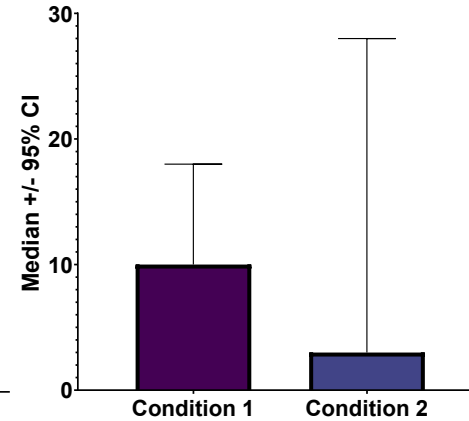
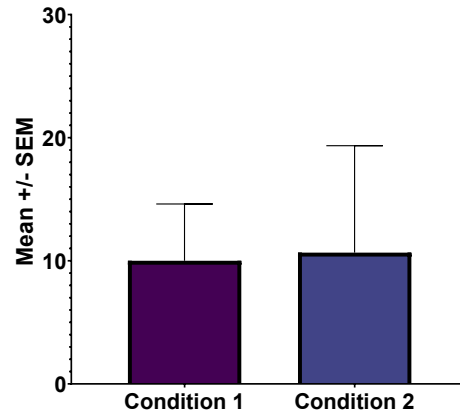
Fair  
Honesty  
Objectivity  
Openness  
Quality

### Reliability

Reproducibility  
Respect  
Responsibility

### Rigor

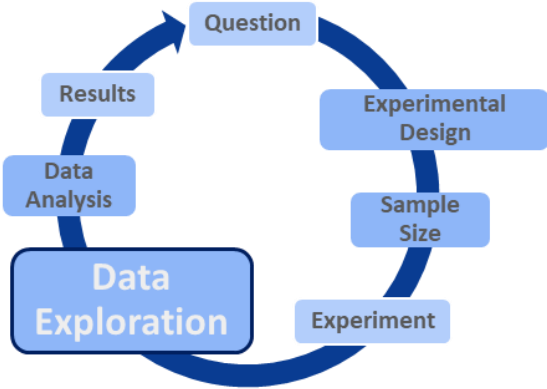
### Transparency



# Data Exploration...



**...Dubious Datasets**



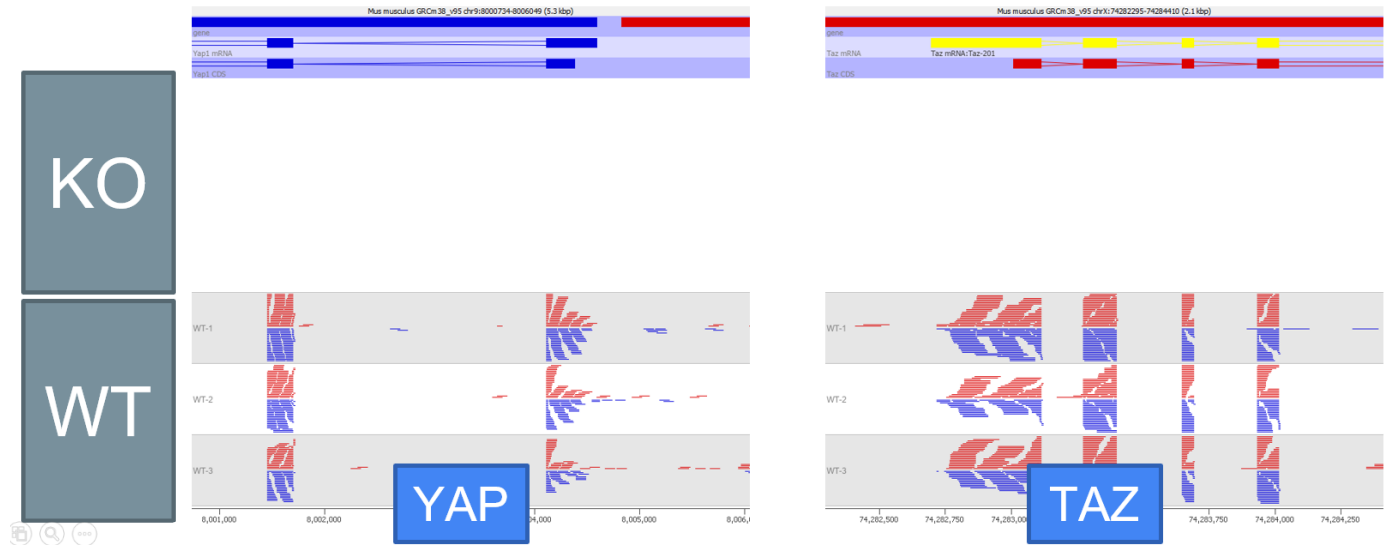
### Accountability

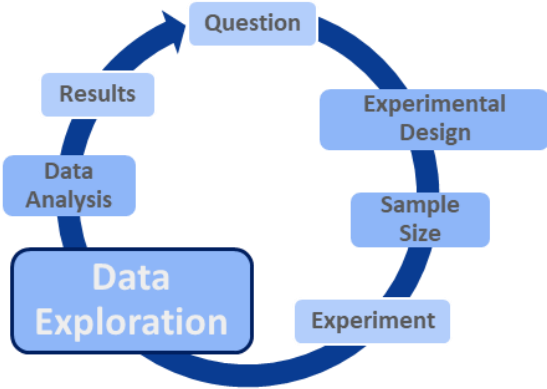
- Accurate
- Care
- Collegiality
- Cooperation
- Ethics**
- Fair
- Honesty
- Objectivity
- Openness
- Quality
- Reliability**
- Reproducibility
- Respect
- Responsibility
- Rigor**
- Transparency**

# Example 1: A Knockout?

**nature  
neuroscience**

**YAP and TAZ control peripheral myelination and the expression of laminin receptors in Schwann cells**



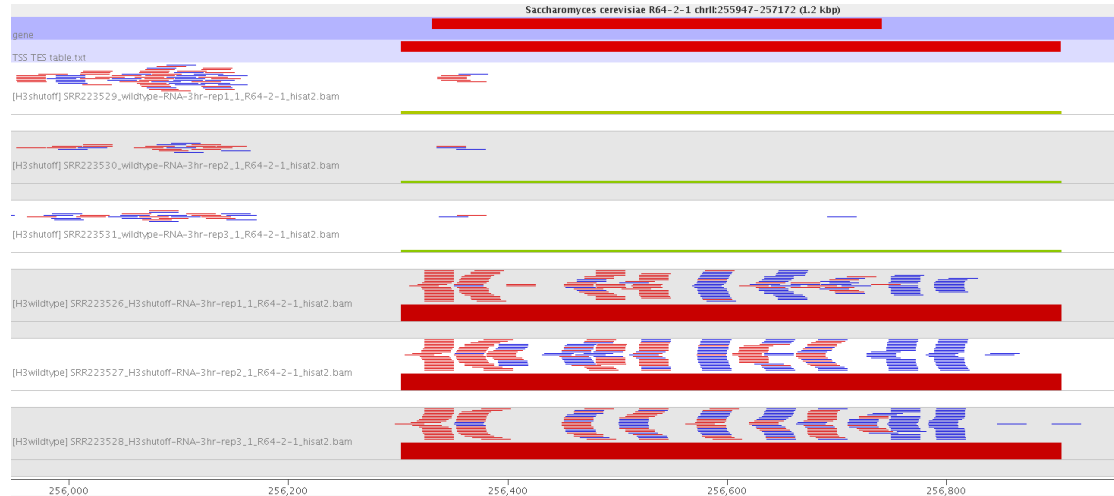


## Example 2: A Case of Mistaken Identity

OPEN ACCESS Freely available online

PLOS GENETICS

### *In Vivo* Effects of Histone H3 Depletion on Nucleosome Occupancy and Position in *Saccharomyces cerevisiae*



WT

KO

#### Accountability

Accurate  
Care  
Collegiality  
Cooperation

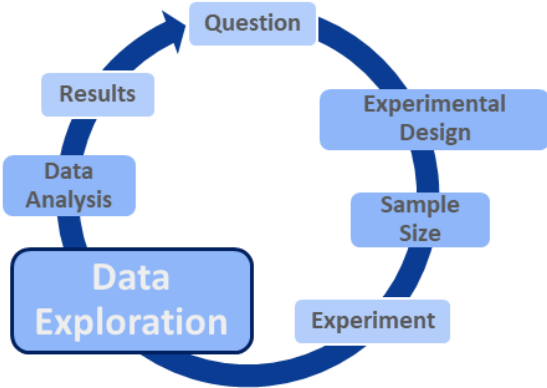
#### Ethics

Fair  
Honesty  
Objectivity  
Openness  
Quality

#### Reliability

Reproducibility  
Respect  
Responsibility

#### Rigor Transparency



## Example 3: Of Mice and Mycoplasma!



DOT1L-mediated murine neuronal differentiation associates with H3K79me2 accumulation and preserves SOX2-enhancer accessibility

### Accountability

Accurate  
Care  
Collegiality  
Cooperation

### Ethics

Fair

Honesty

Objectivity

Openness

Quality

Reliability

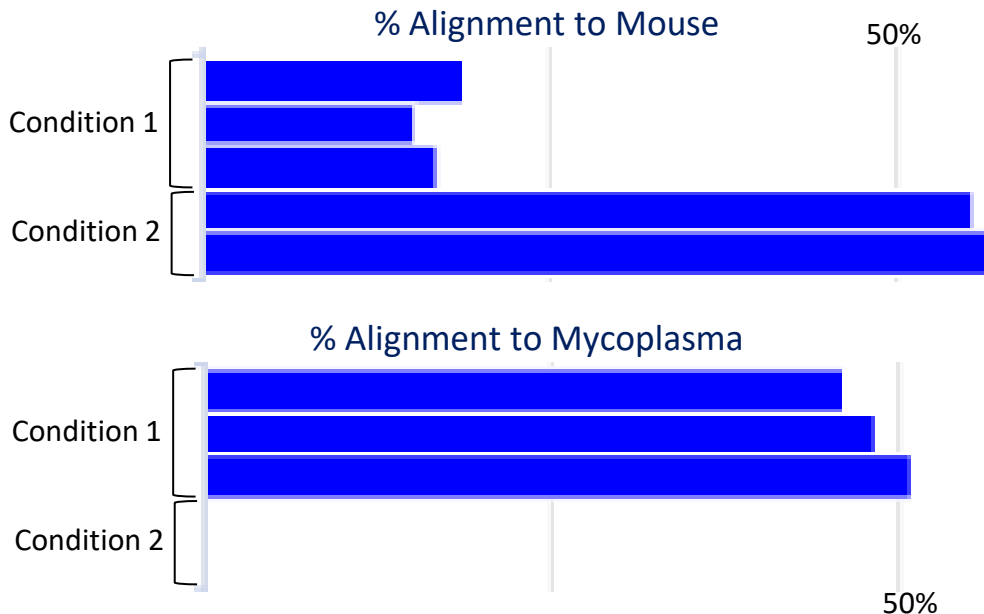
Reproducibility

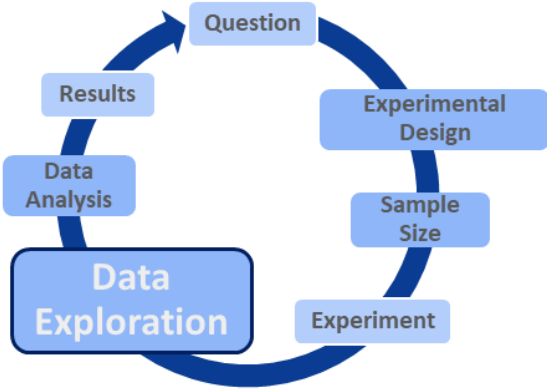
Respect

Responsibility

Rigor

Transparency





#### Accountability

Accurate  
Care  
Collegiality  
Cooperation

#### Ethics

Fair  
Honesty  
Objectivity  
Openness

#### Quality

Reliability  
Reproducibility  
Respect  
Responsibility

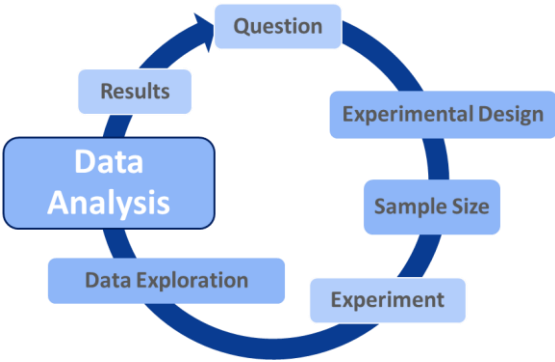
#### Rigor

Transparency

# All highlight the perils of assuming and not exploring....



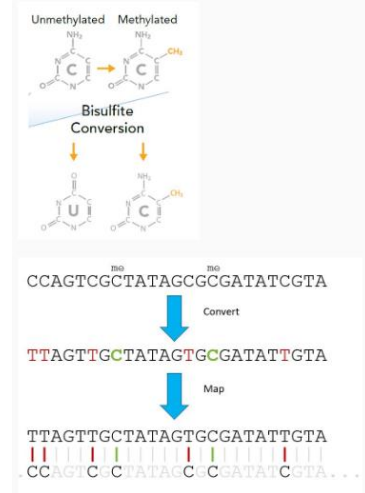
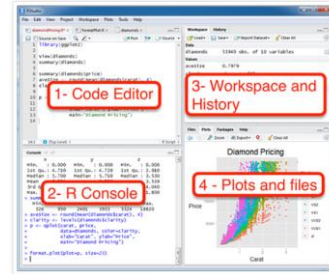
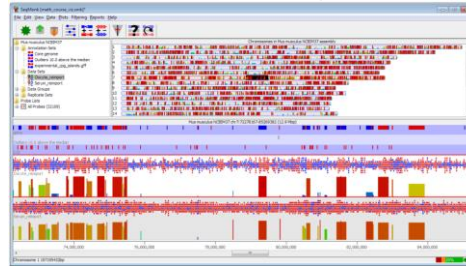
## ...Is it okay if you don't know there's a problem?



