

INTRODUCTION TO EXPERIMENTAL DESIGN AT CRUK-CI

STÉPHANE BALLEREAU

ABBI EDWARDS

ASH SAWLE

CHANDRA CHILAMAKURI

MARK FERNANDEZ

RORY STARK

20 OCTOBER 2021

tinyurl.com/cruk-edesign

Agenda

WHY PERFORM EXPERIMENTS?

WHAT MAKES FOR A WELL DESIGNED EXPERIMENT?

KEY ASPECTS OF EXPERIMENTAL DESIGN

- Experimental variables
- Power: variance and replicates
- Bias: confounding factors, randomisation, and controls

DESIGN PARAMETERS

EXPERIMENTAL DESIGN PROCESS AT CRUK-CI

BREAKOUT SESSIONS: PRACTICALS



UNIVERSITY OF
CAMBRIDGE

2



CANCER
RESEARCH
UK

CAMBRIDGE
INSTITUTE

Why Perform Experiments?

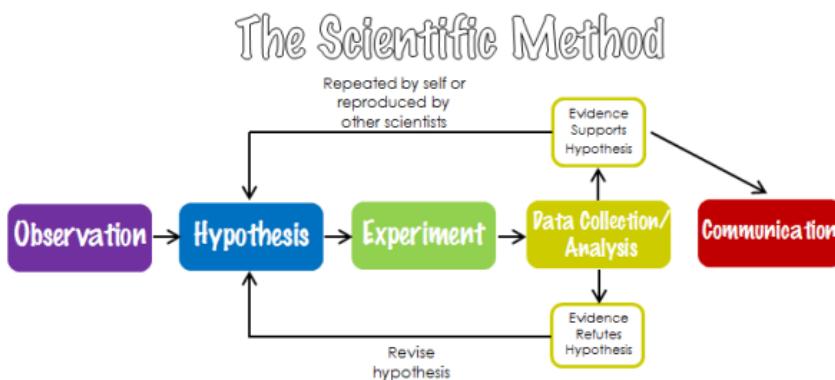
BECAUSE MY SUPERVISOR TOLD ME TO

BECAUSE THEY DID IT IN THIS OTHER PAPER

BECAUSE WE GOT A COOL NEW PIECE OF TECH AND I WANT TO TRY IT OUT

BECAUSE I DON'T KNOW WHAT ELSE TO DO

Experiment is to Test Hypothesis



- Hypothesis
 - A prediction statement
 - Includes clearly defined independent and dependent variables
 - Includes measurable or testable scenario
- Best experiments have clearly defined hypothesis

