INTRODUCTION TO EXPERIMENTAL DESIGN AT CRUK-CI

CHANDRA CHILAMAKURI ABBI EDWARDS
KAMAL KISHORE ASH SAWLE
RORY STARK

31 OCTOBER 2022

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

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

TO GET EVIDENCE (HOPEFULLY) SUPPORTING A HYPOTHESIS

Reproducible Research

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

nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | F

Archive > Volume 483 > Issue 7391 > Comment > Article

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

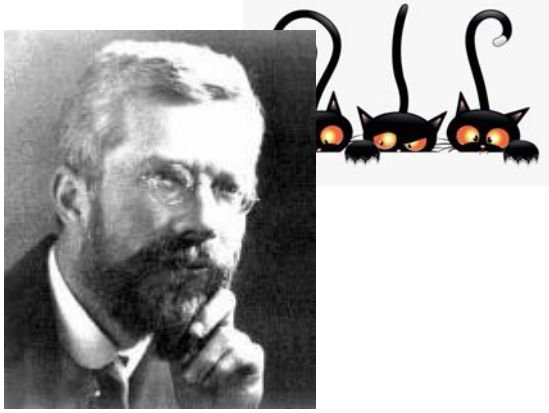
Need for Good Design

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.

Statistical Aspects of Experimental Design

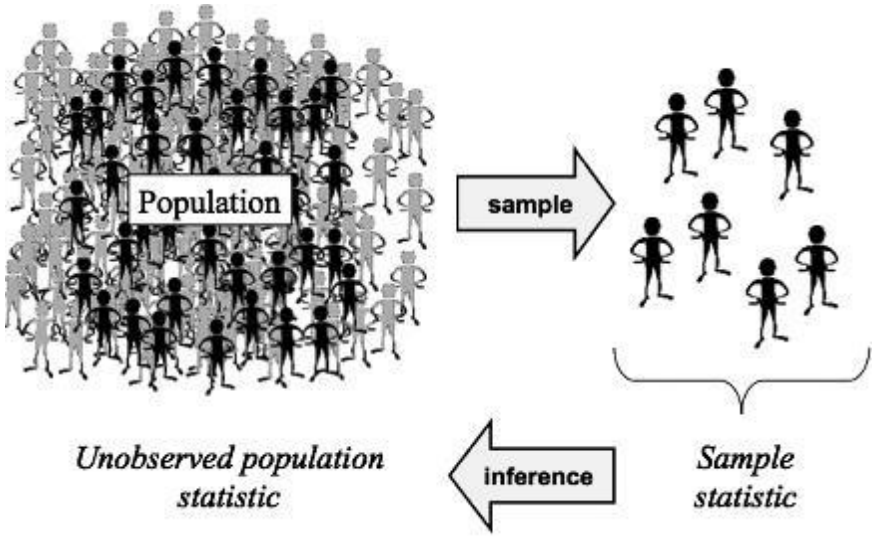
Ronald A. Fisher(1890-1962)



"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."