Accountability  
 Accurate  
 Care  
 Collegiality  
 Cooperation  
 Ethics  
 Fair  
 Honesty  
 Objectivity  
 Openness  
 Quality  
 Reliability  
 Reproducibility  
 Respect  
 Responsibility  
 Rigor  
 Transparency

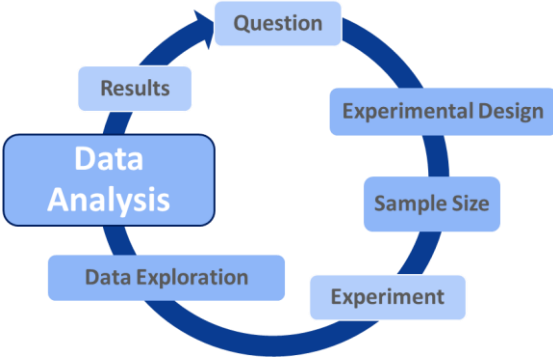
# Analytical Tools

The right tools for the job!



- Know what's out there
- Learn how to use it





Accountability  
Accurate  
Care  
**Collegiality**  
**Cooperation**  
Ethics  
Fair  
Honesty  
Objectivity  
Openness  
**Quality**  
**Reliability**  
**Reproducibility**  
Respect  
**Responsibility**  
**Rigor**  
Transparency

## Help With Analytical Tools

Babraham  
Bioinformatics

### Core Skills Courses

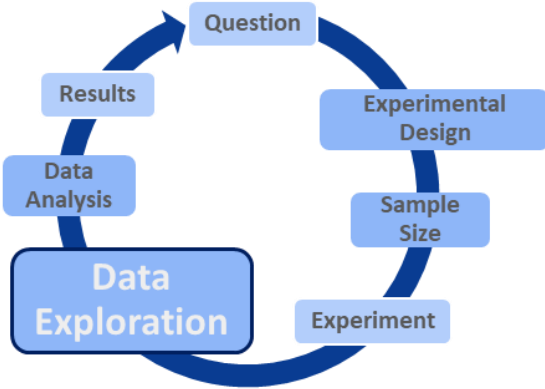
e.g. R Programming

Babraham  
Bioinformatics

### Application Specific Courses

e.g. NGS Analysis

<https://www.bioinformatics.babraham.ac.uk/training.html>



## In a nutshell

**Data exploration** is CRITICAL to:

- Understand our data
- Be confident in our findings

Ensure our results are quality and reliable

... Good for everyone!

### Accountability

Accurate  
Care  
Collegiality  
Cooperation

### Ethics

Fair  
Honesty  
Objectivity  
Openness

### Quality

Reliability  
Reproducibility  
Respect  
Responsibility

Rigor

Transparency

## Accountability

## Accurate

Care  
Collegiality  
Cooperation  
Ethics  
Fair

## Honesty

Objectivity  
Openness  
Quality

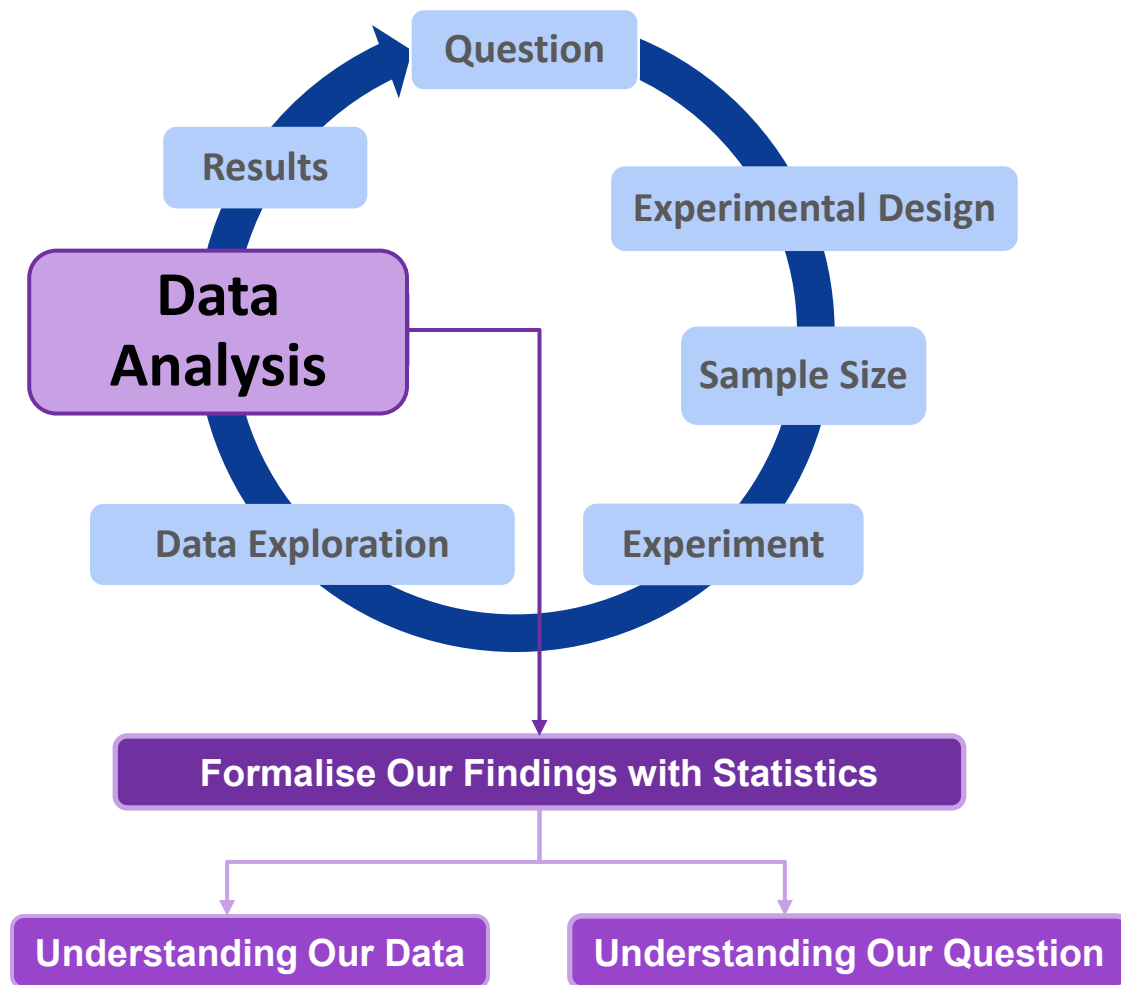
## Reliability

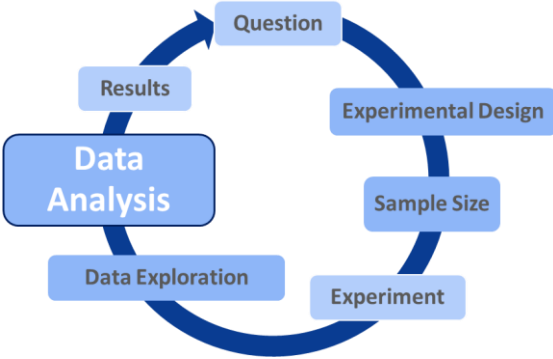
## Reproducibility

Respect  
Responsibility

## Rigor

Transparency





# Statistical Analysis

Translate the hypothesis/question into statistical questions

**By choosing the right test!**

Accountability  
Accurate  
Care  
Collegiality  
Cooperation  
Ethics  
Fair  
Honesty  
Objectivity  
Openness  
Quality  
Reliability  
Reproducibility  
Respect  
Responsibility  
Rigor  
Transparency

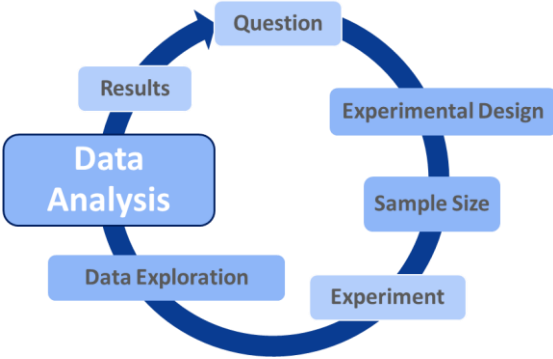
## Healthy approach:

**“It’s not about knowing the name of the test...”**

**...It’s knowing what the test should do”**



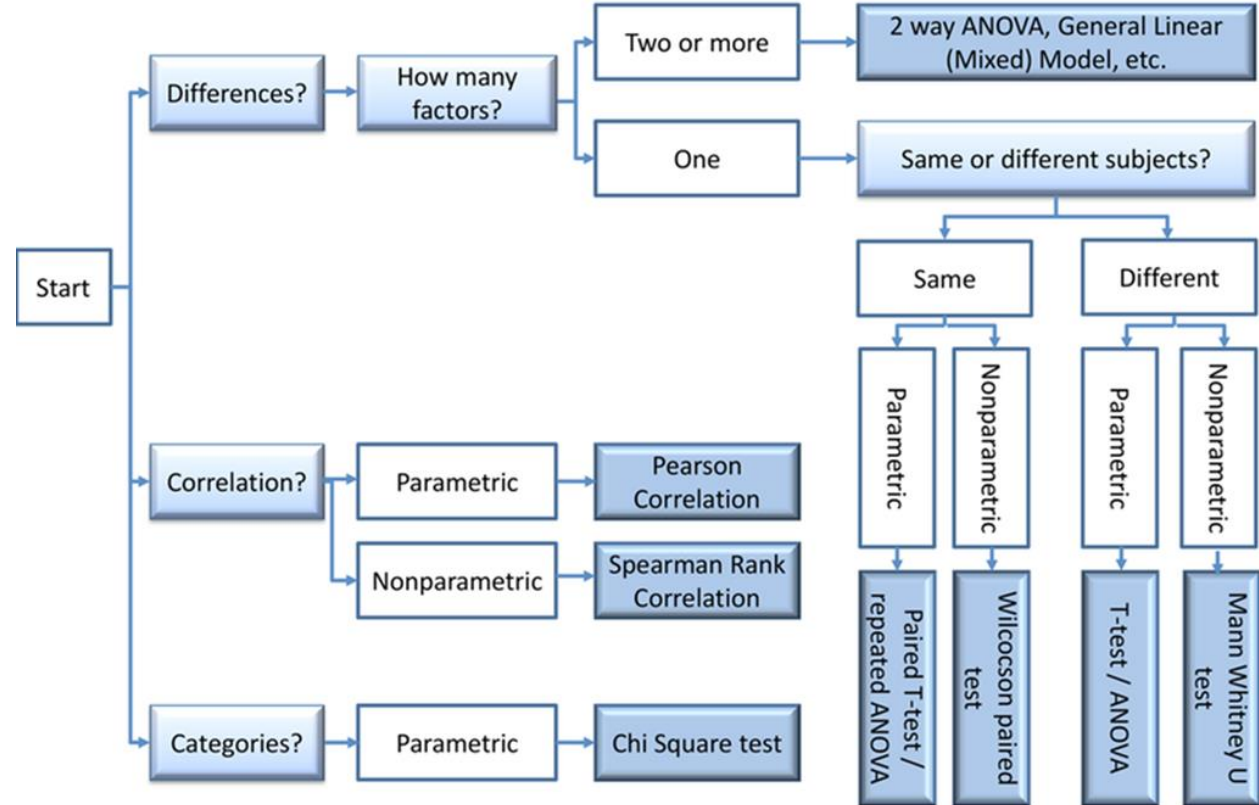
**Anne Segonds-Pichon**

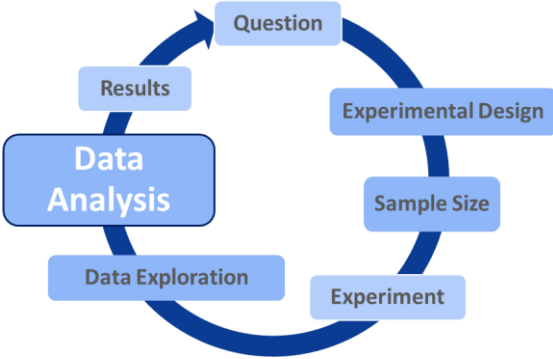


Accountability  
 Accurate  
 Care  
 Collegiality  
 Cooperation  
 Ethics  
 Fair  
 Honesty  
 Objectivity  
 Openness  
 Quality  
 Reliability  
 Reproducibility  
 Respect  
 Responsibility  
 Rigor  
 Transparency