Reproducible Research



UNIVERSITY OF
CAMBRIDGE

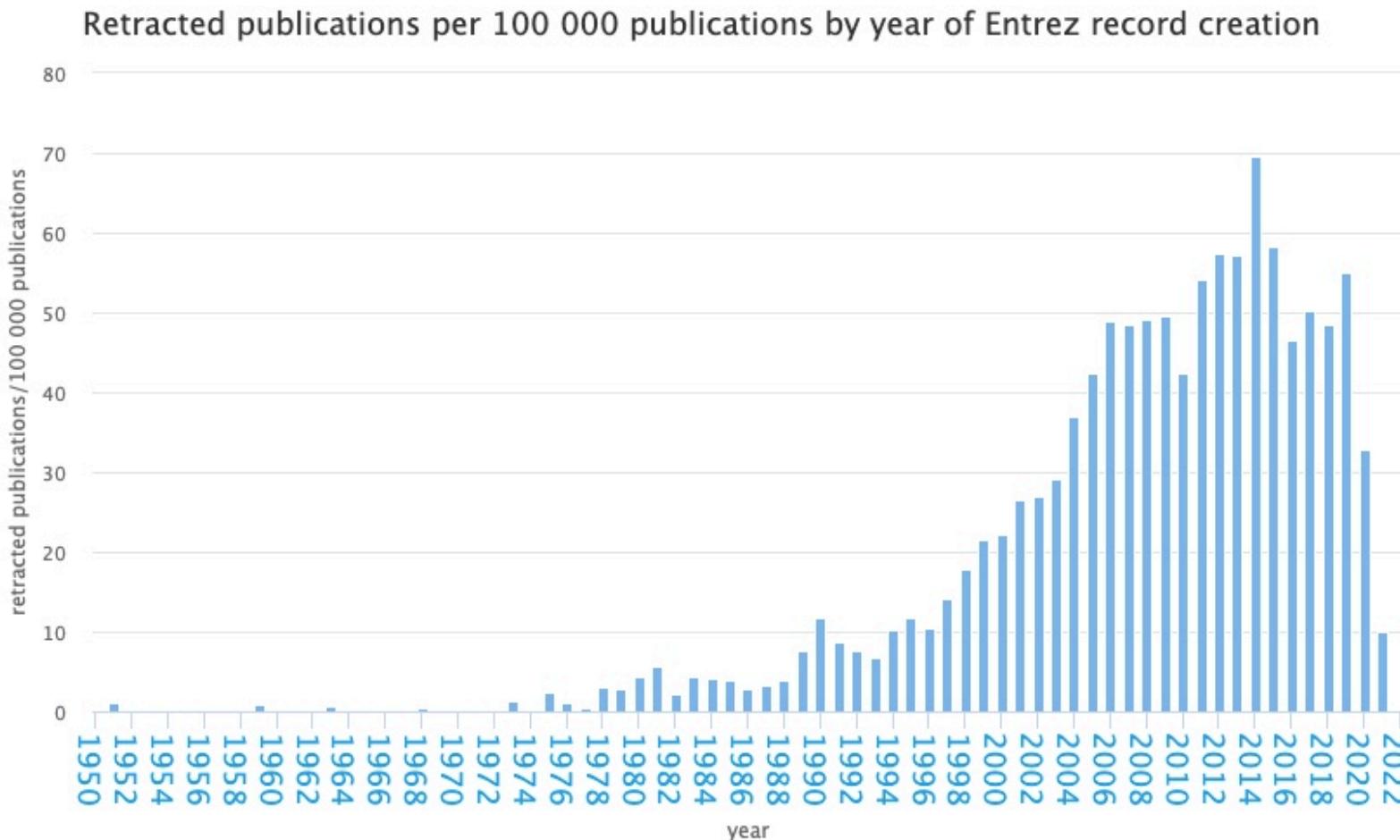
5



CANCER
RESEARCH
UK

CAMBRIDGE
INSTITUTE

Crisis in Reproducible Research



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

The screenshot shows the header of the Nature journal website. The main title 'nature' is in large white serif font, with 'International weekly journal of science' in smaller text below it. A navigation bar includes links for Home, News & Comment, Research, Careers & Jobs, Current Issue, Archive, Audio & Video, and a search icon. Below the navigation is a breadcrumb trail: Archive > Volume 483 > Issue 7391 > Comment > Article. The page content starts with 'NATURE | COMMENT' and a share icon.

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



Need for Good Design



UNIVERSITY OF
CAMBRIDGE

8



CANCER
RESEARCH
UK

CAMBRIDGE
INSTITUTE

Consequences of Poor Experimental Design...

- **Cost** of experimentation. We have a responsibility to CRUK 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.
- **Time and career** of individuals.

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."

"... VERY OFTEN, ... THE MOST ELABORATE STATISTICAL REFINEMENTS POSSIBLE COULD INCREASE THE PRECISION BY ONLY A FEW PERCENT, YET A DIFFERENT DESIGN INVOLVING LITTLE OR NO ADDITIONAL EXPERIMENTAL LABOUR MIGHT INCREASE THE PRECISION TWO-FOLD, OR FIVE-FOLD OR EVEN MORE."

Aspects of Experimental Design

- **EXPERIMENTAL FACTORS**
- **POWER**
 - Sources of Variance
 - Replicates
- **ERRORS**
 - Confounding factors
 - Bias
 - Random errors



