Research Integrity

or *How to be a Good Scientist*

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

v2024-02
An illustration
Research Integrity

MMR = Measles, Mumps and Rubella

Autism
1998 **Link** MMR autism
Wakefield *et al.*
n=12 children

2001: **No link** MMR autism
Dales *et al.*
n=600-1900 children each year over 14 years

2002-2005: **No link** MMR autism
Many more studies

1999 **No link** MMR autism
Taylor *et al.*
n=498 children

1998 **Link** MMR autism
Wakefield *et al.*
n=12 children

1999 **No link** MMR autism
Taylor *et al.*
n=498 children

2001: **No link** MMR autism
Dales *et al.*
n=600-1900 children each year over 14 years

2002-2005: **No link** MMR autism
Many more studies

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

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
Research Integrity

- 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)

---

**Retracted article**

See the retraction notice

![Retraction notice](image)


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

A J Wakefield 1, 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: 9503320  DOI: 10.1016/s0140-6736(97)11096-0

---

**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
Research Integrity

- **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

What does it mean?

- Definitions
- Questioning

In Practice

- Ethics
- Keeping Track

How can we apply it?

Responsibility

- Responsibility

Publications

- Good Science
- Bottom line

Misconduct

- When is it tested?
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

Scientific integrity

Accountability
Responsibility

Honesty
Fair
Objectivity

Rigor
Reproducibility

Accurate

Collegiality

Care

Ethics

Transparency

Reliability
Cooperation

Respect

Quality
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 in practice
Research Integrity: In Practice

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

Experimental Design
Sample Size
Data Analysis
Data Exploration
Experiment
Results
Question
Research Integrity in practice
Experimental design
A Practical Translation of the Question

Communicating Ideas

Types of Design

Avoiding Bias

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

Experimental Design

Question

Sample Size

Results

Data Analysis

Data Exploration

Experiment

A Practical Translation of the Question
Accountability  
Accurate  
Care  
Collegiality  
Cooperation  
Ethics  
Fair  
Honesty  
Objectivity  
Openness  
Quality  
Reliability  
Reproducibility  
Respect  
Responsibility  
Rigor  
Transparency

Clear Communication

Library

RNA seq Library

Library, Cambridge

Input RNA

Fragmentation

Fragmented RNA

Convert to cDNA and add sequencing adaptors

DNA Library
Clear Communication

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

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

Experimental Design

Sample
Sample of mice

Tissue Sample

Sample Size
Data Analysis
Data Exploration
Experiment
Results

Question
**Appropriate Type of Design**

Distinguish real differences from experimental artefacts...

...But experiments can be big and complicated

**Seminal Wellcome Trust GWAS Study:**
- 14,000 cases of 7 diseases & 3000 shared controls
- Each processed at different sites and genotyped on distinct series of plates

**Differences between control and cases are confounded by plate**
Appropriate Type of Design

GenADA multi-site collaborative study:
- 875 Alzheimers patients, 850 controls, 9 sites
- Randomised Block Design

Still have differences between plates  Doesn’t confound the experiment

https://blog.goldenhelix.com/stop-ignoring-experimental-design-or-my-head-will-explode/
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

Objective facts

What confirms your beliefs

What you see
Scientists should be **blind** ... if they want to be **objective**!
Blinding is important to avoid Confirmation bias.
Biases in our sample populations can impact on our conclusions too.
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

Patient information is entered into a computer
The computer randomly assigns patients to two or more groups, helping to prevent bias
Control group receives standard therapy
Investigational group receives new treatment
Random Allocation

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
Randomisation More Broadly

Consider Nuisance Factors

Random Allocation

Also consider randomisation throughout the experiment

https://eda.nc3rs.org.uk/experimental-design-allocation#randomisationstrategy
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
Research Integrity in practice
Sample size
The Right Size to be Confident in Our Results