# Knowing what the Test should do...

## Statistics Decision tree





Accountability  
Accurate  
Care  
**Collegiality**  
**Cooperation**  
Ethics  
Fair  
Honesty  
Objectivity  
Openness  
**Quality**  
**Reliability**  
**Reproducibility**  
Respect  
**Responsibility**  
**Rigor**  
Transparency

## Knowing what the Test should do...

Babraham  
Bioinformatics

### Statistics Using R

Babraham  
Bioinformatics

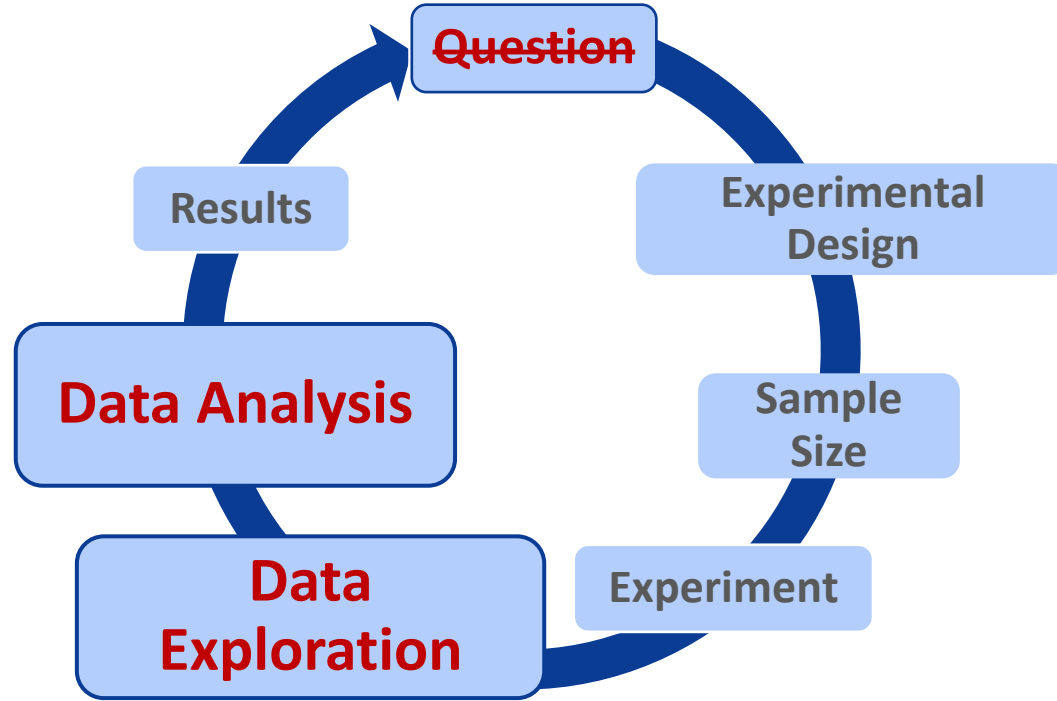
### Statistics Using GraphPad Prism

<https://www.bioinformatics.babraham.ac.uk/training.html>

# Research Integrity

## More than 1 way to Investigate!

Accountability  
Accurate  
Care  
Collegiality  
Cooperation  
Ethics  
Fair  
Honesty  
Objectivity  
Openness  
Quality  
Reliability  
Reproducibility  
Respect  
Responsibility  
Rigor  
Transparency



Exploratory data analysis

Accountability  
Accurate  
Care  
Collegiality  
Cooperation  
**Ethics**  
**Fair**  
**Honesty**  
**Objectivity**  
**Openness**  
**Quality**  
**Reliability**  
**Reproducibility**  
Respect  
**Responsibility**  
**Rigor**  
**Transparency**

## In a nutshell



**Data exploration** should always be a **pivotal step of analysis**

**Stats** helps us **formalise our findings**







# Research Integrity

## What does Ethics mean?

# Research Integrity Ethics and Animal Welfare

## Accountability

Accurate

## Care

Collegiality

Cooperation

## Ethics

Fair

Honesty

Objectivity

Openness

Quality

Reliability

Reproducibility

## Respect

## Responsibility

Rigor

Transparency

- If we need **biological replicates** to be confident in our results, why not have **as many as we can**?
- And what does it have to do with **integrity**?
- Time to talk about the **Home Office**, **ASPA** and the **3 Rs**

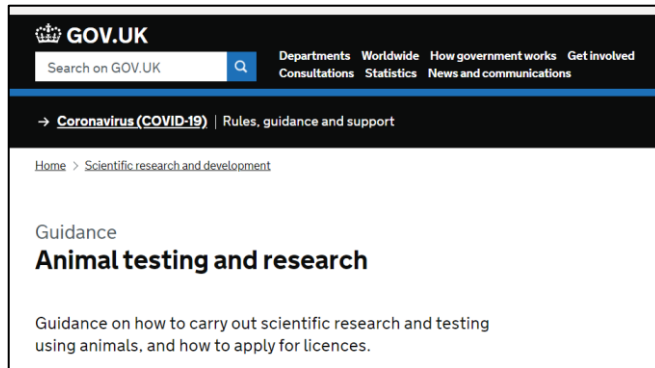


# Research Integrity

## Home Office

**The Home Office** (HO) is a ministerial department of the Government of the UK, responsible for immigration, security and law and order.

- But more importantly: **animal welfare**



By Steph Gray, CC BY-SA 2.0, <https://commons.wikimedia.org/w/index.php?curid=31387711>

# Research Integrity

## The 3 Rs

### Accountability

Accurate

### Care

Collegiality

Cooperation

### Ethics

Fair

Honesty

Objectivity

Openness

Quality

Reliability

Reproducibility

### Respect

### Responsibility

Rigor

Transparency



The diagram consists of three large blue circles arranged horizontally. Each circle contains one of the '3 Rs' in white text. The circles are of equal size and are separated by small gaps.

**Replace**

**Reduce**

**Refine**

# Research Integrity

## The 3Rs: about animal welfare

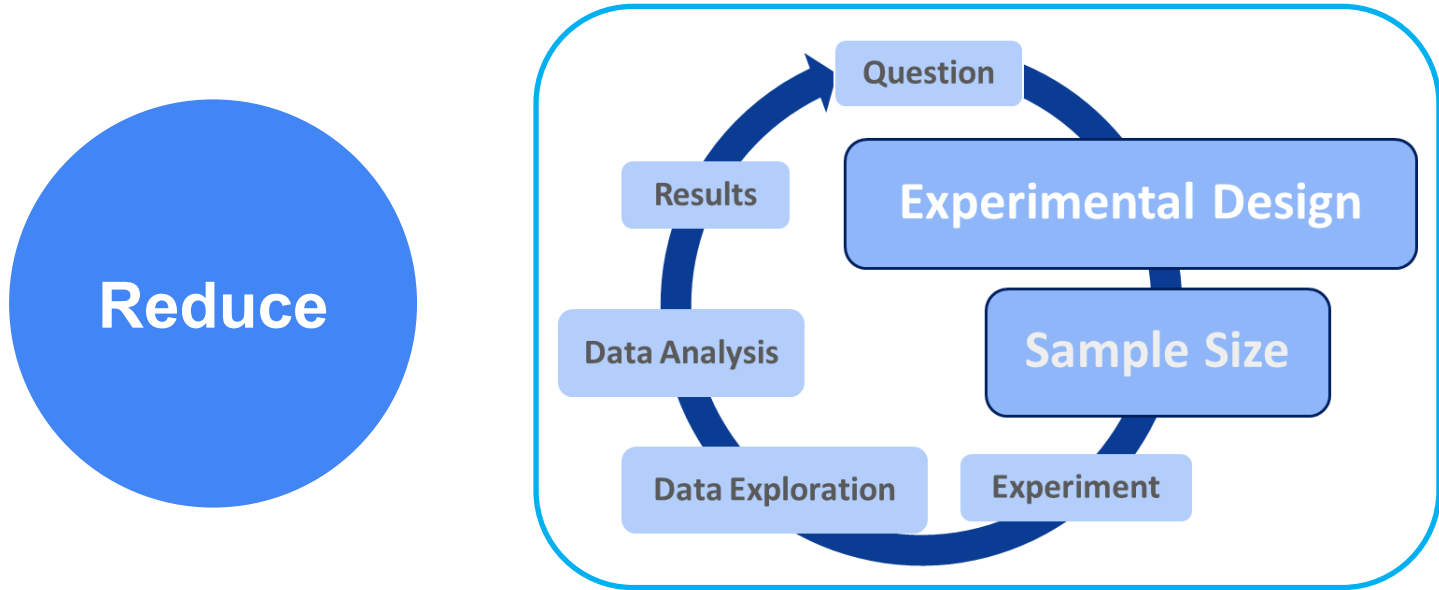
**Replace**



Avoiding or replacing the use of animals in areas where they otherwise would have been used.

# Research Integrity

## The 3Rs: about animal welfare



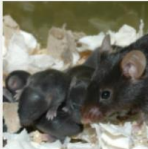
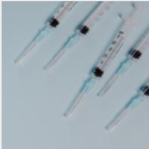
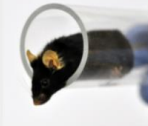



Minimising the number of animals used consistent with scientific aims.

# Research Integrity

## The 3Rs: about animal welfare



Refine

	<b>Breeding and colony management</b> Guidance on re-establishing colonies after a pause (e.g. COVID-19 lockdown).		<b>Blood sampling</b> Techniques for blood sampling in laboratory animal species to ensure the most appropriate technique is chosen.
	<b>Mouse handling: How to pick up a mouse</b> Guidance on non-aversive methods for handling mice, including tutorials, FAQs and practical tips.		<b>Evaluating environmental enrichment</b> Supporting technicians in assessing the welfare impact of new enrichment.
	<b>Grimace scales</b> Posters and other resources on the use of grimace scales to assess pain in laboratory animals.		<b>Rat tickling</b> Resources on rat tickling and how it can be used to promote positive human-animal interactions.

Minimising the pain, suffering, distress or lasting harm that research animals might experience.



# Research Integrity

## The 3 Rs at Babraham: AWERB

[The Hub](#) [People](#) [Committees](#) ▾ [Institute](#) ▾ [Science Facilities](#) ▾



### Animal Welfare & Ethical Review Body

#### Remit

To provide the campus with independent ethical advice on the balance of harms to benefits within scientific projects using animals. To monitor standards of animal care and welfare, to support and advise named persons and licensees working under the Animals (Scientific Procedures) Act and to advise the Establishment Licence Holder on the suitability of Project Licence applications. To develop initiatives and guidelines leading to the widest possible application of the 3Rs (refinement, reduction and replacement) both on the campus and amongst the wider scientific community. In accordance with our commitments to the Concordat on Openness in animal research, any staff member with a concern that falls within the AWERB remit is encouraged to speak to the AWERB chair or any other committee member.

# Good Research in Practice

# Good Research in Practice

## Accountability

Accurate

Care

Collegiality

Cooperation

Ethics

Fair

Honesty

Objectivity

Openness

Quality

Reliability

Reproducibility

Respect

Responsibility

Rigor

Transparency

## The Research Process

- Responsibilities
- Competence
- Project planning
- Quality Control

## Laboratory Practice

- Health and safety
- Handling of samples and materials
- Facilities and equipment
- Documentation of procedures and methods
- Research/work records



Get a GRiP!

'Good Research in Practice' (GRiP) supports and measures the more practical aspects of Research Integrity standards compliance on the ground in the lab.