UNIVERSITY OF
CAMBRIDGE



CANCER
RESEARCH
UK

CAMBRIDGE
INSTITUTE

Experimental Factors



UNIVERSITY OF
CAMBRIDGE

13



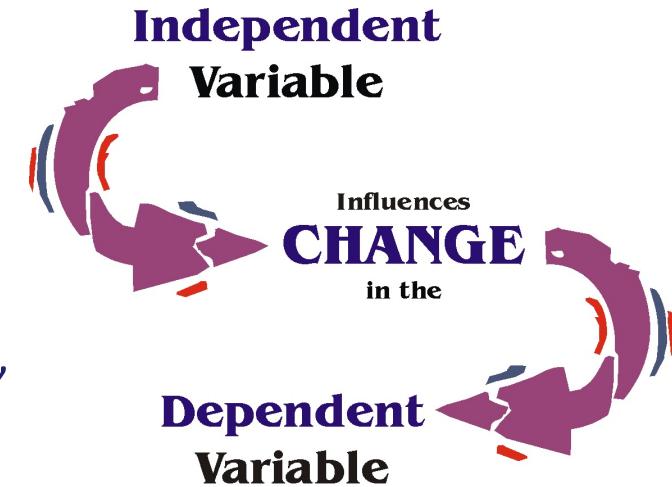
CANCER
RESEARCH
UK

CAMBRIDGE
INSTITUTE

Experimental Factors

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

Capturing Variance



UNIVERSITY OF
CAMBRIDGE

15

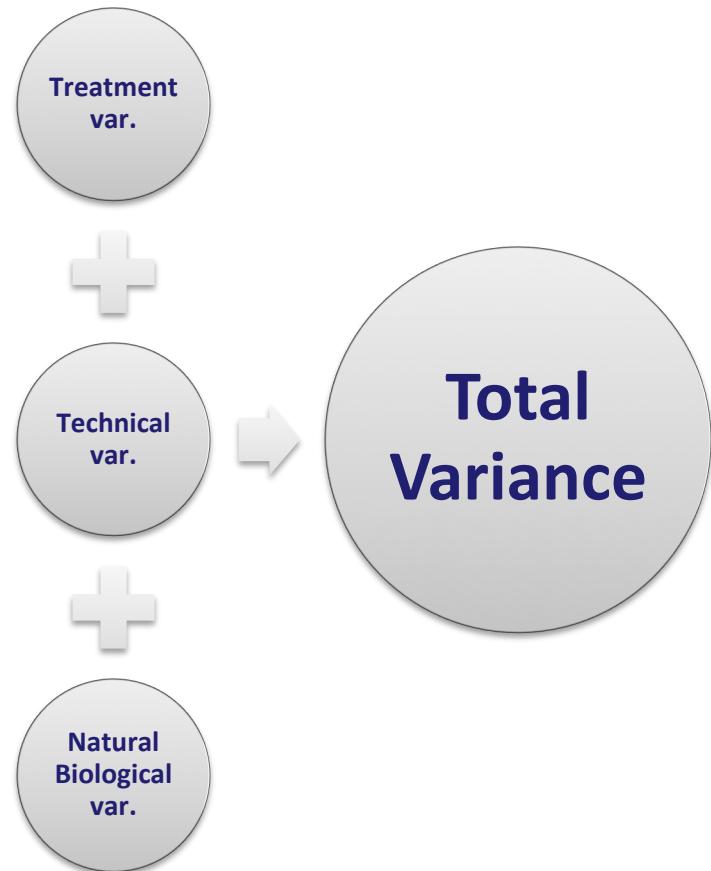


CANCER
RESEARCH
UK

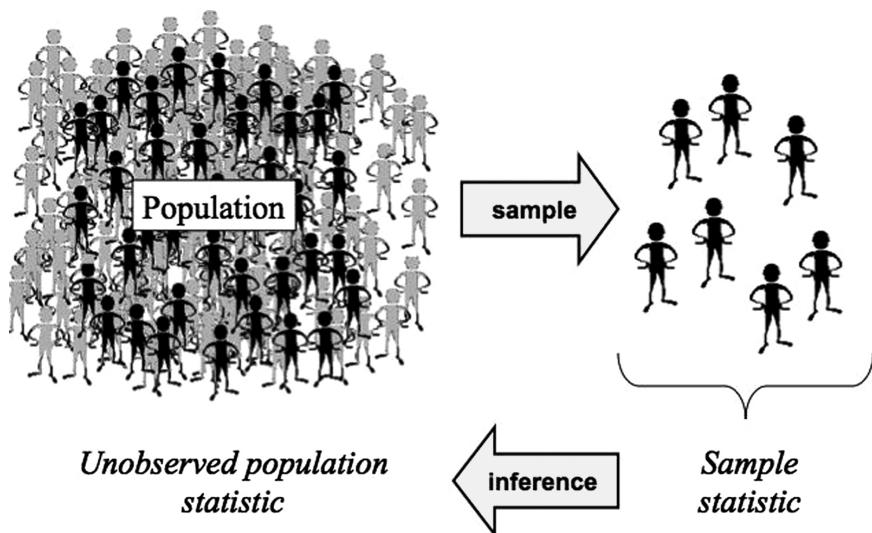
CAMBRIDGE
INSTITUTE

Sources of Variance

- Total Variance
 - Treatment variance
 - Captures the effect of treatment
 - Technical variance
 - Like using different instruments
 - Sample processing in batches
 - Natural Biological variance
 - stochastic in nature



Central Dogma of Statistics



- Two important parameters
 - Mean
 - Variance
- Population mean and variance unknown and are constants
- Estimated using sample
- Estimated mean and variance used for inferring population parameters

P-value gives significance of estimates

Expression Matrix

	Treatment group			Control group		
	T1	T2	T3	C1	C2	C3
Gene 1						
Gene 2						
Gene 3						
Gene n						

EXPRESSION EXPERIMENT ...

