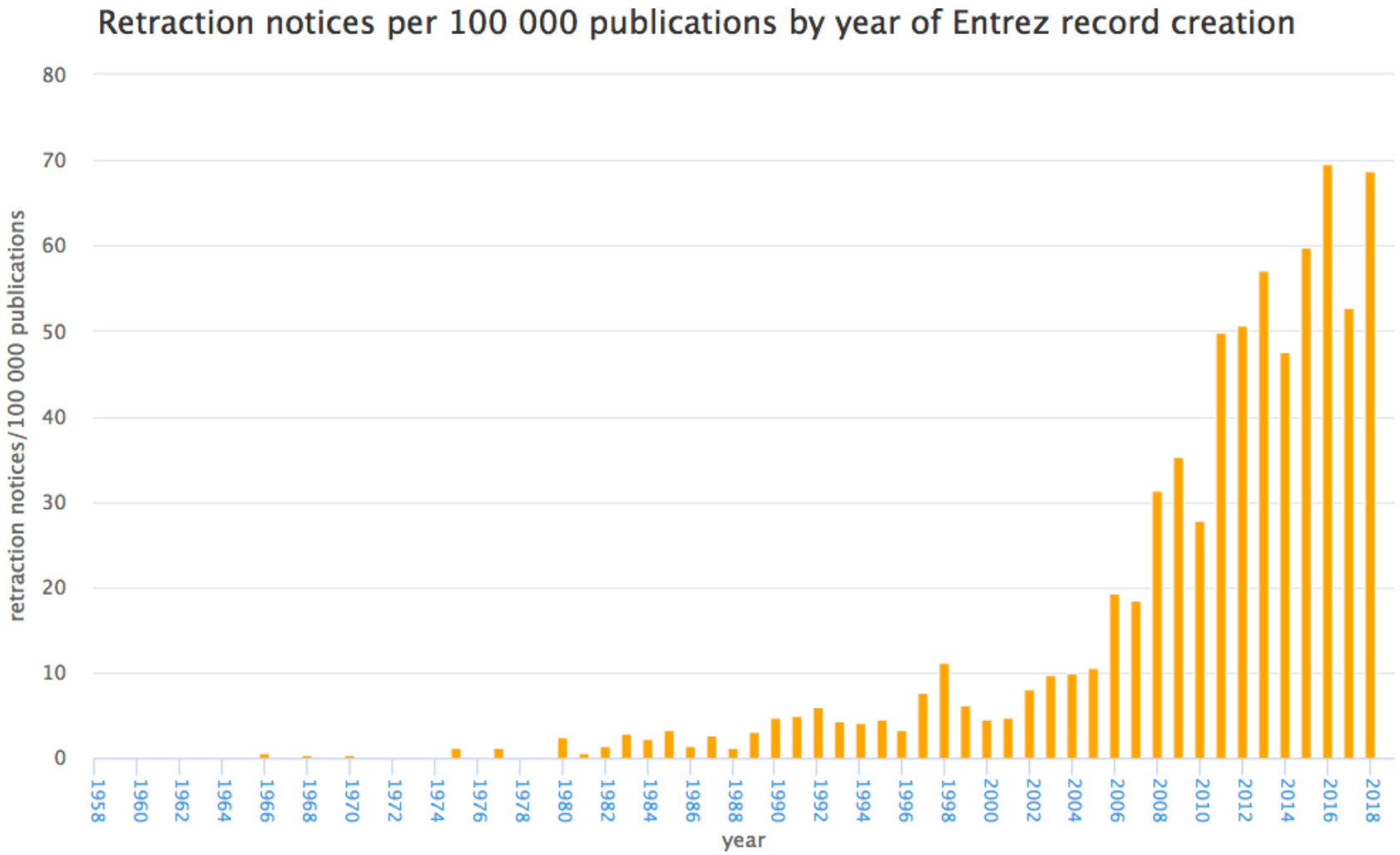
INTRODUCTION TO EXPERIMENTAL DESIGN

From: http://www.hawaii.edu/fishlab/NearsideFrame.htm

Slides adapted from Experimental Design Course, CRUK

Crisis in Reproducible Research



http://neilfws.github.io/PubMed/pmretract/pmretract.html

Consequences of Poor Experimental Design…

- **Cost** of experimentation.
- **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

