

# Introduction to Experimental Design

7 December 2016

Aaron Lun

(based on slides by Rory Stark and Sarah Vowler)

# Agenda

- Why perform experiments?
- Why think about experimental design?
- What makes for a well designed experiment?
- Aspects of experimental design
  - Experimental variables
  - Power: variance and replicates
  - Bias: confounding factors, randomisation, and controls
- Experimental design types

# Why Perform Experiments?

## **Scientific method:**

1. Form a hypothesis about a phenomenon
2. Set up an experiment to test the hypothesis
3. Does data support/refute hypothesis?
4. Go to 1 and repeat.

Statistical analysis deals with **3**...

but depends on experimental design in **2**.

# Why Think About Experimental Design?

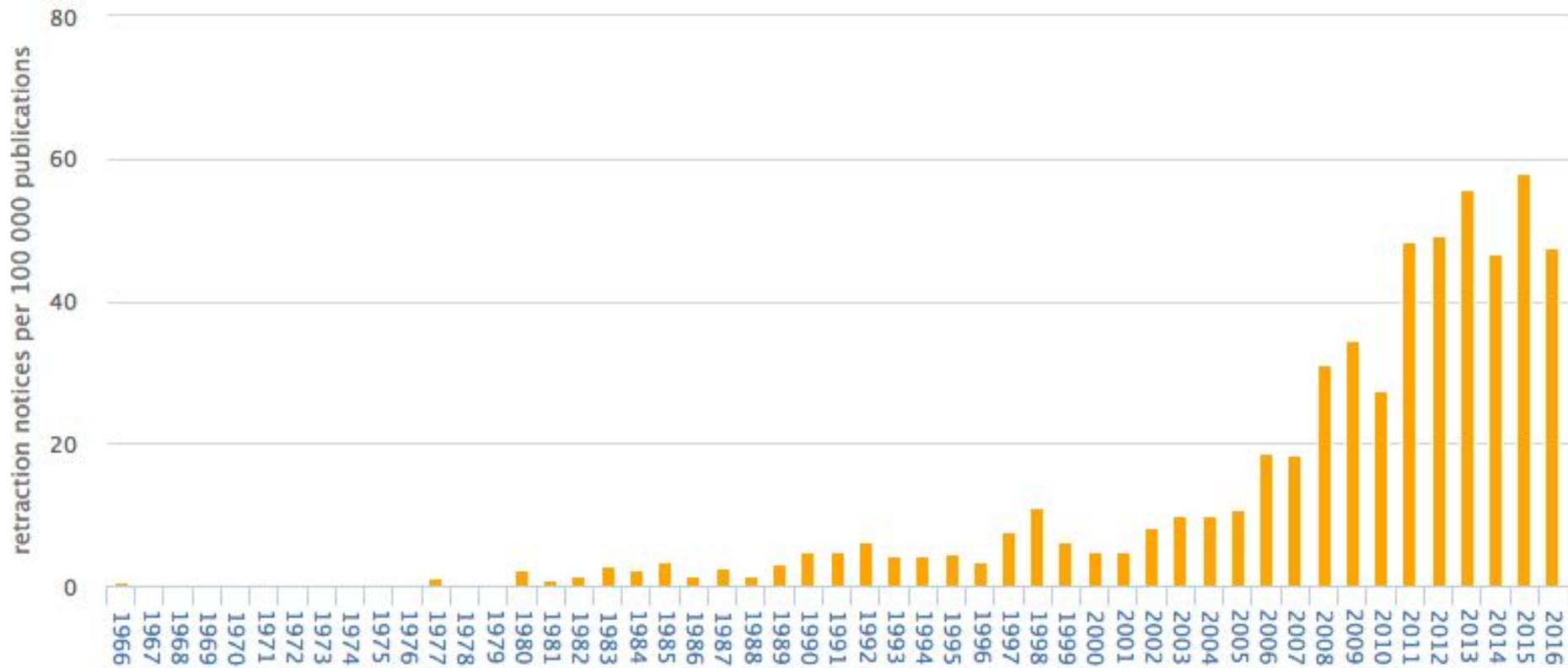
*Does your experiment have the capacity to answer your scientific question?*

*Can the results of your experiment be reproduced by yourself and others?*

If not, you are wasting your own **time** and **money**.