- **TWO GROUP COMPARISON**
- *Statistic IS FUNCTION OF EFFECT SIZE AND SE*
- **P VALUE IS INVERSELY RELATED TO STATISTIC (EG. T - STATISTIC)**

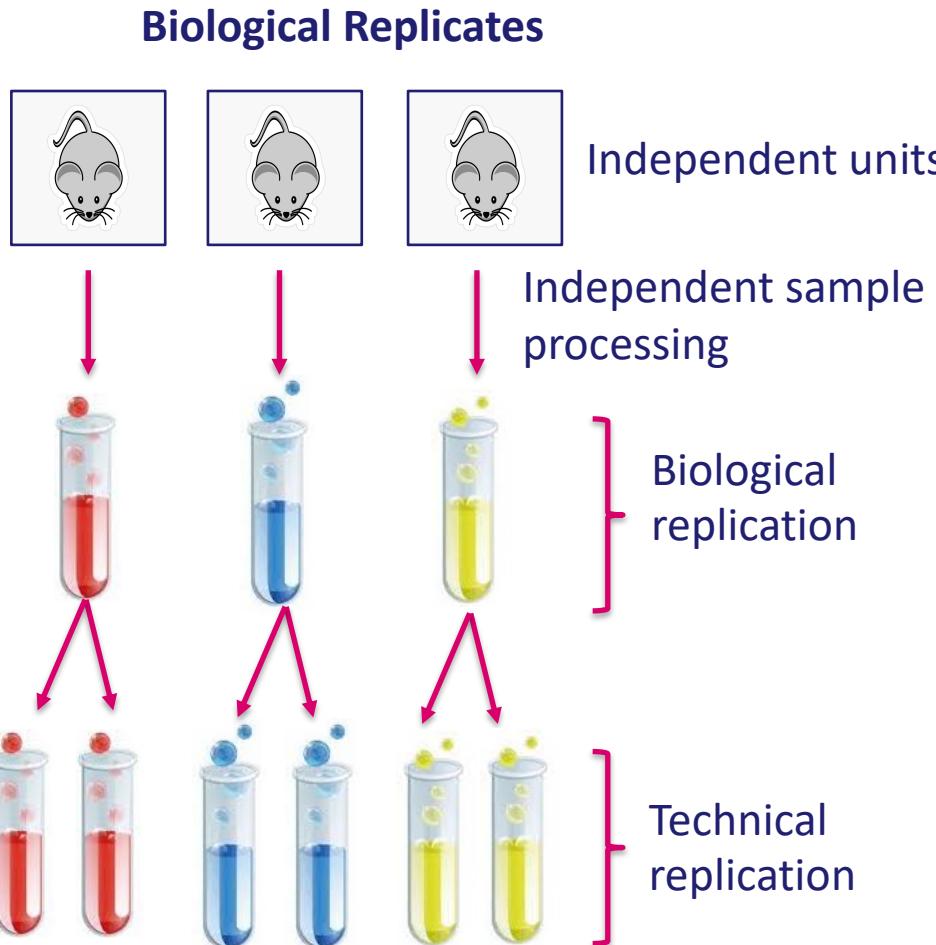
$$Statistic = \frac{\text{Treatment mean} - \text{Control mean} \text{ (Effect size)}}{\text{Standard error (SE)}}$$

$$SE = \text{variance of mean} = \frac{\text{Sample Variance}}{\sqrt{\text{No.of Replicates}}}$$

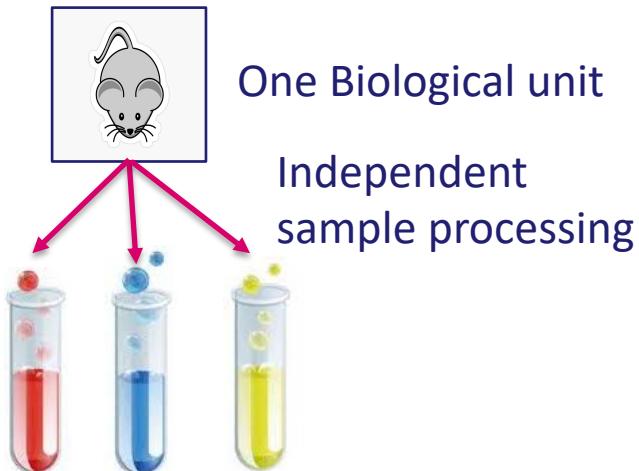
$$p \text{ value} \propto \frac{1}{Statistic}$$



Replicates: Biological /Technical/ Experimental



Experimental replication:
Neither technical nor Biological



Biological replicate : sampling of individuals from a population
in order to make inferences about that population

Technical replicate addresses the measurement error of the assay.