# Keeping track of the research

## How?



# Research Integrity

## Laboratory Notebooks

### **Accountability**

Accuracy

Care

### **Collegiality**

### **Cooperation**

Ethics

Fair

### **Honesty**

Objectivity

### **Openness**

### **Quality**

### **Reliability**

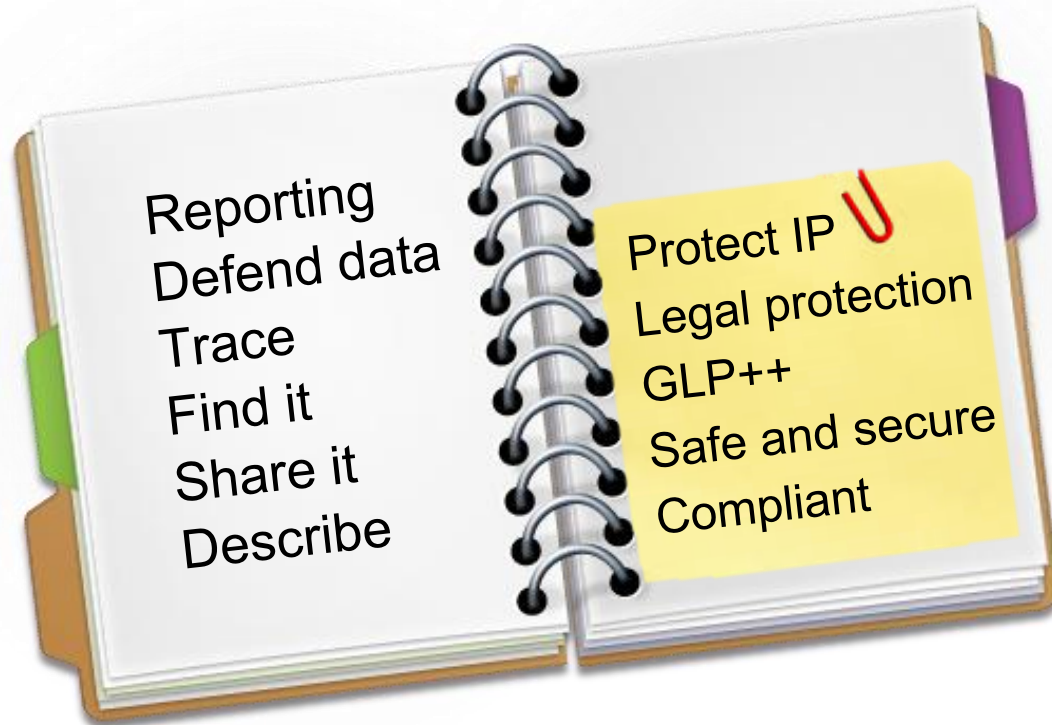
### **Reproducibility**

Respect

### **Responsibility**

### **Rigor**

### **Transparency**



# Research Integrity

## Laboratory Notebooks

**Accountability**

Accuracy

Care

**Collegiality**

**Cooperation**

Ethics

Fair

**Honesty**

Objectivity

**Openness**

**Quality**

**Reliability**

**Reproducibility**

Respect

**Responsibility**

**Rigor**

**Transparency**



[This Photo](#) by Unknown Author is licensed under [CC BY](#)

# Research Integrity

## Laboratory Notebooks

**Accountability**

Accuracy

Care

**Collegiality**

**Cooperation**

Ethics

Fair

**Honesty**

Objectivity

**Openness**

**Quality**

**Reliability**

**Reproducibility**

Respect

**Responsibility**

**Rigor**

**Transparency**





# Research Data Data Storage



**Accountability**

Accuracy

Care

**Collegiality**

**Cooperation**

Ethics

Fair

**Honesty**

Objectivity

**Openness**

**Quality**

**Reliability**

**Reproducibility**

Respect

**Responsibility**

**Rigor**

**Transparency**

Legal  
Requirements

Practicalities



**Accountability**

Accuracy

Care

**Collegiality**

**Cooperation**

Ethics

Fair

**Honesty**

Objectivity

**Openness**

**Quality**

**Reliability**

**Reproducibility**

Respect

**Responsibility**

**Rigor**

**Transparency**

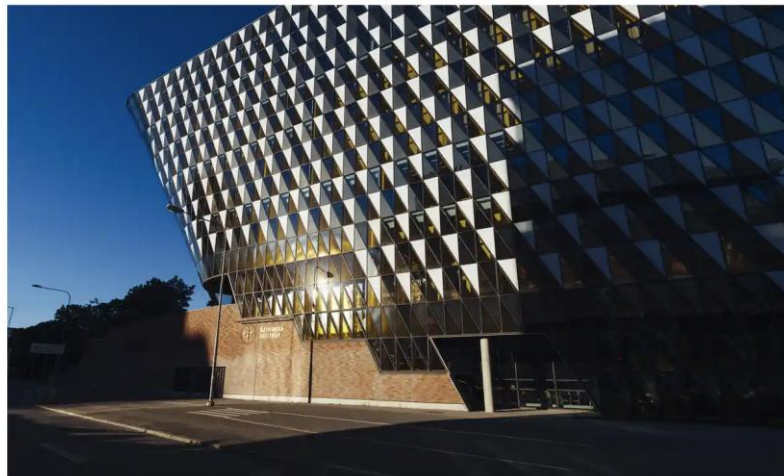
# Research Data

## Data Storage

**Sweden**

Decades of research destroyed after freezer fails at Swedish university

Estimated value of the samples thought to be in the millions as incident reported to police



📷 An internal investigation has been launched at the Karolinska Institutet despite no indication of sabotage. Photograph: Bloomberg/Getty Images

**The Guardian**  
Newspaper of the year

- Interruption in the supply of liquid nitrogen leading to the destruction of samples from multiple institutions.
- Valued in the millions.
- “Those worst affected are those researching leukaemia, they have gathered samples from patients over as much as 30 years,”

# What should we be doing?

## Expectations and Responsibilities



Created by Fajar Studio  
from Noun Project



Created by Ribbla Team  
from Noun Project



Created by Anggara Putra  
from Noun Project



Created by shashank singh  
from Noun Project



Created by Good Wife  
from Noun Project

# Research Data OneNote

**Accountability**

Accuracy

Care

**Collegiality**

**Cooperation**

Ethics

Fair

**Honesty**

Objectivity

**Openness**

**Quality**

**Reliability**

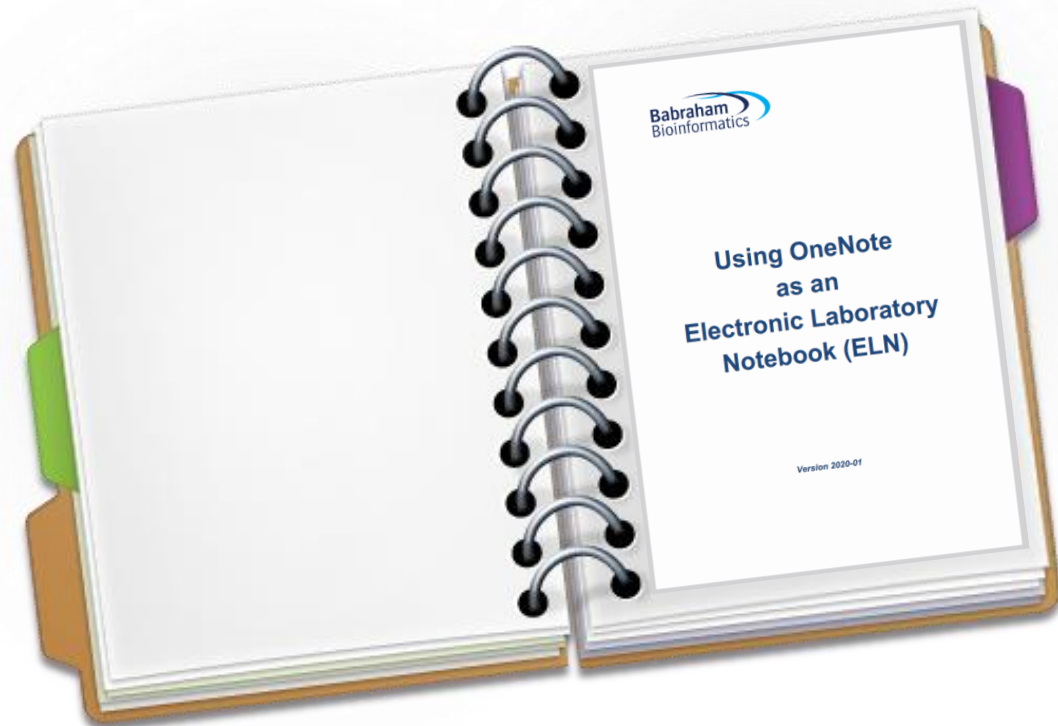
**Reproducibility**

Respect

**Responsibility**

**Rigor**

**Transparency**

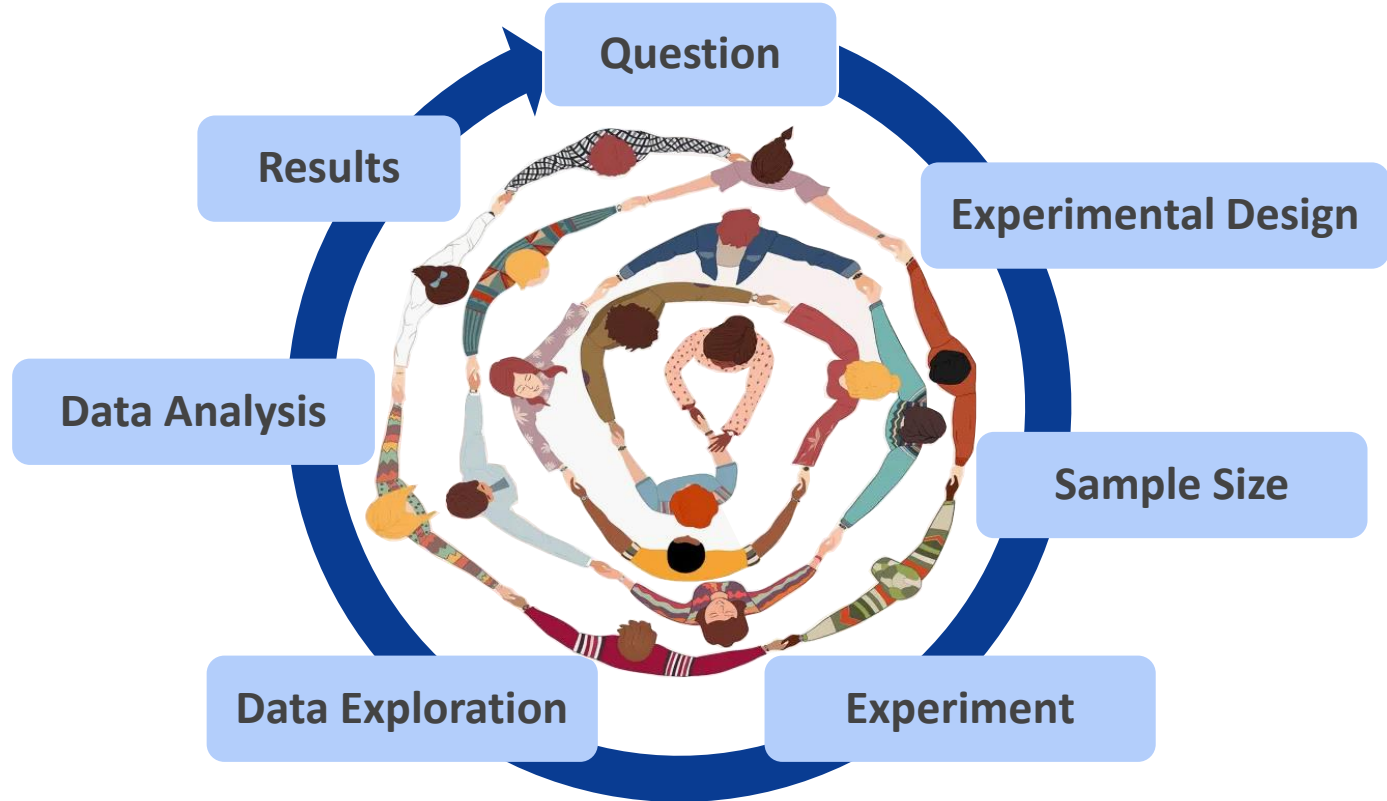


# Responsibility

# Research Integrity

## Wider Responsibility

Accountability  
Accurate  
Care  
Collegiality  
Cooperation  
Ethics  
Fair  
Honesty  
Objectivity  
Openness  
Quality  
Reliability  
Reproducibility  
Respect  
Responsibility  
Rigor  
Transparency



# Research Integrity

## **Accountability**

Accuracy

Care

## **Collegiality**

## **Cooperation**

Ethics

Fair

## **Honesty**

Objectivity

## **Openness**

## **Quality**

## **Reliability**

## **Reproducibility**

Respect

## **Responsibility**

## **Rigor**

## **Transparency**



Photo by Miguel Henriques Unsplash



# Research Integrity

## Questioning

**Accountability**

Accuracy

Care

**Collegiality**

**Cooperation**

Ethics

Fair

**Honesty**

Objectivity

**Openness**

**Quality**

**Reliability**

**Reproducibility**

Respect

**Responsibility**

**Rigor**

**Transparency**



Photo by Ana Municio on Unsplash

# Research Integrity

## Collaboration and Competition

### **Accountability**

Accuracy

Care

### **Collegiality**

### **Cooperation**

Ethics

Fair

### **Honesty**

Objectivity

### **Openness**

### **Quality**

### **Reliability**

### **Reproducibility**

Respect

### **Responsibility**

### **Rigor**

### **Transparency**



Photo by Natalie Pedigo on Unsplash



# Research Integrity

## The Game

**Accountability**

Accuracy

Care

**Collegiality**

**Cooperation**

Ethics

Fair

**Honesty**

Objectivity

**Openness**

**Quality**

**Reliability**

**Reproducibility**

Respect

**Responsibility**

**Rigor**

**Transparency**



Photo by Karthik Balakrishnan on Unsplash

# Research Integrity

## Wider Responsibility and Scientific Community

Accountability  
Accuracy  
Care  
Collegiality  
Cooperation  
Ethics  
Fair  
Honesty  
Objectivity  
Openness  
Quality  
Reliability  
Reproducibility  
Respect  
Responsibility  
Rigor  
Transparency

### Review

This means both:

- Having your work reviewed by others - peer review (journals, papers, conference etc)
- Reviewing other people's work

Reviewing also means more locally: With in your group, department, or culture.

