INTRODUCTION TO EXPERIMENTAL DESIGN AT CRUK-CI

CHANDRA CHILAMAKURI
KAMAL KISHORE
JING SU
RORY STARK
29 OCTOBER 2018

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

PRACTICALS
Why Perform Experiments?
Reproducible Research
Crisis in Reproducible Research

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

http://neilfws.github.io/PubMed/pmretract/pmretract.html
47 of 53 high-profile cancer studies were not reproducible!

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.
A Well-Designed Experiment:

Should have

• CLEAR OBJECTIVES
• FOCUS AND SIMPLICITY
• SUFFICIENT POWER
• RANDOMISED COMPARISONS

And be

• PRECISE
• UNBIASED
• AMENABLE TO STATISTICAL ANALYSIS
• REPRODUCIBLE
“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
- **BIAS**
  - Confounding factors
  - Randomisation wherever a decision is to be made
    - Controls for both measured and unmeasured factors
  - Controls
Experimental Factors
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.

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

INDEPENDENT AND DEPENDENT VARIABLES

- Independent variable (IV): what you change
- Dependent variable (DV): what changes due to IV
- “If (independent variable), then (dependent variable)”
Capturing Variance
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
REPLICATION REQUIRED TO CAPTURE VARIANCE
Sample size and experimental power

WHY DO YOU NEED REPLICATES?

CALCULATING APPROPRIATE SAMPLE SIZES

– Power calculations
– Planning for precision
– Resource equation

EXPERIMENTAL POWER

• **Power**: the *probability* of detecting an *effect* of a specified size if present.
  – Identify and control the *sources of variability*
  – Power calculations estimate sample size required to detect an effect *if degree of variability is known*
  – Using *appropriate numbers* of samples (sample size/replicates)
  – If adding samples increases variability, that alone won’t add power!
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)
Precision, Accuracy, Confounders, and Bias
Precision and Accuracy
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.

HYPOTHETICAL EXAMPLE WOULD BE A STUDY OF COFFEE DRINKING AND LUNG CANCER.

False association

Cause: drinking coffee (independent variable)

Effect/Outcome: lung cancer (dependent variable)

Other factor: smoking (confounding variable)
Confounding Factors

OTHER EXAMPLES:

– Democrats were less satisfied with their sex lives than Republicans. (ABC poll report).

– Overweight (not obese) people have longer life expectancy than thin people (US Centre for Disease Control).
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

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

CONSIDER ALTERNATIVE EXPLANATIONS

CONTROL TECHNICAL EFFECTS:

- Randomisation
  - Statistical analyses assume randomised comparisons
  - May not see issues caused by non-randomised comparisons
  - Make every decision random not arbitrary
  - Caveat: over-randomization can increase error

- Blinding
  - Especially important where subjective measurements are taken
  - Potentially multiple degrees of blinding (eg. double-blinding)
Technical Confounding Factors: Batch Effects

RNA Extraction

Day1, Plate 1

Day2, Plate 2

Day3, Plate 3

Control

Treatment 1

Treatment 2

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

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

**Plate effects by plate**

Left PCA plot shows 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).
Experimental Controls
Experimental Controls

CONTROLLING ERRORS

– Type I: FP
  • Negative controls: should have minimal or no effect

– Type II: FN
  • Positive controls: known effect

TECHNICAL CONTROLS

– Detect/correct technical biases
– Normalise measurements (quantification)
Examples of Experimental Controls

- WILD-TYPE ORGANISM (KNOCKOUTS)
- INACTIVE siRNA (SILENCING)
- VEHICLE (TREATMENTS)
- INPUT: FRAGMENTED CHROMATIN (ChIP)
- SPIKE-INS (QUANTIFICATION/NORMALISATION)
- “GOLD STANDARD” DATAPORTS
- MULTI-LEVEL CONTROLS
  - e.g. contrast Vehicle/Input vs. Treatment/Input
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
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 onerous
  - (Currently) not chargeable
  - Scientists agree process improves experiments!
Experimental Design Meetings - Genomics

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

REQUIREMENTS:
- Email CRIExperimentalDesign@cruk.cam.ac.uk to request meeting
- Fill in Experimental Design Form and return 1 week prior to 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

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

REQUIREMENTS:
- Email ProteomicsProjectDesign@cruk.cam.ac.uk to request meeting
- Fill in ProteomicsMetadataTemplate.xls Your attendance
- 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

• TINYURL.COM/CRUK-EDESIGN
Practicals

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

37