# *Crises in Reproducible Research!!*

Retraction notices per 100 000 publications by year of Entrez record creation



<http://rpubs.com/neilfws/65778>

# 47 of 53 high-profile cancer studies were not reproducible!



NATURE | COMMENT



## Drug development: Raise standards for preclinical cancer research

C. Glenn Begley & Lee M. Ellis

Affiliations | Corresponding author

*Nature* **483**, 531–533 (29 March 2012) | doi:10.1038/483531a

Published online 28 March 2012

# Consequences of Poor Experimental Design...

- **Cost** of experimentation. We have a responsibility to our funding bodies (taxpayers, donors)
- **Limited & Precious** material, esp. clinical samples.
- **Immortalization** of data sets in public databases and methods in the literature. Our bad science begets more bad science.
- **Ethical concerns** of experimentation: animals and clinical samples.

# A Well-Designed Experiment

Should have

- Clear objectives
- Focus and simplicity
- Sufficient power
- Randomised comparisons

And be

- Precise
- Unbiased
- Amenable to statistical analysis
- Reproducible



# Ronald A. Fisher(1890-1962)



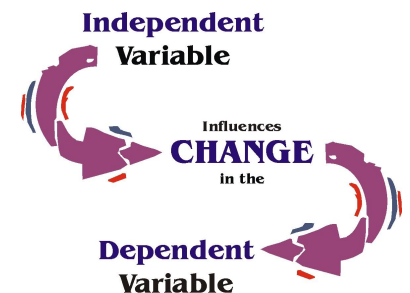
*“To consult the statistician **after** an experiment is finished is often merely to ask him to conduct a **post mortem** examination. He can perhaps say what the experiment died of.” (1938)*

# Aspects of Experimental Design

- Experimental factors
- Minimising Bias
- Sources of variance
- Replicates - why and how many?

# Experimental Factors

- Factors: aspects of experiment that change and **influence the outcome** of the experiment
  - e.g. time, weight, drug, gender, ethnicity, country, plate, cage etc.
  - some of interest and to be varied by the experimenter
  - others not of interest to be kept constant
- Independent and Dependent variables
  - Independent variable (IV): what you change
  - Dependent variable (DV): what changes due to IV
  - “**If** (**independent** variable), **then** (**dependent** variable)”
- Variable type depends on type of measurement:
  - Categorical (**nominal**) , e.g. gender
  - Categorical with ordering (**ordinal**), e.g. tumour grade
  - **Discrete**, e.g. shoe size, number of cells
  - **Continuous**, e.g. body weight in kg, height in cm



# Confounding Factors

- Also known as **extraneous**, **hidden**, **lurking** or **masking** factors, or the **third variable** or **mediator variable**.
- May mask an actual association or **falsely** demonstrate an apparent association between the independent & dependent variables.
- Simple example:
  - test for change in weight between WT and KO mice
  - but all WT mice are male, and all KO mice are female
  - is the difference due to sex or gene knockout?
- Another example:
  - global average temperature increases over the past 4 centuries
  - number of pirates decreases over the same time period
  - lack of pirates causing global warming?

## Genetic Signatures of Exceptional Longevity in Humans

Paola Sebastiani,<sup>1\*</sup> Nadia Solovieff,<sup>1</sup> Annibale Puca,<sup>2</sup> Stephen W. Hartley,<sup>1</sup> Efthymia Melista,<sup>3</sup> Stacy Andersen,<sup>4</sup> Daniel A. Dworkis,<sup>3</sup> Jemma B. Wilk,<sup>5</sup> Richard H. Myers,<sup>5</sup> Martin H. Steinberg,<sup>6</sup> Monty Montano,<sup>3</sup> Clinton T. Baldwin,<sup>6,7</sup> Thomas T. Perls<sup>4\*</sup>