Criticism should be positive, not confrontational.

Your group have a vested interest in you and your science being successful and getting the very right, robust, data.

Actively seek input and questioning

Don't take it personally

Be positive to your criticism of others' science

It makes you and your science better. Of the people know of things, it needs both ways. You do it for them, they do it for you.

As your science progresses, you pass through it more times, pick up issues, find alternative comparisons.

You don't want to be taken down in public for a lot of reasons (for example when a journal scientist took those words to an abstract and conference without sharing with others first and their feelings went against the consensus understanding of the paper discussed - For some professional and dignified equity and understanding, the fact that it was his to keep to you.)

Like meetings etc, you're talking around the subject but you're coming to a decision, how the thing, get the things right before you go to conference/bioscience etc.

I will save you from future embarrassment, time and effort to get through this now.

**What does reviewing the science mean to you?**

At least something on the open science process is available.

### FAIR

Findable  
Accessible  
Interoperable  
Reusable

Ensuring your science is FAIR can maximize the value of your science and your data.

Your data has value above and beyond what you created it for, especially with big datasets.

Make your data as useful as possible (make it work for you and others)

Others can put things out of your data.

This is good for citizens, and ultimately adds value to your research.

Also good for the planet - you're saving what has already been done.

This is about data that are - so we tell you to do it, to make data accessible, BUT, there's a cost and effort and it's not as simple as it sounds - it's a lot of work.

At it we're generating (generating) data that we generate new data. And we generate a lot.

It's important that this is made use of public data as much as possible.


Open data, and more.

Also, make sure - if your work is based on an existing paper (data, that you do a paper) that you're using the data.

This is also a big challenge (generating standards and analyzing the type of data you will likely be collecting)

And also good (check - it will save you a lot of time and money)

**How does this link to research integrity?**

### Pre-prints

The positives and negatives to pre-prints are a bit more nuanced.

The problem is that published data is often out of date.

By the time a paper comes out (6 months +, mostly more),

Take time to get published, findings may have moved on.

Now have bioRxiv.

A way to circumvent the delay in getting research out BUT it's not peer reviewed.

It looks like a paper, but it's not:

- Not peer reviewed
- Not fact checked
- A high proportion on bioRxiv never makes it into a paper.

If we consider that about half of all that is published (and peer reviewed) is wrong, this pushes that likelihood up even more with bioRxiv.

However, it is worth thinking about if your data is time critical.


You may also need to check with your preferred journal - will they preprint you? Not accept if you've published on bioRxiv first?

What are the pros and cons?

pro

cons

**How does this link to research integrity?**

### Open Access

This is a bit of a game.

Open access can have benefits to you, and others.

There are open access journals, or some which you can pay a fee to make open access.

This means your paper is more accessible.

Also accessible to automatic text mining, so even more can be gleaned from your data (see FAIR).

Something like PubMed Central will also make papers open access - with about a 6 month delay - on loads of journals, not just open access ones.

Means you can put out information, including API etc (auto program).

**Remember:** Just because something is in a paper, doesn't mean it's actually true!

What are the pros and cons?

pro

cons

**How does this link to research integrity?**



# Research Integrity

## Review

This means both:

- Having your work reviewed by others – peer review (journals, papers, conferences etc)
- Reviewing other people's work

Reviewing also means more locally: Within your group, department, institute.

Criticism should be positive, not combative.

Your group have a vested interest in you and your science being successful and getting the story right. Listen to them.

Actively seek  
input and  
questioning

Don't take it  
personally

Be positive in  
your criticism  
of others' science

It makes you and your science better. Other people know lots of things. It works both ways. You do it for them, they do it for you.

As your science progresses, you pass through these barriers, pick up issues, find alternative viewpoints.

You don't want to be taken down in a public forum (see how of an example where a junior scientist took their work to an internal oral conference without sharing with others first and their findings were against the common understanding of everyone's research – they were questioned and disproved quickly and embarrassingly. We don't want this to happen to you!). Lab meetings etc can be a safe general discussion opportunity to consider, make mistakes, learn the story, get things right before you go to conference/review etc.

It will save you time, future embarrassment, time and effort to go through this work.

What does reviewing the science mean to you?

1/3 Further training on the peer review process is available.



Birmingham Bioinformatics  
Fostering collaborative relationships between computational biologists and clinicians



## FAIR

Findable  
Accessible  
Interoperable  
Reusable

(which could mean replicate the spirit of the analysis, rather than just jump through mostly the same hoops)

Findable

Accessible

Interoperable

Reusable

Ensuring your science (FAIR) maximises the value of your science and your data.

Your data has value above and beyond what you created it for, especially with big datasets.

Make your data as useful as possible (make it work for you and others).

Others can pull things out of your data.

This is good for citations, and it's really adding value to your research.

Also good for the planet – not repeating what has already been done.

This is about data disclosure – so we need to do this, to make data accessible, BUT, there's a real selling point here to make it as accessible as possible – it benefits everyone.

At it we expose ourselves (exposing) data if we generate new data. And we generate a LOT.

It's also important that if you make use of public data as much as possible.

Saves time, and money.

Also, make sure – if your work is based on an existing paper/data, that you do a PAPER check on the actual data.

This is also a learning opportunity by checking the quality and analysing the type of data you will likely be collecting.

And also a good check – it could save you a lot of time and money!

How does this link to research integrity?



Birmingham Bioinformatics  
Fostering collaborative relationships between computational biologists and clinicians



## Pre-prints

The positives and negatives to pre-prints are a bit more nuanced.

The problem is that published data is often out of date.

By the time a paper comes out (6 months +, mostly more), takes time to get published, findings may have moved on.

Now have bioRxiv.

A way to circumvent the delay in getting research out BUT it's not peer reviewed.

It looks like a paper, but it's not:

Not peer reviewed

Not fact checked

A High proportion on bioRxiv never makes it into a paper.

If we consider that about half of all that is published (and peer reviewed) is wrong, this pushes that likelihood up even more with bioRxiv.

However, it is worth thinking about if your data is time critical. You may also need to check with your preferred journal – will they penalise you/not accept if you've published on bioRxiv first?

What are the pros and cons?

pros

cons

How does this link to research integrity?



Birmingham Bioinformatics  
Fostering collaborative relationships between computational biologists and clinicians



## Open Access

This is a bit of a game.

Open access can have benefits to you, and others.

There are open access journals, or some which you can pay a fee to make open access.

This means your paper is more accessible.

Also accessible to automatic text mining, so even more can be gleaned from your data (see FAIR).

Something like PubMed Central will also make papers open access – with about a 6 month delay – on loads of journals, not just open access ones.

Means you can pull out information, including API etc (auto program).

Remember: Just because something is in a paper, doesn't mean it's actually true!

What are the pros and cons?

pros

cons

How does this link to research integrity?



Birmingham Bioinformatics  
Fostering collaborative relationships between computational biologists and clinicians



# Research Integrity

## Accountability

Accuracy

Care

## Collegiality

## Cooperation

Ethics

Fair

## Honesty

Objectivity

## Openness

## Quality

## Reliability

## Reproducibility

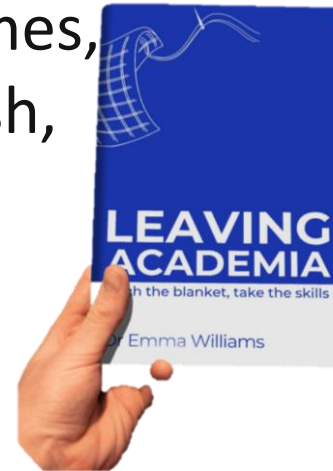
Respect

## Responsibility

Rigor

## Transparency

“Academia is at its best a meritocracy and at its worst a pit of vipers. A grown-up version of the children’s game snakes and ladders. The quest for knowledge and the pushing back of frontiers has created a hierarchal system with super demanding and, at times, conflicting metrics (postdocs must publish, publish, publish but to apply for a lectureship position you need to have a teaching track record).”



# Research Integrity

## In a nutshell

Applying **research integrity principles** is our **responsibility as scientists**

Accountability

Accurate

Care

Collegiality

Cooperation

Ethics

Fair

Honesty

Objectivity

Openness

**Quality**

**Reliability**

**Reproducibility**

Respect

**Responsibility**

Rigor

Transparency



# Research Integrity In the lab

Accountability

Accurate

Care

**Collegiality**

**Cooperation**

**Ethics**

Fair

Honesty

Objectivity

Openness

Quality

Reliability

Reproducibility

**Respect**

Responsibility

Rigor

Transparency



# Research Integrity

## Work culture

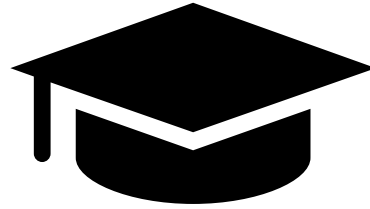
Accountability  
Accurate  
Care  
**Collegiality**  
**Cooperation**  
**Ethics**  
Fair  
Honesty  
Objectivity  
Openness  
Quality  
Reliability  
Reproducibility  
**Respect**  
Responsibility  
Rigor  
Transparency





# Research Integrity

## PhD Students



**It doesn't matter if it 'doesn't work'**

### **Accountability**

Accurate

### **Care**

Collegiality

Cooperation

### **Ethics**

Fair

Honesty

Objectivity

Openness

Quality

Reliability

Reproducibility

### **Respect**

### **Responsibility**

Rigor

Transparency



# Research Integrity

## Accountability

Accurate

## Care

Collegiality

Cooperation

## Ethics

Fair

Honesty

Objectivity

Openness

Quality

Reliability

Reproducibility

## Respect

## Responsibility

Rigor

Transparency

Work hard...



...but what if it is too hard?