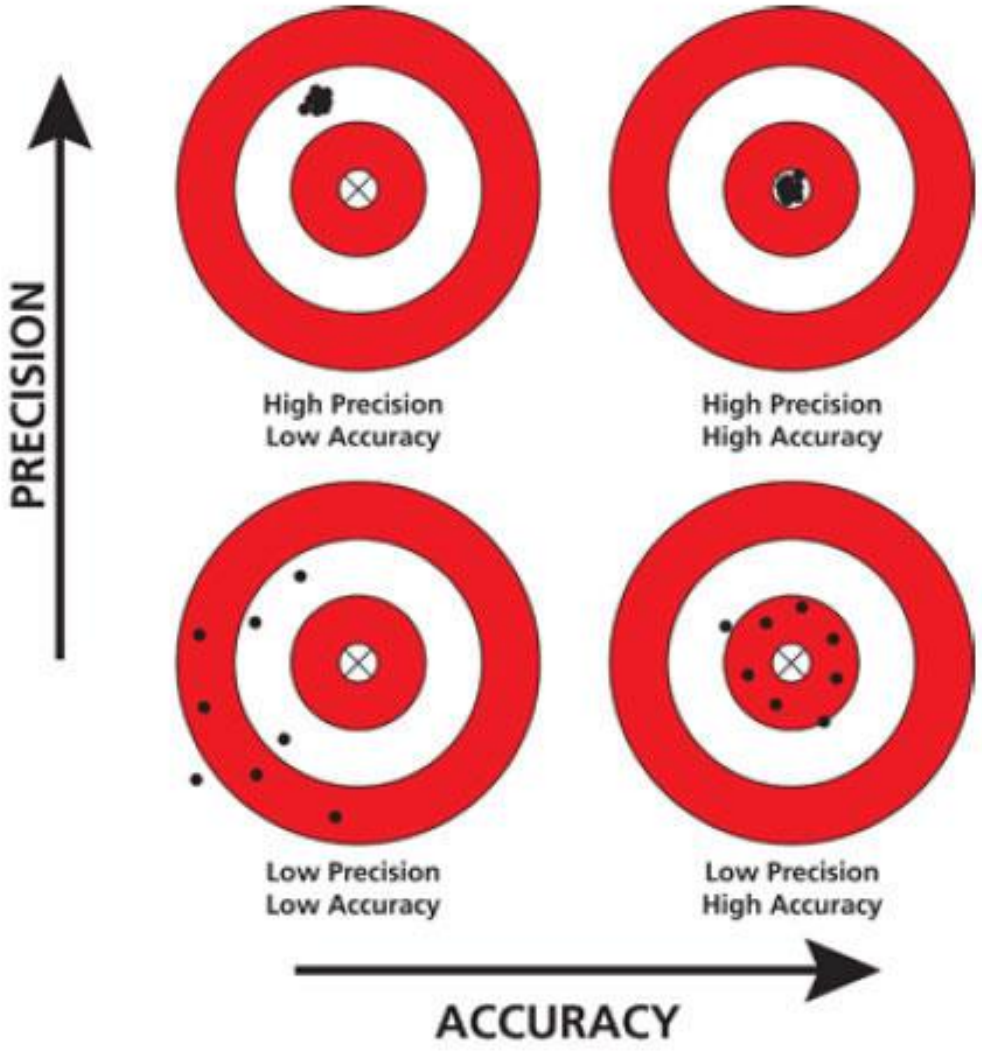
Central Dogma of Parametric Statistics



- Data follow a distribution
- Important parameters
  - Mean
  - Variance
- Population parameters unknown
- Estimated using sampling
- Parameters estimated from samples used for inferring population parameters

CANCER RESEARCH UK | CAMBRIDGE INSTITUTE

Precision and Accuracy



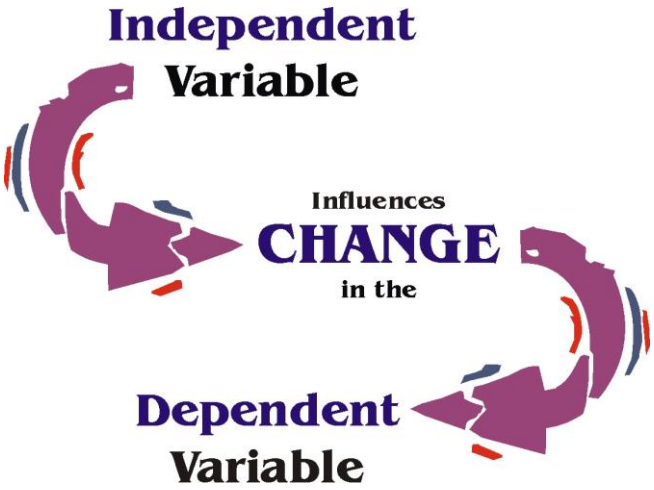
CANCER RESEARCH UK | CAMBRIDGE INSTITUTE

Aspects of Experimental Design

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

Experimental Factors

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:

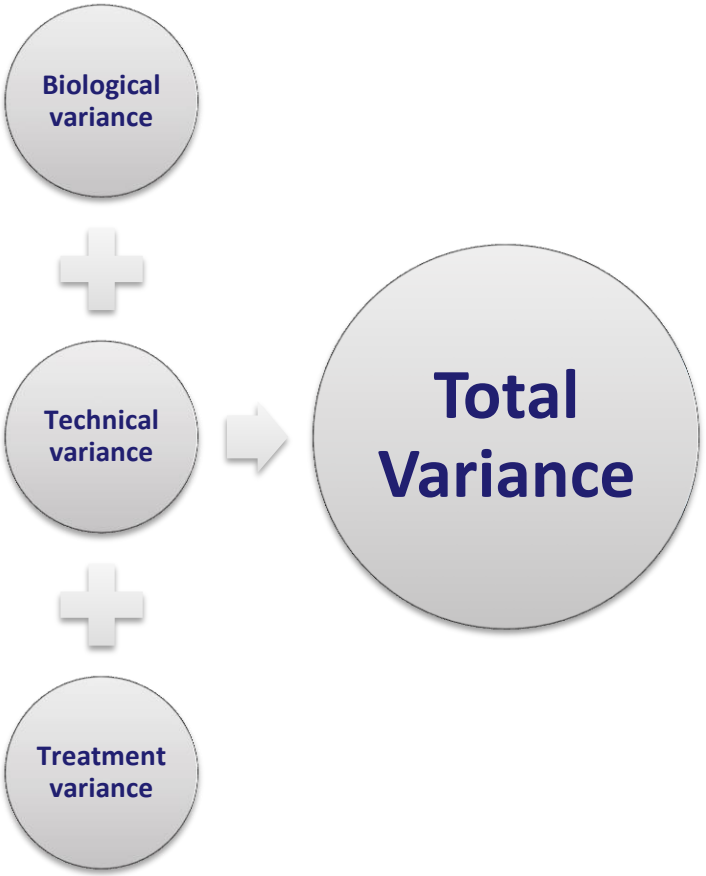
- Categorical (**nominal**) , e.g. gender
- Categorical with ordering (**ordinal**), e.g. tumour grade
- **Discrete**, e.g. shoe size, number of cells
- **Continuous**, e.g. body weight in kg, height in cm

CANCER RESEARCH UK | CAMBRIDGE INSTITUTE

Power: Charactertising Variance using Replication

Sources of Variance

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

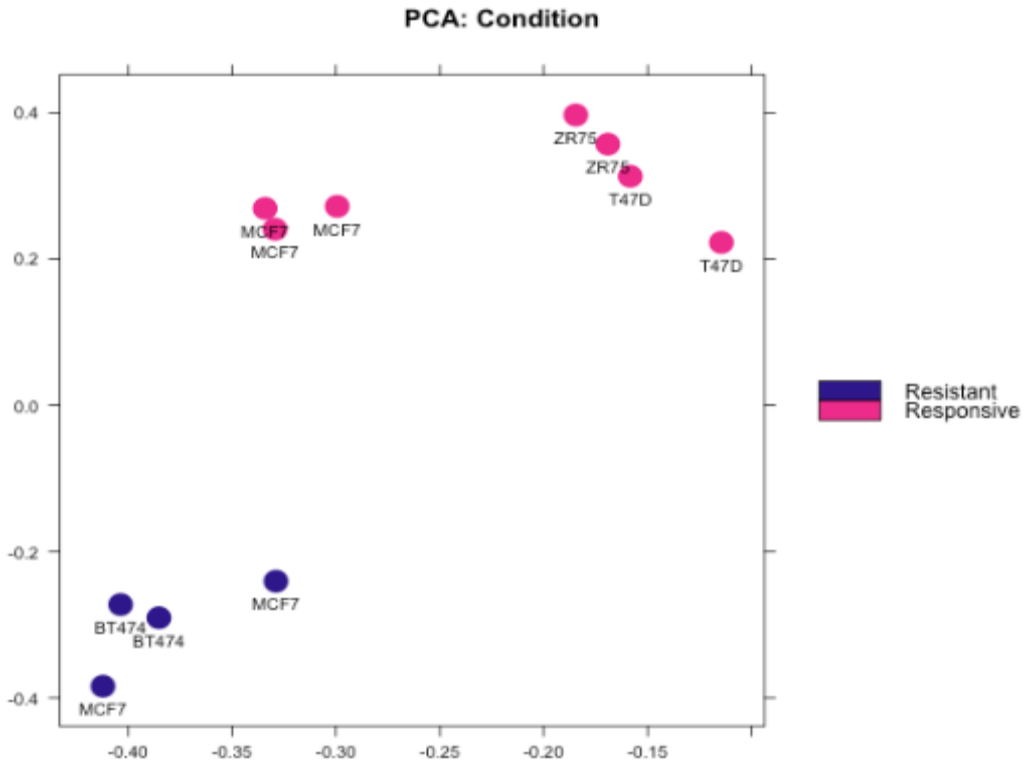


CANCER RESEARCH UK | CAMBRIDGE INSTITUTE

Types of Replication

BIOLOGICAL REPLICATION:

- *In vivo:*
  - Patients
  - Mice
- *In vitro:*
  - Different cell lines
  - Re-growing cells (passages)



TECHNICAL REPLICATION:

- Experimental protocol
- Measurement platform (i.e. sequencer)

CANCER RESEARCH UK | CAMBRIDGE INSTITUTE

Sample size and experimental power

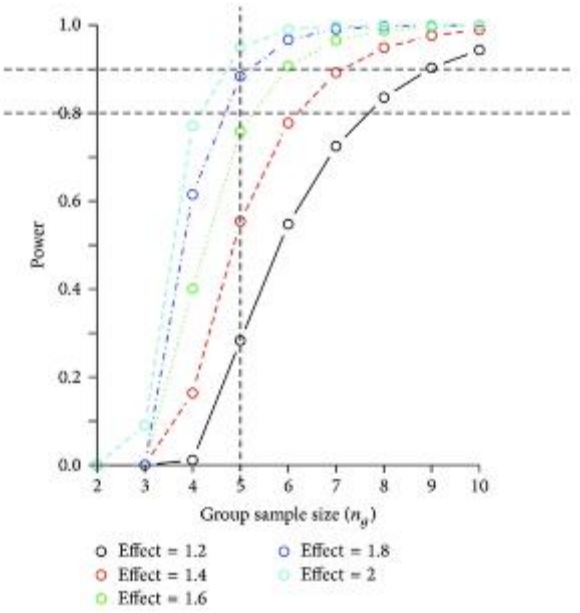
WHY DO YOU NEED REPLICATES?

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

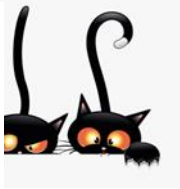


HPB Surg. 2014;2014:310372.

CANCER RESEARCH UK | CAMBRIDGE INSTITUTE

Bias: Counfounders, Randomisation, and Controls

AN EXPRESSION OF CONCERN HAS BEEN PUBLISHED ON THIS PAPER; SEE LAST PAGE



Sciencexpress Report

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]*

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

CANCER RESEARCH UK | CAMBRIDGE INSTITUTE

http://www.the-scientist.com/blog/display/57558/

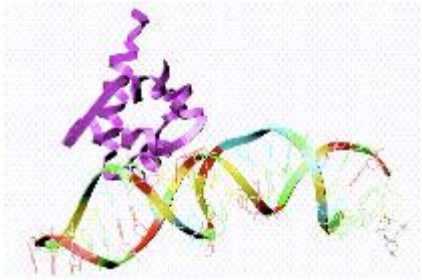
Confounding Factors

WHAT COULD ACCOUNT FOR THESE FINDINGS?

- Democrats were less satisfied with their sex lives than Republicans.
  (ABC poll report)
- Overweight people have longer life expectancy than thin people
  (US Centre for Disease Control)

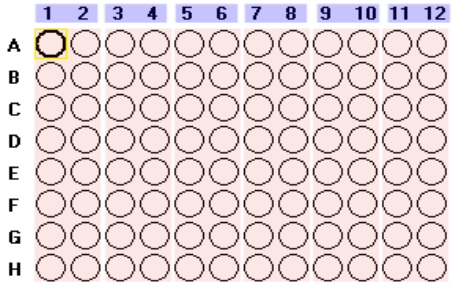
CANCER RESEARCH UK | CAMBRIDGE INSTITUTE