<sup>1</sup>Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118, USA. <sup>2</sup>IRCCS Multimedica, Milano, Italy; Istituto di Tecnologie Biomediche, Consiglio Nazionale delle Ricerche, Segrate, 20122, Italy. <sup>3</sup>Department of Medicine, Boston University School of Medicine, Boston, MA 02118, USA. <sup>4</sup>Section of Geriatrics, Department of Medicine, Boston University School of Medicine and Boston Medical Center, Boston, MA 02118, USA. <sup>5</sup>Department of Neurology, Boston University School of Medicine, Boston, MA 02118, USA. <sup>6</sup>Departments of Medicine and Pediatrics, Boston University School of Medicine and Boston Medical Center, Boston, MA 02118, USA. <sup>7</sup>Center for Human Genetics, Boston University School of Medicine, Boston, MA 02118, USA.

- GWAS study: 800 centenarians vs. controls
- Found 150 SNPs predicting centenarians with 77 % accuracy
- Problem: they used **different SNP chips** for centenarians and controls
- Retracted in 2011 following independent review and QC of data

# Technical Confounding Factors: Batch Effects



RNA Extraction

Day1, Plate 1

	1	2	3	4	5	6	7	8	9	10	11	12
A												
B												
C												
D												
E												
F												
G												
H												

Control

Day2, Plate 2

	1	2	3	4	5	6	7	8	9	10	11	12
A												
B												
C												
D												
E												
F												
G												
H												

Treatment 1

Day3, Plate 3

	1	2	3	4	5	6	7	8	9	10	11	12
A												
B												
C												
D												
E												
F												
G												
H												

Treatment 2

The difference between Control, Treatment 1 and Treatment 2 is confounded by **day** and **plate**.

# Confounding Factors

- **Inadequate management and monitoring** of confounding factors
  - experimental data becomes difficult to interpret if it's not just the the factor of interest that varies.
  - one of the most common causes of researchers wrongly assuming that a correlation leads to a causality.
- Confounding factors introduce biases into the results, potentially resulting in misleading conclusions.
- If a study does not consider confounding factors, **don't believe it!**

# Forms of Bias

Type of Bias	Description
<b>Selection bias</b>	Systematic differences between <b>baseline characteristics</b> or <b>treatment groups</b> that are being compared.
<b>Performance bias</b>	Systematic differences between groups in exposure to factors other than the interventions of interest (aka <b><i>confounding</i></b> or <b><i>extraneous</i></b> factors).
<b>Attrition bias</b>	Systematic differences between groups due to samples being <b>withdrawn</b> from the study or <b>excluded</b> from the analyses.
<b>Detection or Measurement bias</b>	Systematic differences between groups in how outcomes are assessed or determined, e.g. <b>measurement errors</b> and inefficient use of data.
<b>Reporting bias</b>	Systematic differences between reported and unreported findings due to manipulation in the reporting of findings such as <b>selective or distorted reporting</b> , e.g. papers with more 'interesting results' are more likely to be submitted and accepted for publication.



# Solutions

- Controls
- Randomisation
- Blinding

# Why use controls?

*Is the observed effect caused by my factor of interest?*

- Collect data for control samples
- Collect data for “treatment” samples
- Ensure that **only** the factor of interest differs between control and treatment
- Effect of factor determined by comparing between control and treatment values

# Experimental Controls

- Negative controls = no expected effect:
  - e.g., a WT control when studying a KO phenotype
  - avoid “type I” errors, i.e., false positives
- Positive controls = expected known effect
  - e.g., a qPCR reaction with known template
  - avoid “type II” errors, i.e., false negatives
- Technical controls
  - e.g., standard curve for qPCR, spike-ins
  - Detect/correct technical biases, normalize

# Examples of Experimental Controls




- Wild-type organism (knockouts)
- Inactive siRNA (silencing)
- Vehicle (treatments)
- Input: fragmented chromatin (ChIP)
- Spike-ins (quantification/normalisation)
- “Gold standard” datapoints
- Multi-level controls
  - e.g. contrast Vehicle/Input vs. Treatment/Input