# Research Integrity

## Accountability

Accurate

## Care

Collegiality

Cooperation

## Ethics

Fair

Honesty

Objectivity

Openness

Quality

Reliability

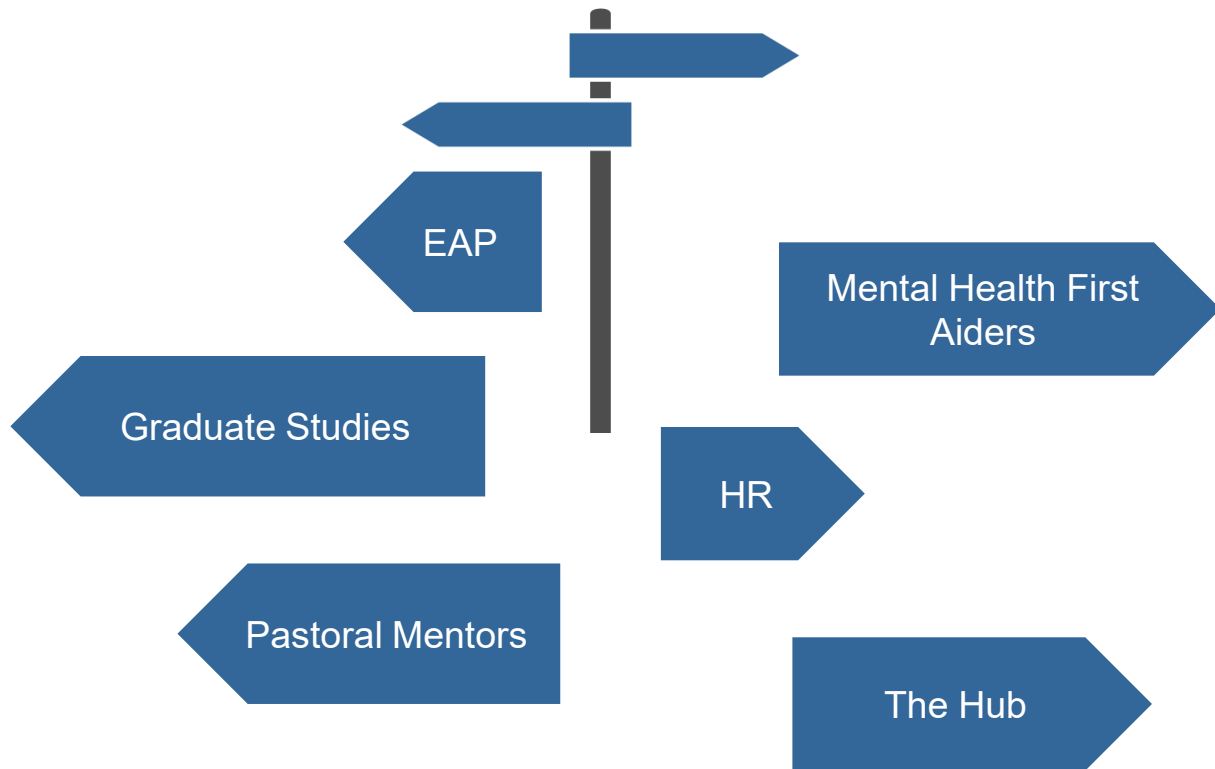
Reproducibility

## Respect

## Responsibility

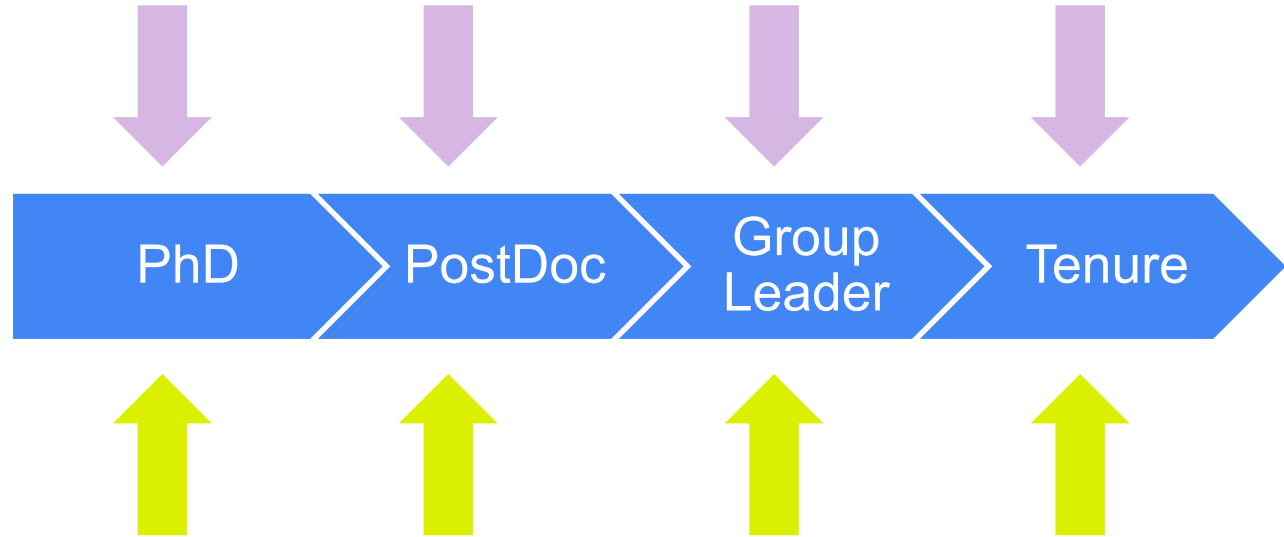
Rigor

Transparency



# Research Integrity

Accountability  
Accurate  
Care  
Collegiality  
Cooperation  
Ethics  
Fair  
Honesty  
Objectivity  
Openness  
Quality  
Reliability  
Reproducibility  
Respect  
Responsibility  
Rigor  
Transparency



# Research Integrity



Which roles are most important or responsible for Research Integrity?

Accountability  
Accurate  
Care  
Collegiality  
Cooperation  
Ethics  
Fair  
Honesty  
Objectivity  
Openness  
Quality  
Reliability  
Reproducibility  
Respect  
Responsibility  
Rigor  
Transparency

Director

PostDoc

PhD  
student

Head  
of ISP

Technician

Research  
Assistant

H&S  
officer

Lab  
Manager

Group  
Leader

Facilities

Animal  
Technician

Leadership  
Team

Visiting  
Scientist

Tenure

# Generative AI

# Have You Used Generative AI?

Accountability

Accurate

Care

Collegiality

Cooperation

Ethics

Fair

Honesty

Objectivity

Openness

Quality

Reliability

Reproducibility

Respect

Responsibility

Rigor

Transparency



Gemini



Copilot



ChatGPT



Apple Intelligence

A

Yes

B

No

# Have You Used Generative AI?

Q: How often do you use generative AI tools (such as ChatGPT) at work?

- I use them every day
- I use them more than once a week
- I use them occasionally
- I've used them only a few times
- Never

Respondents who study AI



Respondents who use AI in research



Respondents who don't use AI in research



AI and science: what 1,600 researchers think

<https://www.nature.com/articles/d41586-023-02980-0#correction-0>

Accountability

Accurate

Care

Collegiality

Cooperation

Ethics

Fair

Honesty

Objectivity

Openness

Quality

Reliability

Reproducibility

Respect

Responsibility

Rigor

Transparency

# Why Use Generative AI?

Accountability

Accurate

Care

Collegiality

Cooperation

Ethics

Fair

Honesty

Objectivity

Openness

Quality

Reliability

Reproducibility

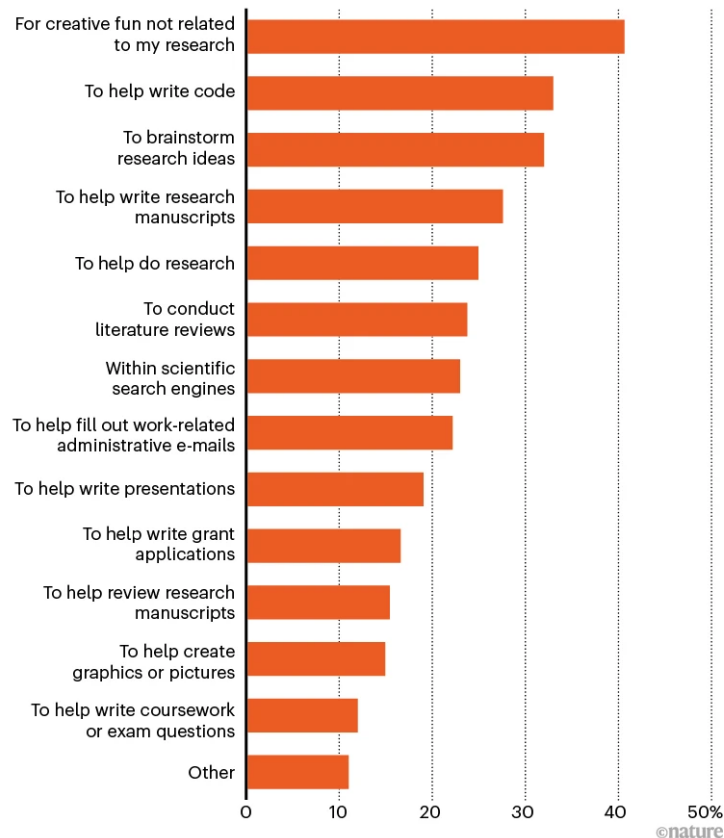
Respect

Responsibility

Rigor

Transparency

Q: What do you use generative AI tools (such as ChatGPT and other large language models) for? (Choose all that apply.)





# Considerations For Generative AI

## Will it Actually Help?

Accountability

Accurate

Care

Collegiality

Cooperation

Ethics

Fair

Honesty

Objectivity

Openness

Quality

Reliability

Reproducibility

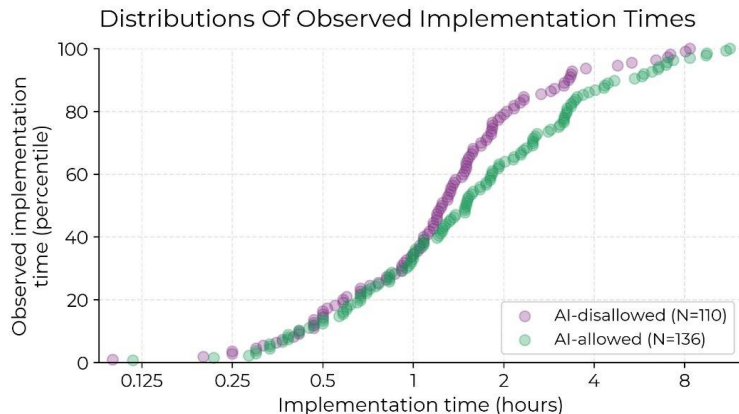
Respect

Responsibility

Rigor

Transparency

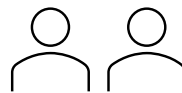
### Software Developers



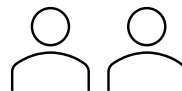
Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity, Becker *et al* 2025

### Scientists

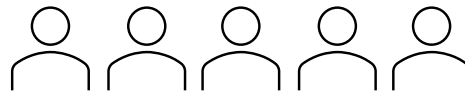
From 9 Self-Reported “Correct” Answers



Correct



Partially Correct



Incorrect

Attendees of the Joint International Bioinformatics Hackathon 2025.

# Considerations For Generative AI

## Accountability

Accurate

## Care

Collegiality

Cooperation

## Ethics

Fair

Honesty

Objectivity

Openness

Quality

Reliability

Reproducibility

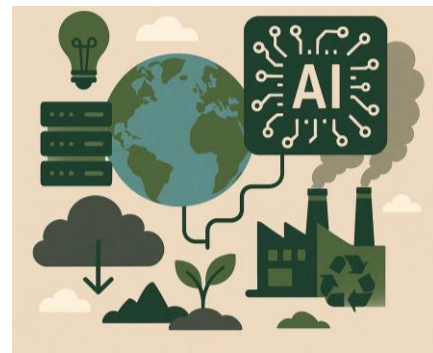
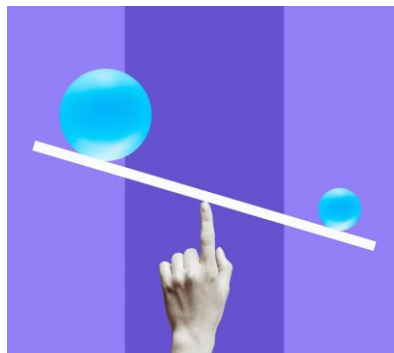
Respect