CANCER
RESEARCH
UK

CAMBRIDGE
INSTITUTE

Sample size and experimental power

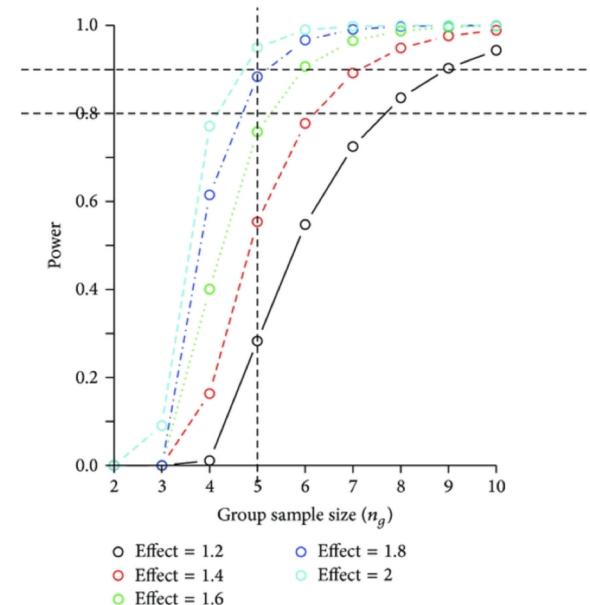
WHY DO YOU NEED REPLICATES?

CALCULATING APPROPRIATE SAMPLE SIZES

- Power calculations
- Planning for precision
- Resource equation

EXPERIMENTAL POWER

- **Power:** Probability of detecting an effect, if there is a true effect present to detect.
- Statistical power increases with ...
 - Higher sample sizes
 - Higher effect size
 - Low variance
 - Higher alpha values



HPB Surg. 2014;2014:310372.