- Statistical Power
- Replicates
<table>
<thead>
<tr>
<th><strong>Accountability</strong></th>
<th><strong>Accurate</strong></th>
<th><strong>Care</strong></th>
<th><strong>Collegiality</strong></th>
<th><strong>Cooperation</strong></th>
<th><strong>Ethics</strong></th>
<th><strong>Fair</strong></th>
<th><strong>Honesty</strong></th>
<th><strong>Objectivity</strong></th>
<th><strong>Openness</strong></th>
<th><strong>Quality</strong></th>
<th><strong>Reliability</strong></th>
<th><strong>Reproducibility</strong></th>
<th><strong>Respect</strong></th>
<th><strong>Responsibility</strong></th>
<th><strong>Rigor</strong></th>
<th><strong>Transparency</strong></th>
</tr>
</thead>
</table>

**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?"

**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)?”

<table>
<thead>
<tr>
<th><strong>How long did the child spend looking?</strong></th>
<th><strong>How many mice do you look at? Sample size</strong></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th><strong>How big is the tool?</strong></th>
<th><strong>How big is the difference? The absolute effect</strong></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th><strong>How messy is the basement?</strong></th>
<th><strong>How messy the data are? Variability</strong></th>
</tr>
</thead>
</table>

*Statistical power: an analogy*
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)?”

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

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

How messy the data are? **Variability**

Little to no control

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

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

A  Yes      B  No

The Importance of Statistical Power

The Problem of the “Inflation Effect”...

The TRUE effect size = 1

- n=3, power = 0.157
- n=50, power = 0.999

https://royalsocietypublishing.org/doi/10.1098/rsos.140216
Defining Replicates

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

Technical versus biological replicates

- **Technical**: n=1
- **Biological**: n=3
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....
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

A Faithful and Informative View of Our Data

Understanding Our Data
Presenting Our Data Honestly

Question
Experimental Design
Sample Size
Experiment
Data Analysis
Results

Data Exploration
Data Exploration

Understanding Our Data:
- The Biology
- The Quality

Builds Our Understanding and Confidence
Data Exploration: Understanding Our Data

Often we summarise our data to key values

<table>
<thead>
<tr>
<th>N</th>
<th>182</th>
</tr>
</thead>
<tbody>
<tr>
<td>X Mean</td>
<td>54.26</td>
</tr>
<tr>
<td>Y Mean</td>
<td>47.83</td>
</tr>
<tr>
<td>X SD</td>
<td>16.76</td>
</tr>
<tr>
<td>Y SD</td>
<td>26.93</td>
</tr>
<tr>
<td>Correlation</td>
<td>-0.06</td>
</tr>
</tbody>
</table>

Can be really useful...

...And also really not!
Data Exploration: Anscombe’s Quartet

“A computer should make both calculations and graphs”

<table>
<thead>
<tr>
<th>Dataset 1</th>
<th>Dataset 2</th>
<th>Dataset 3</th>
<th>Dataset 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>Y1</td>
<td>X2</td>
<td>Y2</td>
</tr>
<tr>
<td>10.00</td>
<td>8.04</td>
<td>10.00</td>
<td>9.14</td>
</tr>
<tr>
<td>8.00</td>
<td>6.95</td>
<td>8.00</td>
<td>8.14</td>
</tr>
<tr>
<td>13.00</td>
<td>7.58</td>
<td>13.00</td>
<td>8.74</td>
</tr>
<tr>
<td>9.00</td>
<td>8.81</td>
<td>9.00</td>
<td>8.77</td>
</tr>
<tr>
<td>11.00</td>
<td>8.33</td>
<td>11.00</td>
<td>9.26</td>
</tr>
<tr>
<td>14.00</td>
<td>9.96</td>
<td>14.00</td>
<td>8.10</td>
</tr>
<tr>
<td>6.00</td>
<td>7.24</td>
<td>6.00</td>
<td>6.13</td>
</tr>
<tr>
<td>4.00</td>
<td>4.26</td>
<td>4.00</td>
<td>3.10</td>
</tr>
<tr>
<td>12.00</td>
<td>10.84</td>
<td>12.00</td>
<td>9.13</td>
</tr>
<tr>
<td>7.00</td>
<td>4.82</td>
<td>7.00</td>
<td>7.26</td>
</tr>
<tr>
<td>5.00</td>
<td>5.68</td>
<td>5.00</td>
<td>4.74</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>X3</th>
<th>Y3</th>
<th>X4</th>
<th>Y4</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.00</td>
<td>7.46</td>
<td>8.00</td>
<td>6.58</td>
</tr>
<tr>
<td>8.00</td>
<td>6.77</td>
<td>8.00</td>
<td>5.76</td>
</tr>
<tr>
<td>13.00</td>
<td>12.74</td>
<td>8.00</td>
<td>7.71</td>
</tr>
<tr>
<td>9.00</td>
<td>7.11</td>
<td>8.00</td>
<td>8.84</td>
</tr>
<tr>
<td>11.00</td>
<td>7.81</td>
<td>8.00</td>
<td>8.47</td>
</tr>
<tr>
<td>14.00</td>
<td>8.84</td>
<td>8.00</td>
<td>7.04</td>
</tr>
<tr>
<td>6.00</td>
<td>6.08</td>
<td>8.00</td>
<td>5.25</td>
</tr>
<tr>
<td>4.00</td>
<td>5.39</td>
<td>19.00</td>
<td>12.50</td>
</tr>
<tr>
<td>12.00</td>
<td>8.15</td>
<td>8.00</td>
<td>5.56</td>
</tr>
<tr>
<td>7.00</td>
<td>6.42</td>
<td>8.00</td>
<td>7.91</td>
</tr>
<tr>
<td>5.00</td>
<td>5.73</td>
<td>8.00</td>
<td>6.89</td>
</tr>
</tbody>
</table>