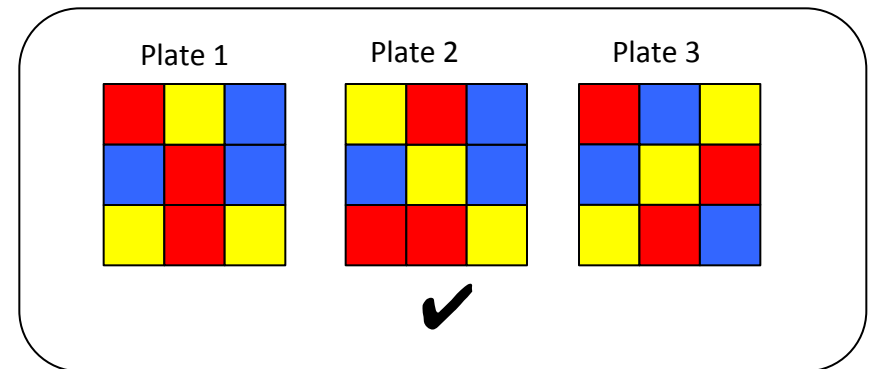
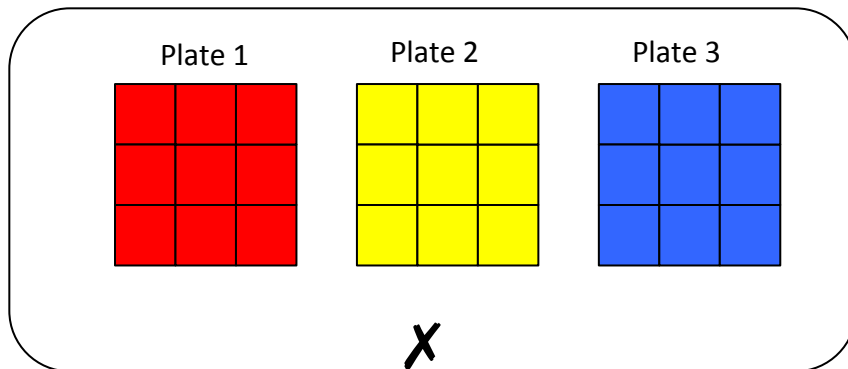
# Randomization

- Some variables cannot be easily controlled
  - e.g., random fluctuations in measuring devices
  - logistics are not feasible
- Randomize to eliminate systematic biases
  - e.g., can't process all samples in a single day
  - randomize the samples to be processed across days, avoid biases due to time effect

# Randomised Block Design

- **Blocking** is the arranging of *experimental units* in groups (blocks) that are similar to one another.

 Control     Treatment 1     Treatment 2



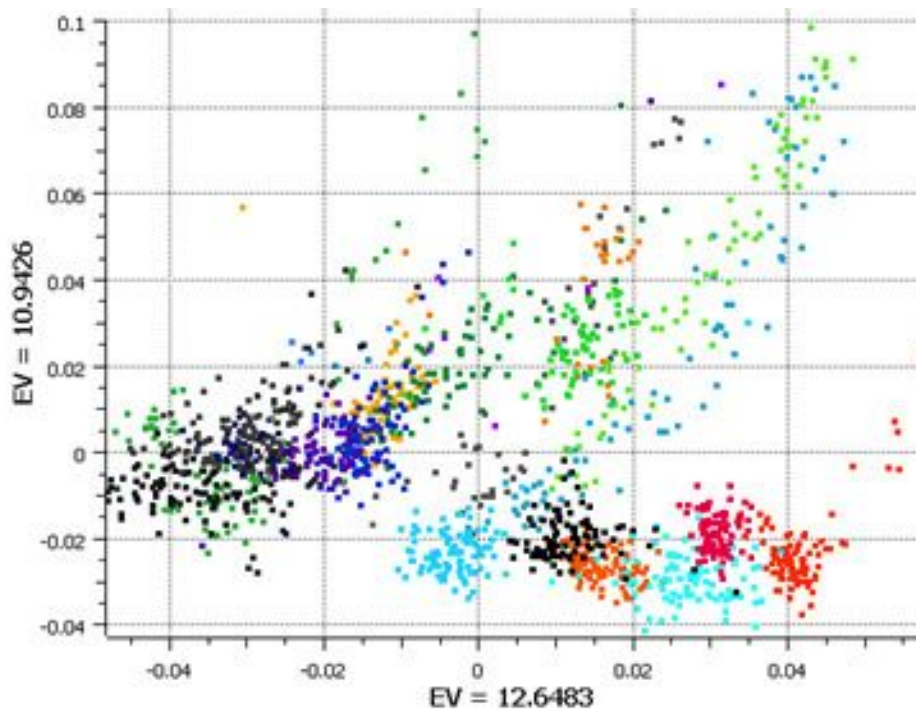
- RBD across plates so that each plate contains spatially randomised **equal proportions** of:
  - Control
  - Treatment 1
  - Treatment 2controlling plate effects.