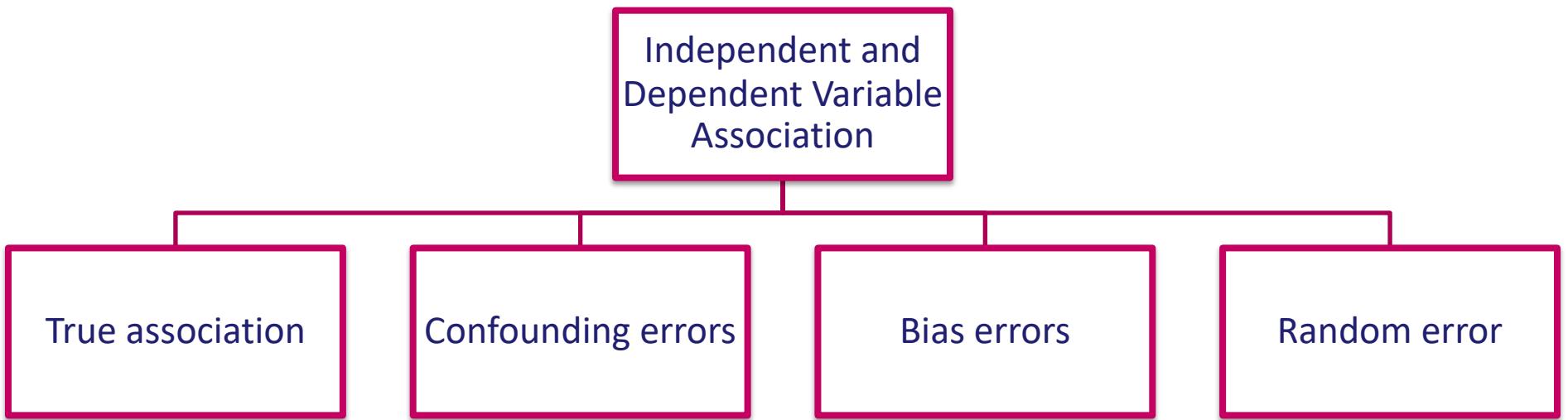
CANCER
RESEARCH
UK

CAMBRIDGE
INSTITUTE

Errors in the studies: Confounding, bias, chance



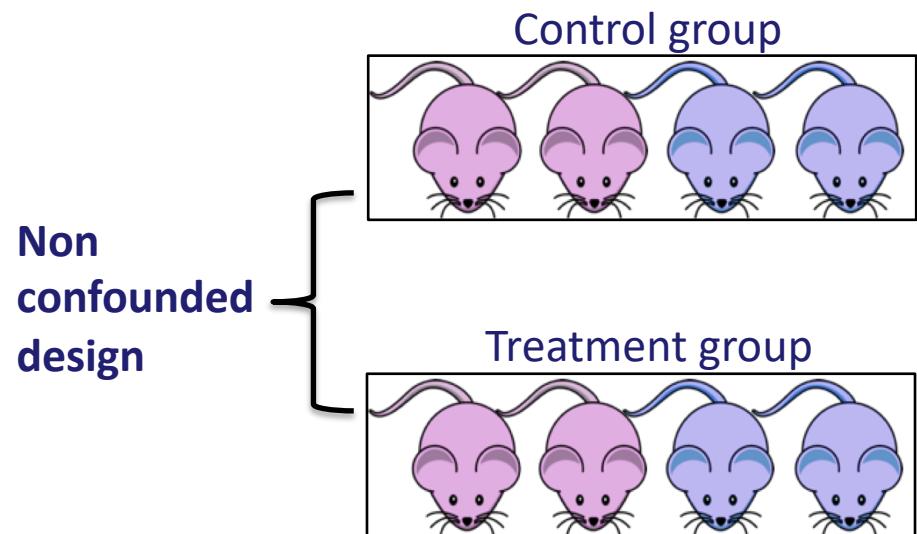
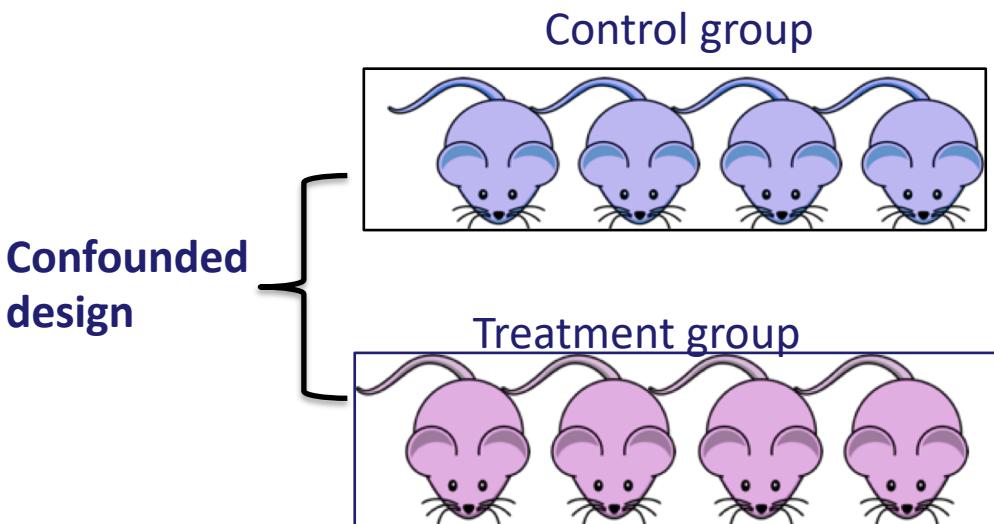
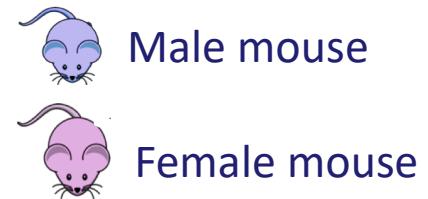
Experimental Errors



Confounding Errors: Thought experiment

Hypothesis: Drug X increases mouse weight

Resources: 8 mice (4 male + 4 female)



$$Statistic = \frac{\text{Treatment mean} - \text{Control mean} \text{ (Effect size)}}{\text{Standard error (SE)}}$$

23



CANCER
RESEARCH
UK

CAMBRIDGE
INSTITUTE

How to deal with confounding factors?

- **KNOWN CONFOUNDING FACTORS**
 - E.g.: Sex, Processing batch, Age, strain, smoking status ...
 - Careful planning of the experiment
 - Take into account during the analysis (E.g.: Multivariate analysis)
 - Randomization
- **UNKNOWN CONFOUNDING FACTORS**
 - Many confounding factors can not be observed
 - A way to deal with unknown confounders is randomization (E.g. randomly allocate samples to groups)