4 datasets
Each consisting of X and Y variable
All 4 datasets have the same summary statistics...
Data Exploration: Anscombe’s Quartet

The average and spread of the conditions appears the same

<table>
<thead>
<tr>
<th>Dataset 1</th>
<th>Dataset 2</th>
<th>Dataset 3</th>
<th>Dataset 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>Y1</td>
<td>X2</td>
<td>Y2</td>
</tr>
<tr>
<td>N</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Mean</td>
<td>9.00</td>
<td>7.50</td>
<td>9.00</td>
</tr>
<tr>
<td>STD</td>
<td>3.31</td>
<td>2.03</td>
<td>3.31</td>
</tr>
<tr>
<td>SEM</td>
<td>1.00</td>
<td>0.612</td>
<td>1.00</td>
</tr>
<tr>
<td>X3</td>
<td>Y3</td>
<td>X4</td>
<td>Y4</td>
</tr>
<tr>
<td>N</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Mean</td>
<td>9.00</td>
<td>7.50</td>
<td>9.00</td>
</tr>
<tr>
<td>STD</td>
<td>3.31</td>
<td>2.03</td>
<td>3.31</td>
</tr>
<tr>
<td>SEM</td>
<td>1.00</td>
<td>0.612</td>
<td>1.00</td>
</tr>
</tbody>
</table>

The relationship between X & Y can be described the same

<table>
<thead>
<tr>
<th>Correlation (r)</th>
<th>Line of Best Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>x1 vs. y1</td>
<td>r = 0.8164</td>
</tr>
<tr>
<td>x2 vs. y2</td>
<td>r = 0.8162</td>
</tr>
<tr>
<td>x3 vs. y3</td>
<td>r = 0.8163</td>
</tr>
<tr>
<td>x4 vs. y4</td>
<td>r = 0.8165</td>
</tr>
</tbody>
</table>
Data Exploration: Anscombe’s Quartet

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

What about the relationship between X & Y?

$r = 0.816 \quad Y = 0.500 \times X + 3.00$
Can You Predict The Data Structure?

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

<table>
<thead>
<tr>
<th>Line</th>
<th>A</th>
<th>N</th>
<th>182</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>X Mean</td>
<td>54.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Y Mean</td>
<td>47.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td>X SD</td>
<td>16.76</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Y SD</td>
<td>26.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Correlation</td>
<td>-0.06</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Circle</th>
<th>B</th>
<th>C</th>
<th>Unstructured</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>D</td>
<td>I Don’t Know...A Dinosaur!</td>
</tr>
</tbody>
</table>

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

A

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

And More!

https://www.research.autodesk.com/publications/same-stats-different-graphs/
Data Exploration: Seeing is Believing!

Beware of statistical or graphical summaries...

...without proper exploration & visualisation!