Responsibility

Rigor

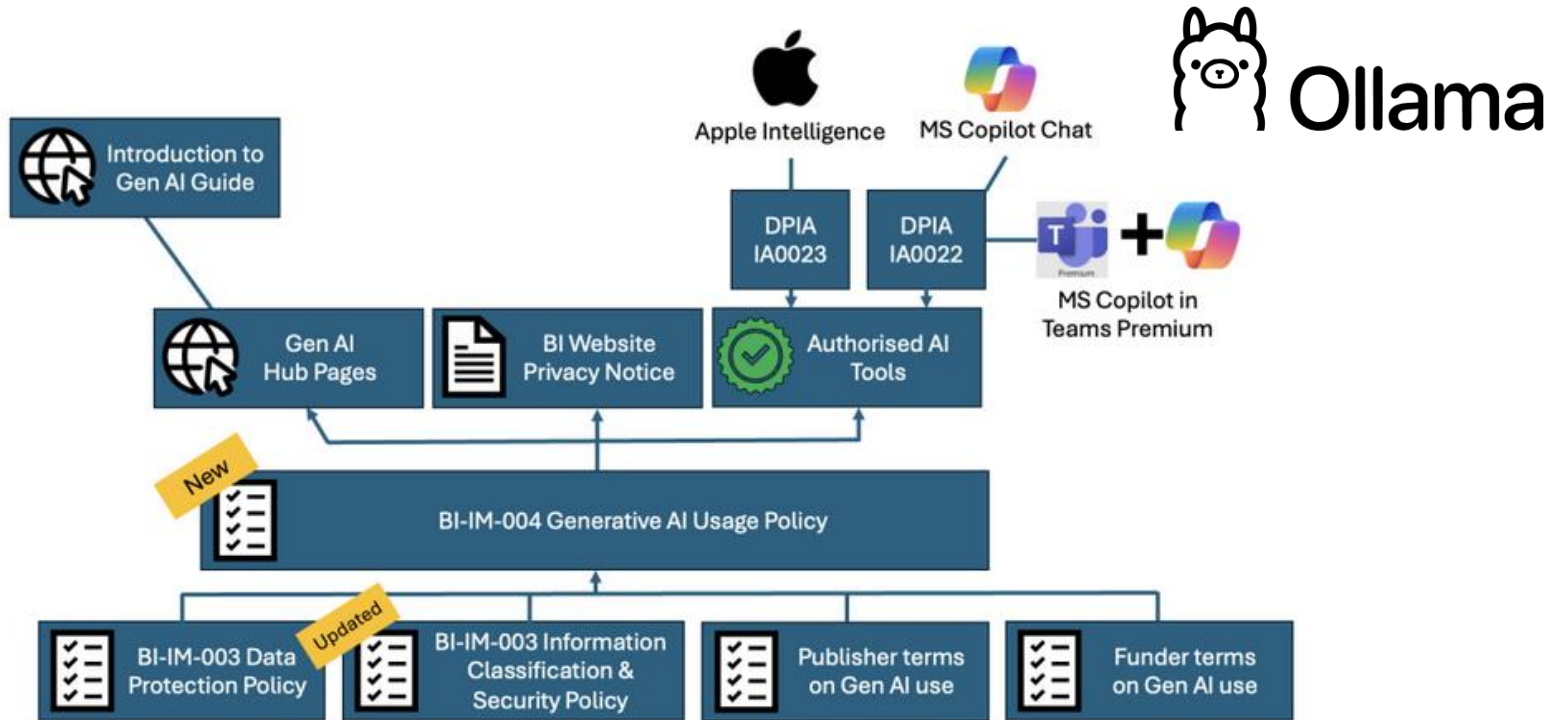
Transparency

## Will it Harm?



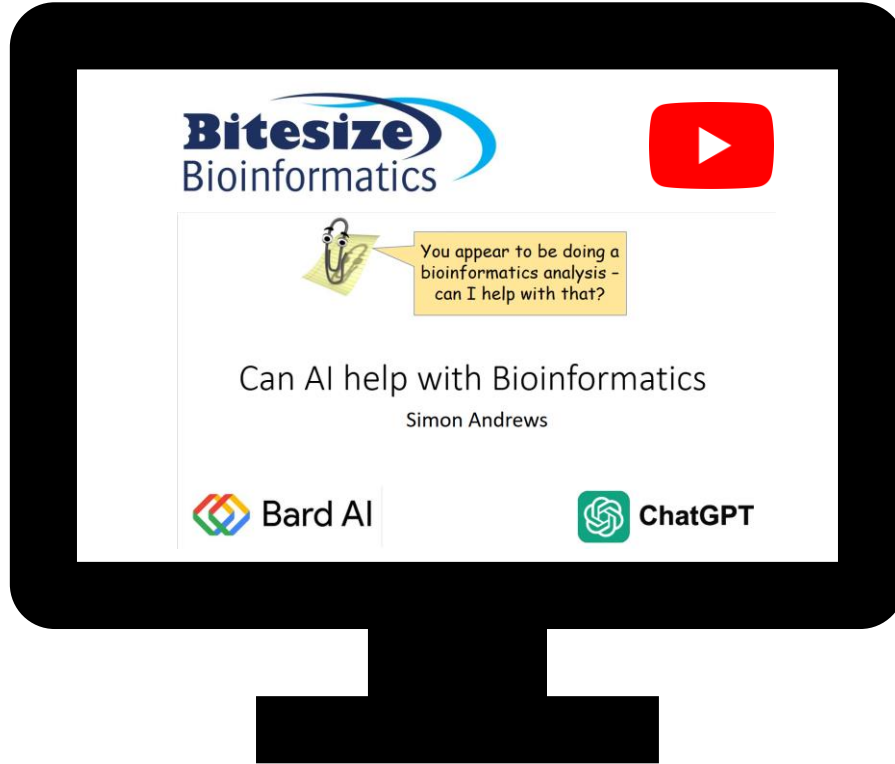
Ultimately it is our responsibility

# Generative AI at Babraham



See The Generative AI Pages on The Hub

# Generative AI and Bioinformatics



Babraham  
Bioinformatics

**New  
Course  
Coming  
Soon**

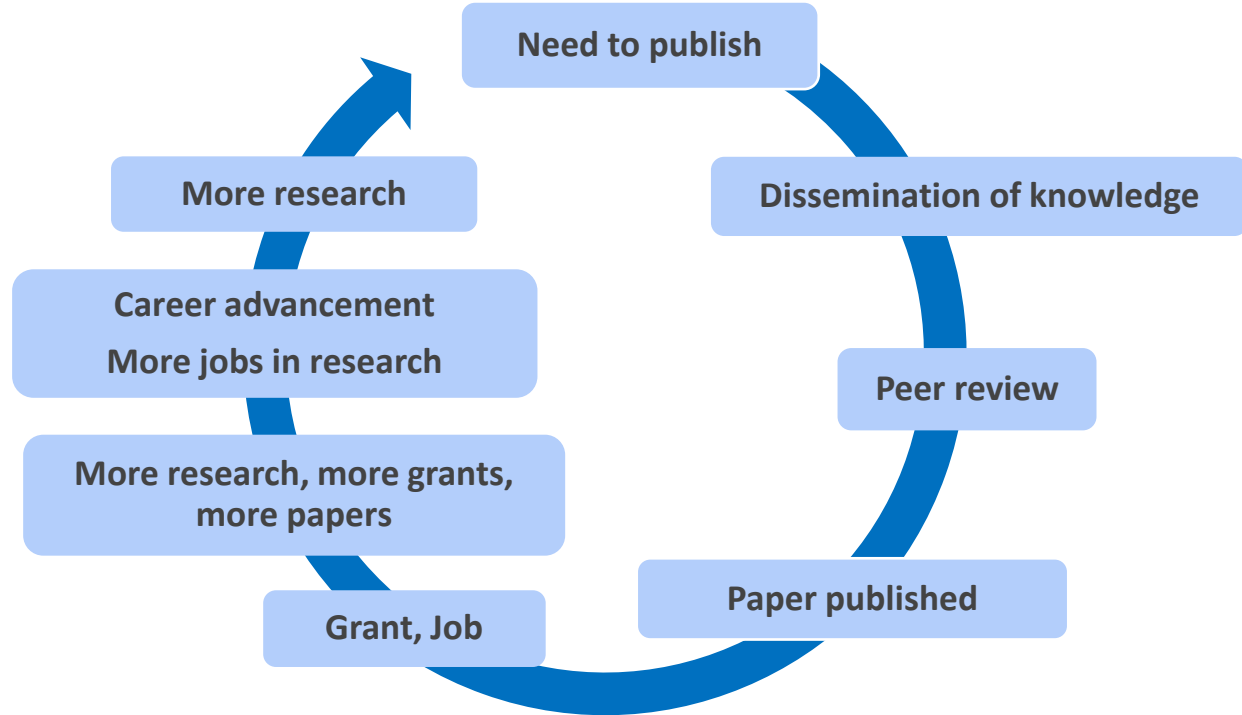
# How our integrity may be tested

# Publications

# Publish or Perish?

# Publications: the good

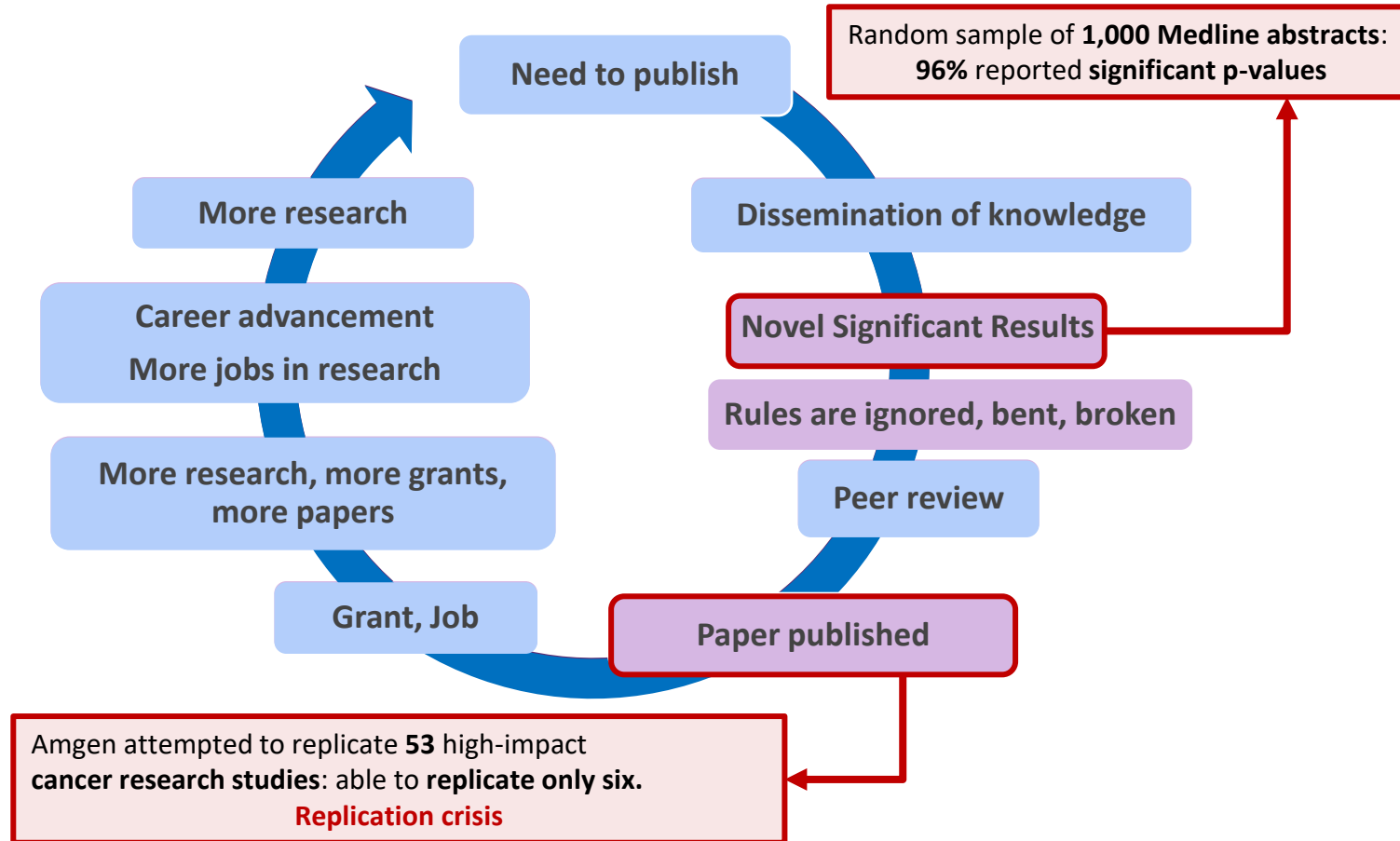
Accountability  
Accurate  
Care  
**Collegiality**  
**Cooperation**  
Ethics  
Fair  
Honesty  
Objectivity  
**Openness**  
Quality  
Reliability  
Reproducibility  
Respect  
Responsibility  
Rigor  
Transparency





# Publications: the bad

Accountability  
Accurate  
Care  
**Collegiality**  
**Cooperation**  
Ethics  
Fair  
Honesty  
Objectivity  
**Openness**  
Quality  
Reliability  
Reproducibility  
Respect  
Responsibility  
Rigor  
Transparency



# Publications: the bad

## The most cited paper to be retracted

**Pluripotency of mesenchymal stem cells derived from adult marrow**

> *Nature*. 2002 Jul 4;418(6893):41-9. doi: 10.1038/nature00870. Epub 2002 Jun 20.

### Accountability

Accurate  
Care  
Collegiality  
Cooperation

### Ethics

Fair  
Honesty  
Objectivity  
Openness  
Quality  
Reliability  
Reproducibility  
Respect  
Responsibility  
Rigor  
Transparency

June 2002

Paper published in *Nature*

June 2024

**Retraction** of the paper

December 2019

New concerns raised by  
Elisabeth Bik

<https://pubpeer.com/publications/DF95522E3585E37663CAD1972E70BD#>

**22 years!**

Concerns about  
reproducibility and  
duplicated images

Primarily by *New Scientist*

Investigation by University  
Retraction of related paper

June 2007  
Erratum published

# Publications: the bad

## The most cited paper to be retracted

### Pluripotency of mesenchymal stem cells derived from adult marrow

> Nature. 2002 Jul 4;418(6893):41-9. doi: 10.1038/nature00870. Epub 2002 Jun 20.