# Randomised Block Design

**Good** design example: Alzheimer's study from GlaxoSmithKline

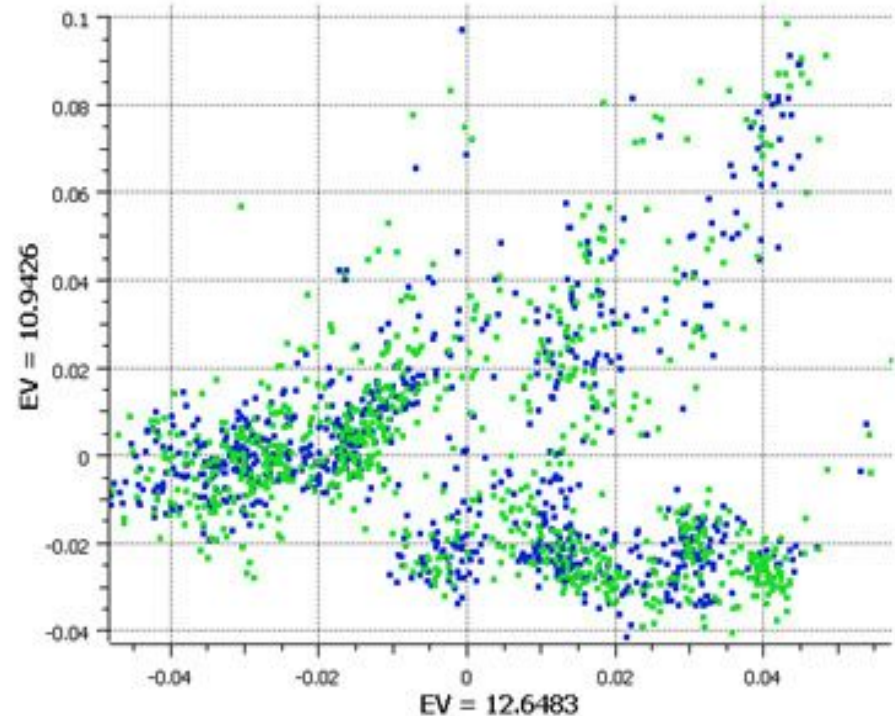
## Plate effects by plate

Left PCA plot show *large plate effects*.  
Each colour corresponds to a different plate



## Plate effects by case/control

Right PCA plot shows each plate cluster contains *equal proportions* of cases (blue) and controls (green).



# Blinding

Especially important where subjective measurements are taken

Unconscious (or conscious) biases of the experimenter may affect the measurements

Blind the labels so that person doing the measuring can only use the data

- e.g., wine tasting, scoring phenotypes...



# Sources of Variation

- Biological “noise”
  - Biological processes are inherently stochastic
  - Single cells, cell populations, individuals, organs, species....
  - Timepoints, cell cycle, synchronized vs. unsynchronized
- Technical noise
  - Reagents, antibodies, temperatures, pollution
  - Platforms, runs, operators
- Consider in advance and control