<table>
<thead>
<tr>
<th>Correlation (r)</th>
<th>Line of Best Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>x1 vs. y1</td>
<td>$r = 0.8164, Y = 0.5001X + 3.000$</td>
</tr>
<tr>
<td>x2 vs. y2</td>
<td>$r = 0.8162, Y = 0.5000X + 3.001$</td>
</tr>
<tr>
<td>x3 vs. y3</td>
<td>$r = 0.8163, Y = 0.4997X + 3.002$</td>
</tr>
<tr>
<td>x4 vs. y4</td>
<td>$r = 0.8165, Y = 0.4999X + 3.002$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>182</td>
</tr>
<tr>
<td>X Mean</td>
<td>54.26</td>
</tr>
<tr>
<td>Y Mean</td>
<td>47.83</td>
</tr>
<tr>
<td>X SD</td>
<td>16.76</td>
</tr>
<tr>
<td>Y SD</td>
<td>26.93</td>
</tr>
<tr>
<td>Correlation</td>
<td>-0.06</td>
</tr>
</tbody>
</table>

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

Exercise
Values from 3 conditions from one experiment

What do you think of this graph?
Representing Dataset 1: 
**Create** the most informative graph

https://tinyurl.com/RIdataExp  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
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

Experiment

Data Analysis

Sample Size

Experimental Design

Results

Question

Mean +/- SEM

Median +/- 95% CI

Median

Condition 1
Condition 2
Condition 1
Condition 2
Condition 1
Condition 2
Data Exploration...

...Dubious Datasets
Example 1: A Knockout?

YAP and TAZ control peripheral myelination and the expression of laminin receptors in Schwann cells.
Example 2: A Case of Mistaken Identity

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

% Alignment to Mouse

<table>
<thead>
<tr>
<th>Condition</th>
<th>% Alignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition 1</td>
<td>50%</td>
</tr>
<tr>
<td>Condition 2</td>
<td>50%</td>
</tr>
</tbody>
</table>

% Alignment to Mycoplasma

<table>
<thead>
<tr>
<th>Condition</th>
<th>% Alignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition 1</td>
<td>50%</td>
</tr>
<tr>
<td>Condition 2</td>
<td>50%</td>
</tr>
</tbody>
</table>
All highlight the perils of assuming and not exploring....

...Is it okay if you don’t know there’s a problem?
Analytical Tools

The right tools for the job!

- Know what's out there
- Learn how to use it
Help With Analytical Tools

Core Skills Courses
e.g. R Programming

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

Formalise Our Findings with Statistics

Understanding Our Data
Understanding Our Question

Question

Experimental Design

Sample Size

Experiment

Data Exploration

Data Analysis

Results

Experimental Design

Sample Size

Experiment

Data Exploration

Data Analysis

Results

Question
Translate the hypothesis/question into statistical questions

By choosing the right test!

Healthy approach:

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

...It’s knowing what the test should do”
Knowing what the Test should do...

Statistics Decision tree

- Start
  - Differences?
    - How many factors?
      - Two or more
        - 2 way ANOVA, General Linear (Mixed) Model, etc.
      - One
        - Same or different subjects?
          - Same
            - Parametric
              - Pearson Correlation
            - Nonparametric
              - Spearman Rank Correlation
          - Different
            - Parametric
              - Paired T-test / repeated ANOVA
            - Nonparametric
              - Wilcoxon paired test
            - T-test / ANOVA
            - Mann Whitney U

Accountability
Accurate
Care
Collegiality
Cooperation
Ethics
Fair
Honesty
Objectivity
Openness
Quality
Reliability
Reproducibility
Respect
Responsibility
Rigor
Transparency
Knowing what the Test should do...

Statistics Using R

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
Data exploration should always be a pivotal step of analysis.

Stats helps us formalise our findings.
Research Integrity

What does Ethics mean?
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*
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
Research Integrity
The 3 Rs

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

Replace
Reduce
Refine

https://wellcome.org/sites/default/files/wtp057673_0.pdf
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

Reduce

Minimising the number of animals used consistent with scientific aims.
Research Integrity
The 3Rs: about animal welfare

Refine

Minimising the pain, suffering, distress or lasting harm that research animals might experience.
Research Integrity
The 3 Rs at Babraham: AWERB