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

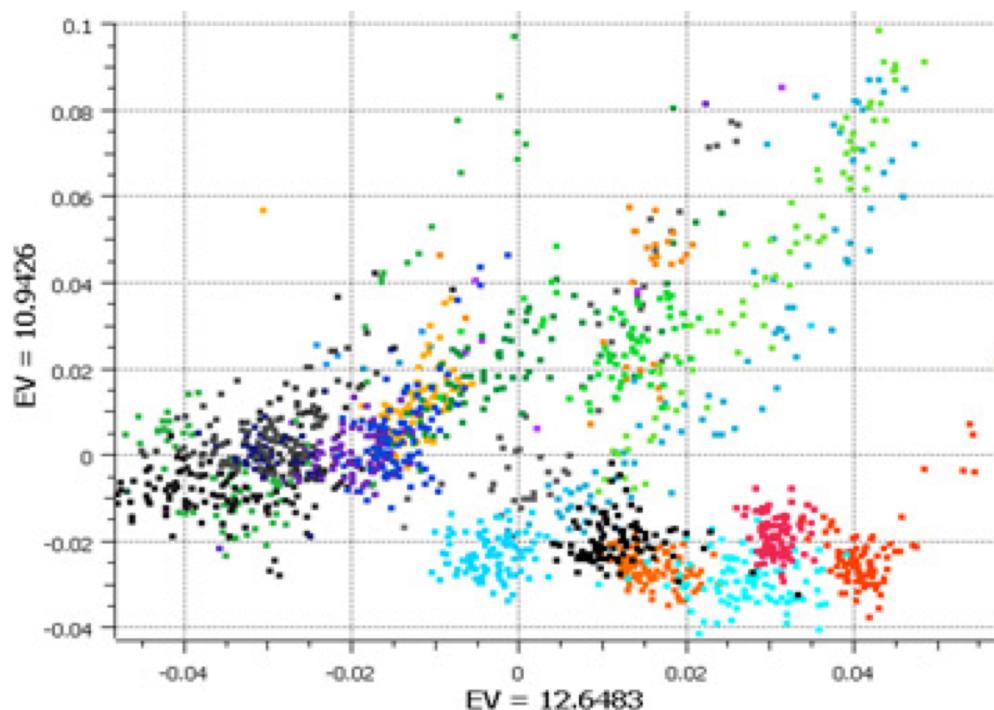
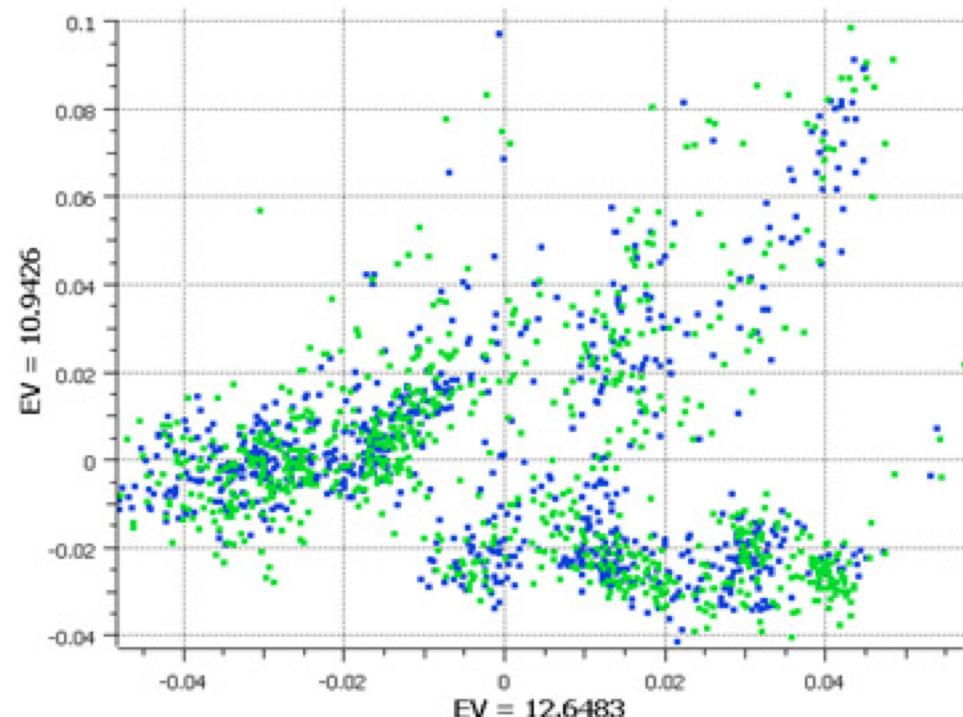


Plate effects by case/control

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



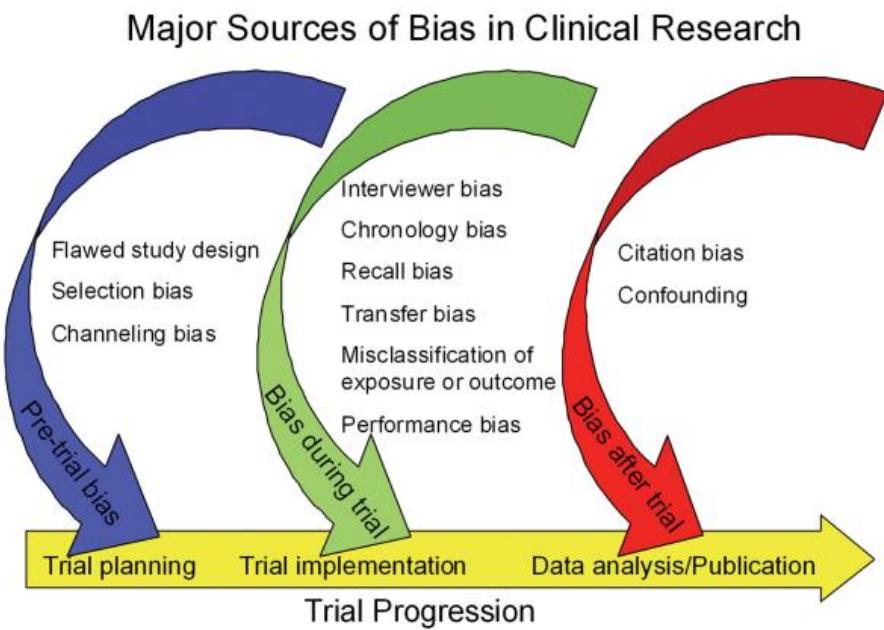
Genetic Signatures of Exceptional Longevity in Humans