	Published	Accesses	Citations
Original Article	20 June 2002	27k	4492
Erratum Note	14 June 2007	6289	39
Retraction Note	17 June 2024	19k	1

(Numbers as of Oct. 2024)

Most authors agree with the retraction\*

<https://retractionwatch.com/the-retraction-watch-leaderboard/top-10-most-highly-cited-retracted-papers/>

No. 7 is Andrew Wakefield original MMR paper

#### Accountability

Accurate

Care

Collegiality

Cooperation

#### Ethics

Fair

Honesty

Objectivity

Openness

Quality

Reliability

Reproducibility

Respect

Responsibility

Rigor

Transparency

# Publications: the ugly

## Generative AI Concerns

Accountability

Accurate

Care

Collegiality

Cooperation

Ethics

Fair

Honesty

Objectivity

Openness

Quality

Reliability

Reproducibility

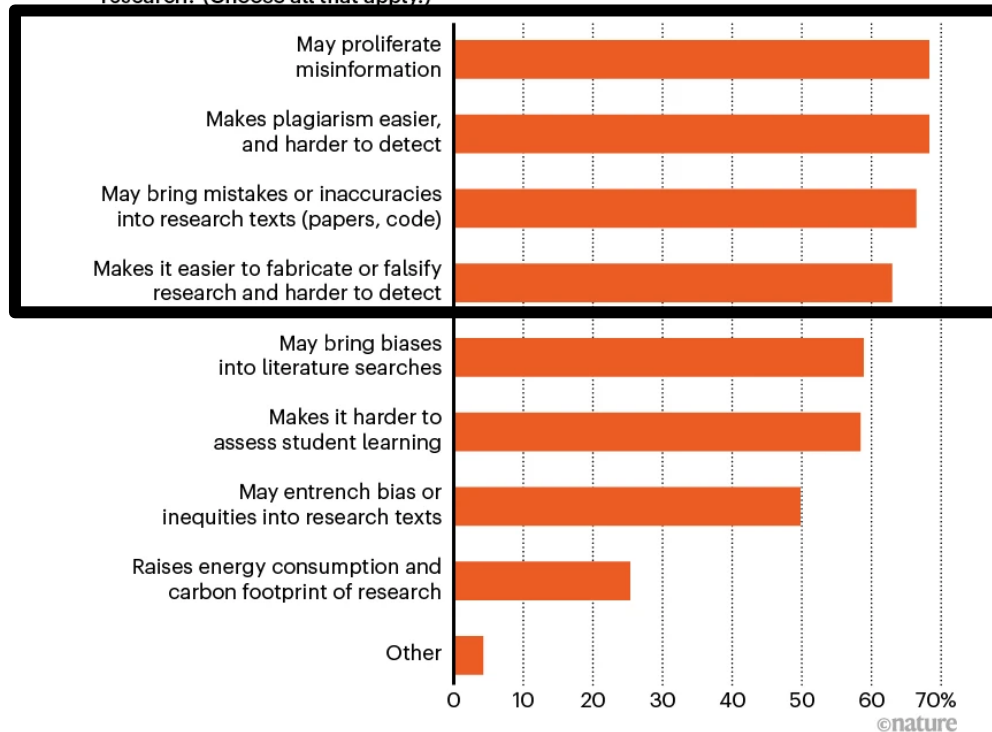
Respect

Responsibility

Rigor

Transparency

Q: Where do you think generative AI may have negative impacts on research? (Choose all that apply.)



# Publications: the ugly Paper Mills

## Accountability

Accurate

Care

Collegiality

Cooperation

## Ethics

Fair

Honesty

Objectivity

Openness

Quality

Reliability

Reproducibility

Respect

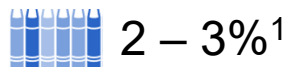
Responsibility

Rigor

Transparency



## Estimated Prevalence



2 – 3%<sup>1</sup>



11%<sup>2</sup>



2 – 46%<sup>3</sup>

1. <https://www.nature.com/articles/d41586-023-03464-x>

2. <https://www.medrxiv.org/content/10.1101/2023.05.06.23289563v2>

3. <https://publicationethics.org/node/55256>



In 2023 Hindawi closed 4 journals because they were “heavily compromised by paper mills”

# Publications: the balance

## **Accountability**

Accurate

Care

Collegiality

Cooperation

## **Ethics**

Fair

Honesty

Objectivity

Openness

Quality

Reliability

Reproducibility

Respect

Responsibility

Rigor

Transparency

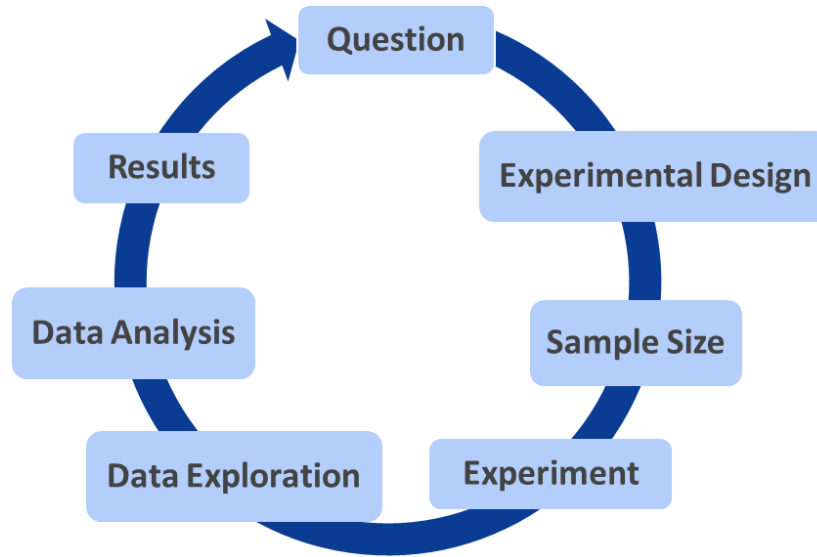


# Misconduct

# Research Integrity

## Misconduct

**Accountability**  
**Accurate**  
**Care**  
**Collegiality**  
**Cooperation**  
**Ethics**  
**Fair**  
**Honesty**  
**Objectivity**  
**Openness**  
**Quality**  
**Reliability**  
**Reproducibility**  
**Respect**  
**Responsibility**  
**Rigor**  
**Transparency**





# Misconduct: How?

Accountability

Accurate

Care

Collegiality

Cooperation

Ethics

Fair

**Honesty**

Objectivity

Openness

Quality

Reliability

Reproducibility

Respect

Responsibility

Rigor

Transparency



There are **many ways to misbehave** when it comes to research

**Plagiarism**

**Fabrication and falsification**

**Inappropriate image manipulation**

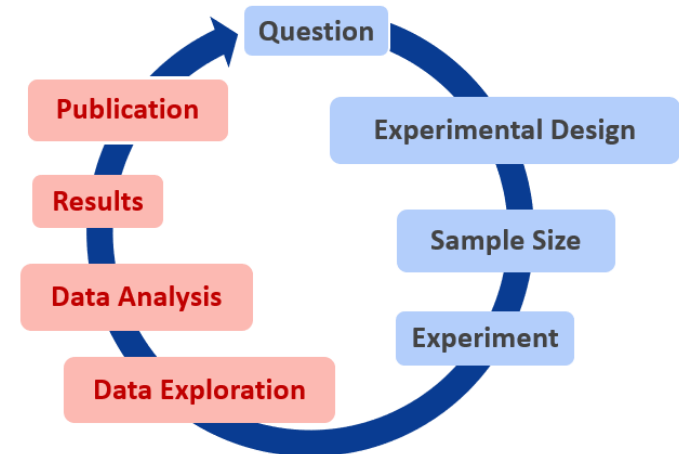
**Non-publication of data**

**Faulty data-gathering procedures**

**Poor data storage and retention**

**Misleading authorship**

**Sneaky publication practices**



# Misconduct: Why?

## Accountability

Accurate  
Care

Collegiality  
Cooperation

Ethics  
Fair

Honesty  
Objectivity

Openness  
Quality

Reliability  
Reproducibility

Respect  
Responsibility

Rigor  
Transparency



T



R

A



G

E



D

I

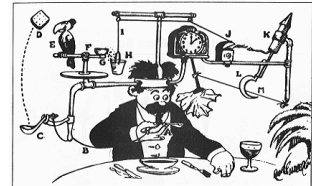


E

S



Tina Gunsalus



# Misconduct: Spot the TRAGEDIES



**Group and authority pressure**

**Entitlement**

**Deception**

**Ambition**

**Incrementalism**

**Rationalisation**

**Embarrassment**

**Temptation**

**Stupid systems**

**A Simple Request from their Professor**

Verify the numbers in a figure matched the spreadsheet

**A Questionable Outcome**

Paper published with the new student as a co-author

**They accept the credit**  
**Why?**



**The New Student**



**The Professor**

A real case from Tina Gunsalus

<https://www.nature.com/articles/d41586-018-05145-6>

# Misconduct: In a nutshell

Accountability

Accurate

Care

Collegiality

Cooperation

Ethics

Fair

**Honesty**

Objectivity

Openness

Quality

Reliability

Reproducibility

Respect

Responsibility

Rigor

Transparency



“I never met anyone who said, yeah, you know, that was the day I woke up and decided to you know, put my career at risk, potentially go to jail, embarrass my family, lose my job.”



Tina Gunsalus

**We need to be aware of our feelings and circumstances  
And acknowledge how they might influence us**

# The Only Acceptable Tragedy!\*

**STEPS** 

*Tragedy • The Dance - How To...*



**STEP 1** - Put your hands up beside your ears in 'Tragedy' shock!



**STEP 2a** - Raise your right arm.



**STEP 2b** - Raise your left arm so both arms are up.



**STEP 3** - Clasp both hands together on your heart.



**STEP 4** - Stretch both arms out straight in front with your palms up.



**STEP 5a** - Put your hands to your head as in STEP 1 and then step to your right.



**STEP 5b** - Repeat the same move to your left. Repeat the same two moves again.



**STEP 6a** - Step to the left and swing your left arm out with your right hand on your hip.



**STEP 6b** - Step to the left and swing your right arm out with your left hand on your hip.



**STEP 7** - Hold your right arm out and stop the traffic!



**STEP 8** - Keep your right arm out, turn to your right and roll your left shoulder three times.

<https://www.youtube.com/watch?v=OiwDHHcHPh0>

\*To be fully transparent we should note the Bee Gees did it first!

# The bottom line

# Research Integrity

## It's about being a good scientist



**Research integrity**  
is about **owning every step of our research, and**  
**benefits everyone.**



# Research Integrity

```
graph TD; RI[Research Integrity] --> Q[Questioning]; RI --> D[Definitions]; RI --> P[Publications]; RI --> M[Misconduct]; RI --> R[Responsibility]; RI --> E[Ethics]; RI --> K[Keeping Track]; RI --> I[In Practice]; RI --> B[Bottom line]; RI --> GS[Good Science]
```

What does it mean?

Questioning

Definitions

Publications

Misconduct

When is it tested?

Responsibility

Keeping Track

Ethics

In Practice

Bottom line

How can we apply it?

Good Science



**What do you think now?  
Which words are  
most important/synonymous with integrity?**



# Research Integrity

## More than words

A word cloud of research integrity principles arranged in a circular pattern. The words are in various sizes and orientations, creating a sense of movement and interconnectedness. The principles include:

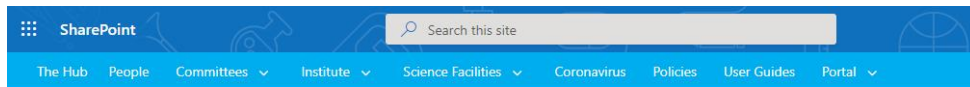
- Responsibility
- Honesty
- Fair
- Objectivity
- Rigor
- Reproducibility
- Accurate
- Collegiality
- Cooperation
- Quality
- Reliability
- Ethics
- Openness
- Care
- Scientific integrity
- Transparency
- Respect
- Accountability



The Hub

# Research Integrity

## Useful resources



Research Integrity

Research Integrity Review Good Research in Practice Human Research Resources Research Forum



Any enquiries regarding Research Integrity can be sent to

[research.integrity@babraham.ac.uk](mailto:research.integrity@babraham.ac.uk)

[Dr Martin Turner](#) is the senior member of staff responsible for overseeing research integrity and is the Institute's first point of contact for anyone wanting more information.

[Mr. Simon Jones](#) is the confidential liaison for whistle-blowers or any other person wishing to raise concerns about the integrity of research being conducted under the auspices of The Babraham Institute.



Trevor Smith  
Health & Safety & QA Manager



Priya Schoenfelder  
Health & Safety Advisor and BI Deputy Bi...



Heather Bath  
Contracts Manager



Created by Fajar Studio  
from Noun Project

BI-HR-001 Code of Conduct  
BI-HR-005 Disciplinary Policy  
BI-COR-010 Whistleblowing Policy  
BI-RES-001 Authorship Policy  
BI-RES-004 Research Misconduct Policy  
BI-RES-005 Research Integrity Policy  
Record Retention Policy  
Data Management Guidance

# Research Integrity

## Useful resources



### Training Courses



[www.bioinformatics.babraham.ac.uk](http://www.bioinformatics.babraham.ac.uk)



MS Teams channel and mailing  
list. Contact

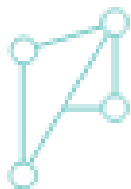
[Richard.Acton@babraham.ac.uk](mailto:Richard.Acton@babraham.ac.uk)

# Research Integrity

## Useful resources

nature

Retraction Watch



**PUBPEER**  
The online Journal club

**UKRIO**  
RESEARCH INTEGRITY OFFICE



“Integrity is doing  
the right thing even  
when no one is  
watching.”

— C.S. Lewis

“Integrity is doing the right thing when you don’t have to—when no one else is looking or will ever know—when there will be no congratulations or recognition for having done so.”

Excerpted from [\*Shattering the Glass Slipper: Destroying Fairy-tale Thinking Before It Destroys You.\*](#)

© 2003 Charles Marshall