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

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

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

Reporting
Defend data
Trace
Find it
Share it
Describe

Protect IP
Legal protection
GLP++
Safe and secure
Compliant
Accountability
Accuracy
Care
Collegiality
Cooperation
Ethics
Fair
Honesty
Objectivity
Openness
Quality
Reliability
Reproducibility
Respect
Responsibility
Rigor
Transparency

Research Integrity
Laboratory Notebooks

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

This Photo by Unknown Author is licensed under CC BY-ND
Research Data

Data Storage

Legal Requirements

Practicalities

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

- 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 Ribbit 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
Reproductibility
Respect
Responsibility
Rigor
Transparency

Question

Results

Experimental Design

Sample Size

Data Exploration

Experiment

Data Analysis
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
Research Integrity

Review

Activity seeks input and questioning
Don't take it personally
Be positive in your criticism of others' science

FAIR

Findable
Accessible
Interoperable
Reproducible

Reviewing your science in FAIR can maximize the value of your science and your data.

You data is 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 others, and ultimately adds 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 doesn't need to be perfect to make data accessible as possible - it benefits everyone.

All 4 we recommend providing (numerical) data as you generate new data, and we generate in CF.

It's also important that FAIR use of public data as much as possible.

Some (e.g. code) are not perfect, but we should try to make it as perfect as possible.

You should also make sure of work is based on existing paper/data, that you do a 90% check on the initial data.

This is also a testing opportunity in terms of checking and applying the type of data you can share publicly.

How does this link to research integrity?

Pre-prints

What are the pros and cons?

How does this link to research integrity?
Research Integrity

In a nutshell

Applying research integrity principles is our responsibility as scientists
Research Integrity

In the lab

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

https://www.nature.com/articles/nj7593-263a
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’
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

Graduate Studies
Pastoral Mentors
EAP
Mental Health First Aiders
HR
The Hub

Mental Health First Aiders
HR
Pastoral Mentors
EAP
Graduate Studies
Research Integrity

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

https://forbetterscience.com/2020/07/07/science-misconduct/
Accountability
Accurate
Care
Collegiality
Cooperation
Ethics
Fair
Honesty
Objectivity
Openness
Quality
Reliability
Reproducibility
Respect
Responsibility
Rigor
Transparency

Research Integrity

PhD ➔ PostDoc ➔ Group Leader ➔ Tenure

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?
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

Need to publish
Dissemination of knowledge
Peer review
Paper published
Grant, Job
More research, more grants, more papers
Career advancement
More jobs in research
More research

Publications: the good
Publications: the bad

- Need to publish
- Dissemination of knowledge
- Peer review
- Novel Significant Results
  - Rules are ignored, bent, broken
- Career advancement
  - More jobs in research
- More research
  - More grants, more papers
- Grant, Job
- Paper published

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

Amgen attempted to replicate 53 high-impact cancer research studies: able to replicate only six. **Replication crisis**

Random sample of 1,000 Medline abstracts: 96% reported significant p-values
Publications: the ugly
Paper Mills

Estimated Prevalence

- 2 – 3%¹
- 11%²
- 2 – 46%³

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

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
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
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

https://undsci.berkeley.edu/article/socialsideofscience_06
Misconduct: Why?

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

https://www.nature.com/articles/d41586-018-05145-6
Misconduct: Spot the TRAGEdies

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?

A real case from Tina Gunsalus
https://www.nature.com/articles/d41586-018-05145-6
“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.”

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

Tina Gunsalus
The Only Acceptable Tragedy!*

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 - Cross 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.

https://penntoday.upenn.edu/news/pursuit-scientific-truth-adversarial-collaboration-Tetlock-Clark
Research Integrity

What does it mean?

Questioning
Definitions

Responsibility

In Practice
Keeping Track
Ethics

Publications
Misconduct

Good Science

Bottom line

When is it tested?

How can we apply it?
What do you think now?
Which words are most important/synonymous with integrity?
Scientific integrity  
Accountability  
Care  
Openness  
Respect  
Ethics  
Reliability  
Cooperation  
Quality  
Responsibility  
Honesty  
Fair  
Objectivity  
Rigor  
Reproducibility  
Accurate  
Collegiality
Research Integrity

Useful resources

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.

Any enquiries regarding Research Integrity can be sent to research.integrity@babraham.ac.uk

Trevor Smith
Health & Safety & QA Manager

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