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

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

Bias errors



- **BIAS: Systematic Error in design/implementation of study**
 - Selection bias: Error in selecting the samples/enrolling the patients
 - Measurement bias: Error in collecting data
- **How to deal with bias?**
 - Randomization
 - Blinding

Random errors

- **RANDOM ERRORS DEVIATE FROM THE TRUTH WITHOUT SPECIFIC DIRECTION**
 - Can not be avoided but can be reduced by repeated experiments.
 - Can also be reduced by using precise apparatus
 - Can be reduced by increasing the number of replicates

Experimental Controls



UNIVERSITY OF
CAMBRIDGE

29



CANCER
RESEARCH
UK

CAMBRIDGE
INSTITUTE

Experimental Controls

CONTROLLING ERRORS

- Type I: False Positives (reject true H₀)
 - Use Negative controls: A group that should have minimal or no effect
- Type II: False Negative (fail to reject a false H₀)
 - Use Positive controls: A group where known response expected

TECHNICAL CONTROLS

- Detect/correct technical biases
- Normalise measurements (quantification) e.g. spike-ins

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

Design Parameters for Sequencing Experiments



UNIVERSITY OF
CAMBRIDGE

32



CAMBRIDGE
INSTITUTE

Design Issues: Sequencing Experiments

PLATFORMS

LIBRARY PREPS

MULTIPLEXING AND POOLING STRATEGIES

SINGLE-END VS PAIRED END

SEQUENCING DEPTH

- Coverage
- Lanes

VALIDATION

- Knock-downs
- Pull-downs

Experimental Design process at CRUK-CI



CRUK-CI Experimental Design Process

- Students required to take (this) Experimental Design class
- All sequencing and proteomics experiments require experimental design review meeting
 - Simple form with key aspects of experiment
 - Attended by Scientists, Genomics/Proteomics Core, Bioinformatics Core, Statistician
 - Project opened in LIMS afterwards
- Randomisation and Layouts
 - Checkpoint for experiment
 - Project cleared for sample submission
- Keys:
 - Form and meeting not difficult
 - (Currently) not chargeable
 - Scientists agree process improves experiments!

Experimental Design Meetings - Genomics

WEDNESDAY 30 MIN SLOTS (2:00-3:00PM) WITH BIOINFORMATICS GENOMICS/ CORES

REQUIREMENTS:

- Email CRIExperimentalDesign@cruk.cam.ac.uk to request meeting
- Fill in [Experimental Design Form](#) and return at least couple of days before meeting
- **Your attendance**
- Provide **project background** (a few slides from you)

DISCUSSION:

- Planning, time-scale, cost, aims, scope, questions
- Choosing the correct technology
- Effect size & Sample-size calculation?
- Sample collection and processing methods
- Sample information (meta-data) collection
- Randomisation, Blocking and Replication issues
- Technical issues e.g. what sequencing depth?
- Will Bioinformatics Core help with/do analysis?
- Analysis deliverables

Experimental Design Meetings - Proteomics

- FRIDAY 45 MIN SLOTS (10-11:30AM)) WITH BIOINFORMATICS PROTEOMICS CORES

REQUIREMENTS:

- For proteomics EDMs: Email ProteomicsProjectDesign@cruk.cam.ac.uk to request meeting
- Provide **project background** (a few slides from you)

DISCUSSION:

- Planning, time-scale, cost, aims, scope, questions
- Choosing the correct technology
- Sample collection and processing methods
- Sample information (meta-data) collection
- Randomisation, Blocking and Replication issues
- Will Bioinformatics Core help with/do analysis?
- Analysis deliverables

Experimental Design Guide

- [HTTPS://SHAREPOINT.CRI.CAMRES.ORG/SITES/BIOINFORMATICS/PUBLIC/INTRODUCTIONTOEXPERIMENTALDESIGN/EXPERIMENTALDESIGNMANUAL.PDF](https://sharepoint.cri.camres.org/sites/bioinformatics/public/introductiontoexperimentaldesign/experimentaldesignmanual.pdf)
- [TINYURL.COM/CRUK-EDESIGN](http://tinyurl.com/cruk-edesign)

Practicals

1. **Genomic/Clinical:** Identification of prognostic biomarkers in human prostate cancer patients ([Stéphane](#))
2. **RNA-seq/Animal:** Effects of mutant vs wildtype HHEX in liver and brain development ([Abbi/Ash](#))
3. **Quantitative Proteomics/Cultured Cells:** AR interactome differences between drug responsive/resistant conditions ([Kamal/Mark](#))
4. **ChIP-seq/Animal:** Evolution of transcription factor binding in mouse strains ([Chandu/Ash](#))