Technical Confounding Factors: Batch Effects



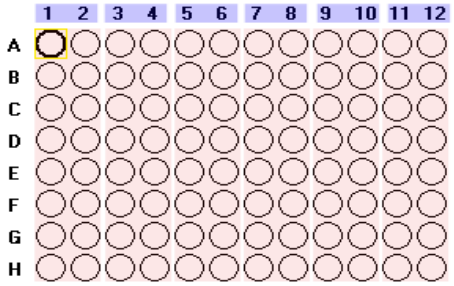
RNA Extraction

Day1, Plate 1



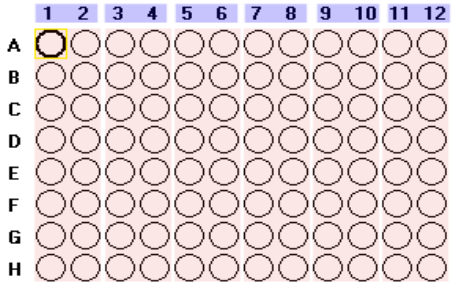
Control

Day2, Plate 2



Treatment 1

Day3, Plate 3



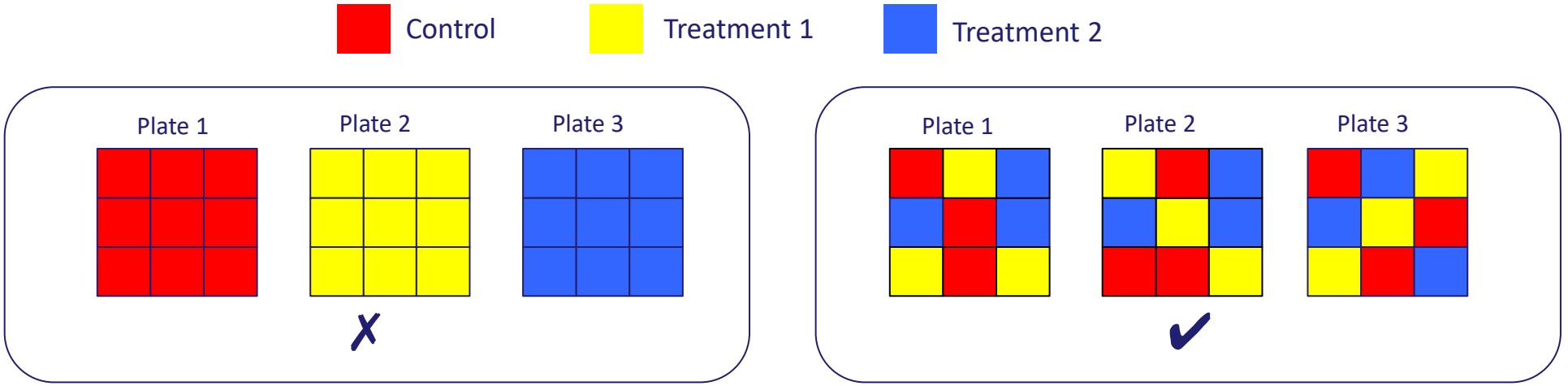
Treatment 2

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

CANCER RESEARCH UK | CAMBRIDGE INSTITUTE

Randomised Block Design

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



RBD across plates so that each plate contains spatially randomised **equal proportions** of:

- Control
- Treatment 1
- Treatment 2

controlling plate effects.