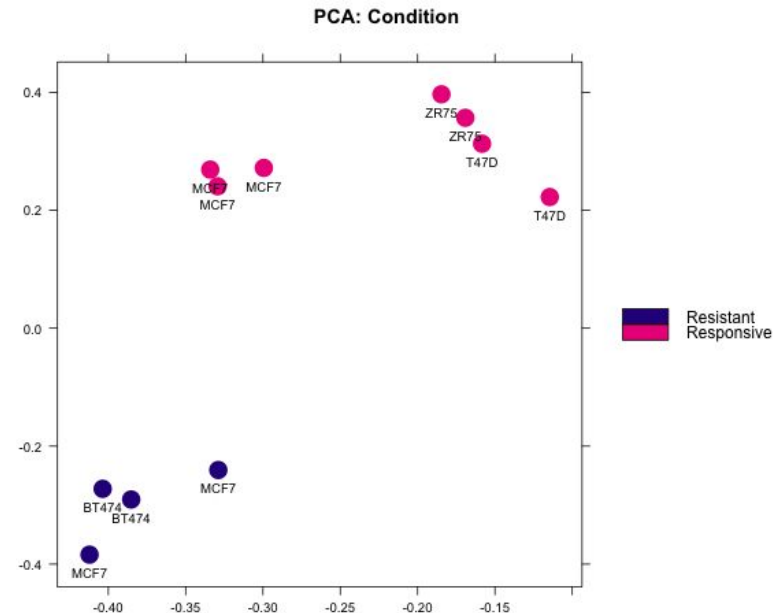
# Replicates

*How can I be sure that my effect is real and reproducible, and not just due to random variability?*

Do it again, and again, and again...

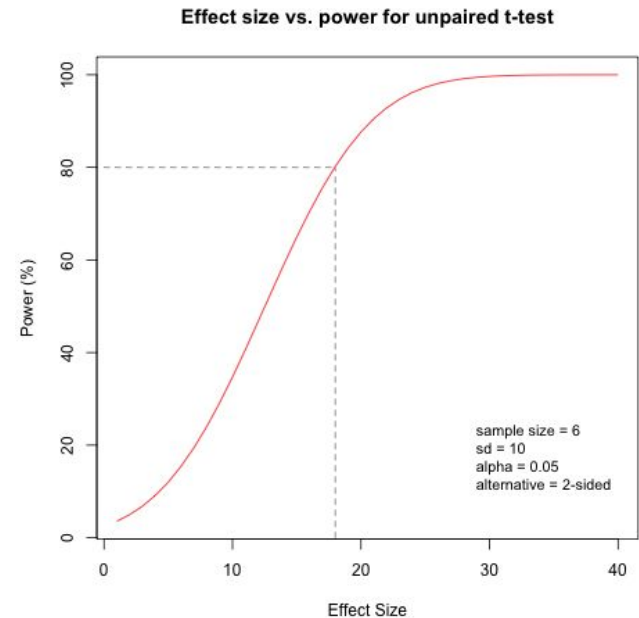
# Types of Replication

- Biological replication:
  - accounts for biological variability
  - *In vivo*:
    - Patients
    - Mice
  - *In vitro*:
    - Different cell lines
    - Re-growing cells (passages)
- Technical replication:
  - accounts for variability in experimental protocol, measurement platform (i.e. sequencer)
  - easier to generate, but less useful



# How many samples?

- Depends on variability and effect size
  - smaller effects = more samples
  - greater var = more samples
  - (assuming that adding more samples doesn't increase the variability)
- Calculating appropriate sample sizes
  - Power calculations
  - Resource equation
- Power: the **probability** of detecting an **effect** of a specified size if present.
  - Calculation requires knowledge of variability; some expectation of the effect; and knowledge of the statistical analysis and acceptable error rate
  - determine **appropriate numbers** of samples (sample size/replicates)
- Mead's resource equation: get enough samples to estimate the error well
  - usually around 10-20 "degrees of freedom" for simple experiments
  - i.e., 10-20 more samples than groups (WT/KO/treated/untreated, etc.)