CANCER RESEARCH UK | CAMBRIDGE INSTITUTE

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

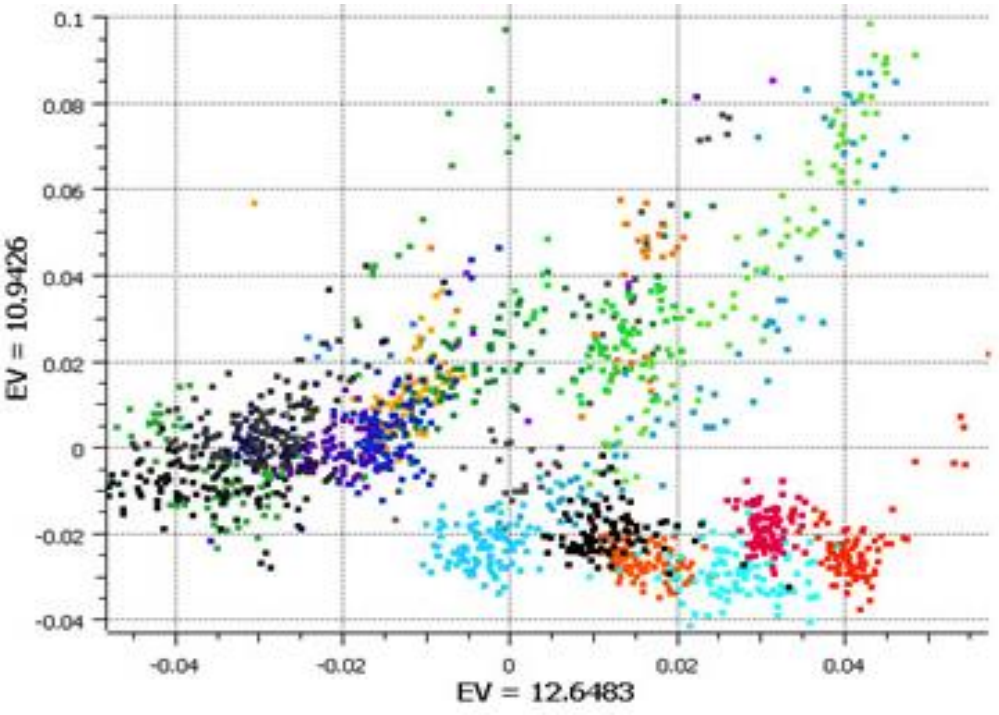
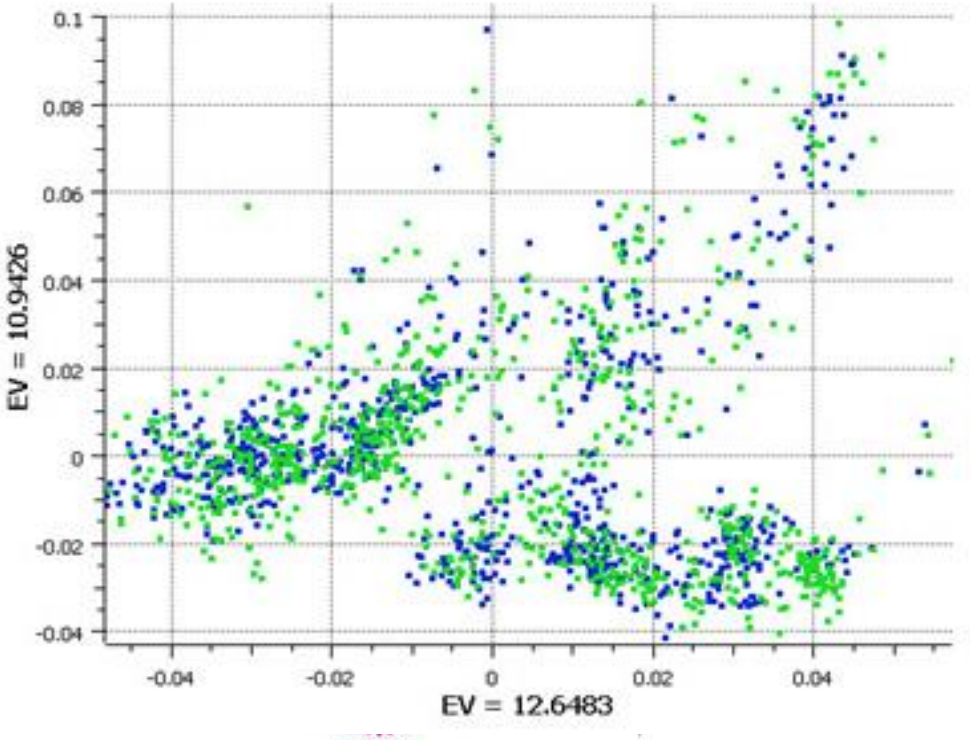


Plate effects by *case/control*

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



http://blog.goldenhelix.com/?p=322

Experimental Controls

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

CANCER RESEARCH UK | CAMBRIDGE INSTITUTE

Design Parameters for Sequencing Experiments

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

CANCER RESEARCH UK | CAMBRIDGE INSTITUTE

Experimental Design process at CRUK-CI

CRUK-CI Experimental Design Process

- Students required to take (this) Experimental Design class
- Experimental design review meetings for sequencing and proteomics experiments require
  - 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
  - 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:

- 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/INRODUCTIONTOEXPERIMENTALDESIGN/EXPERIMENTALDESIGNMANUAL.PDF
- **TINYURL.COM/CRUK-EDESIGN**

Practicals

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