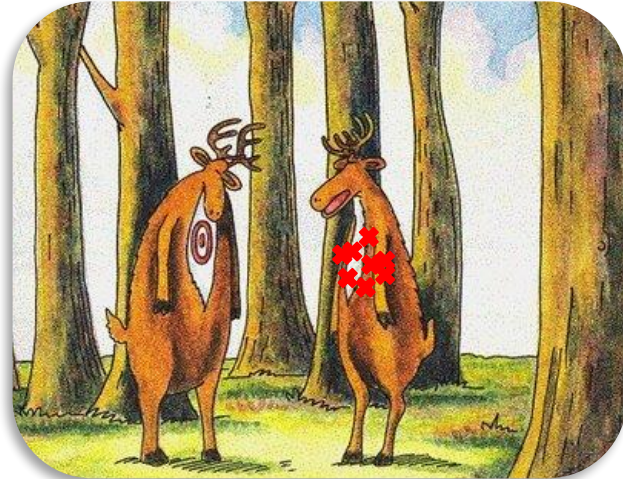
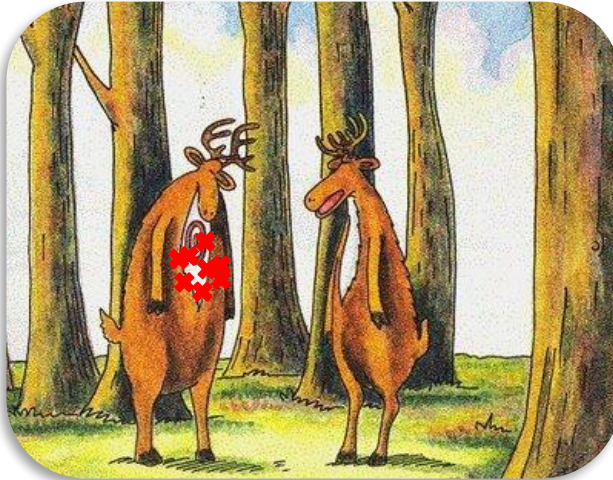


# Precision, Accuracy & Bias

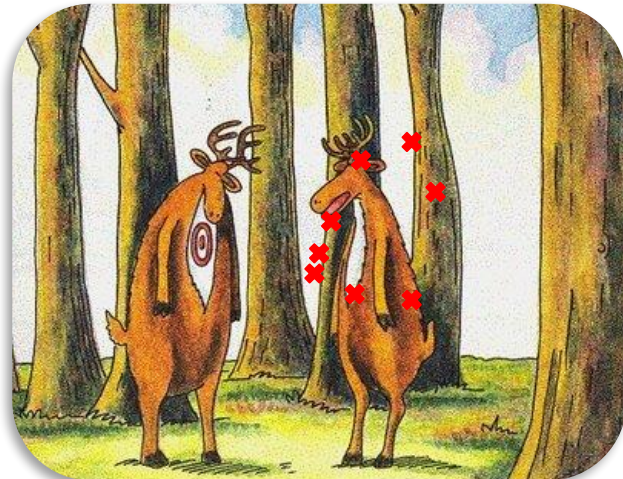
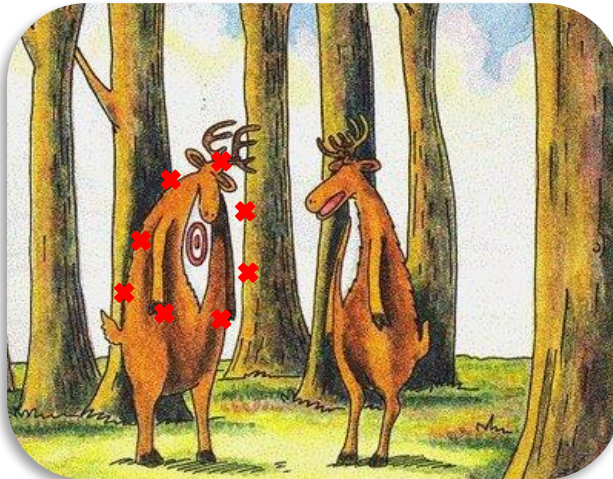
**Accurate**

**Biased**

**Precise**



**Imprecise**



# Types of Experimental Designs

- Block designs: randomisation
- Matched: tumour/normal
- Factorial/multifactorial designs
- Nested designs
- Time series

<https://rawgit.com/bioinformatics-core-shared-training/experimental-design/master/ExperimentalDesignManual.pdf>

# Design Issues: Sequencing Experiments

- Platforms (MiSeq, HiSeq, etc.)
- Library preps
- Multiplexing and pooling strategies
- Single-end vs paired end
- Sequencing depth
  - Coverage
  - Lanes
- Validation