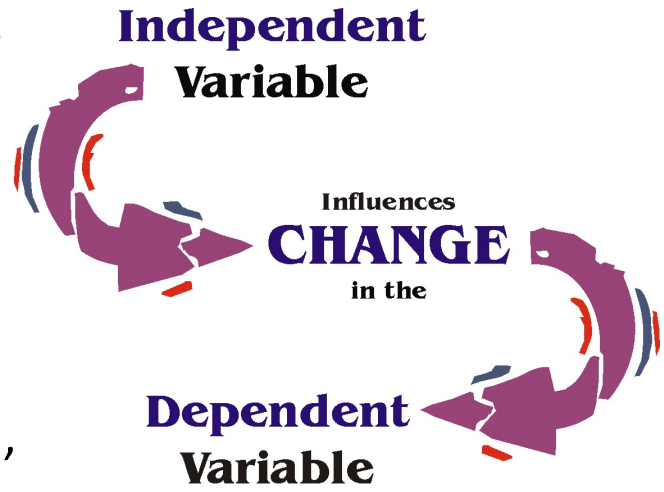
Ronald A. Fisher(1890-1962)

"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."

"… 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."

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)'



Sources of Variation

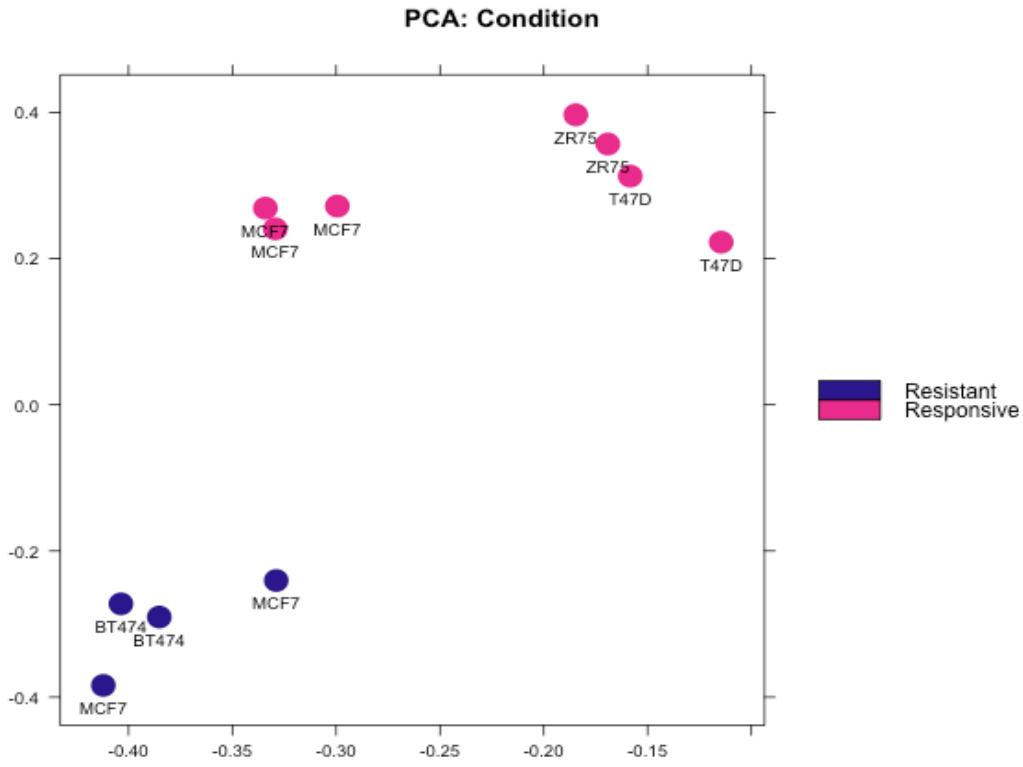
- Biological “noise”
  - Biological processes are inherently stochastic
  - Single cells, cell populations, individuals, organs, species….
  - Timepoints, cell cycle, synchronized vs. unsynchronized
- Technical noise
  - Reagents, antibodies, temperatures, pollution
  - Platforms, runs, operators

- Consider in advance and control
- *Replication required to capture variance*

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)

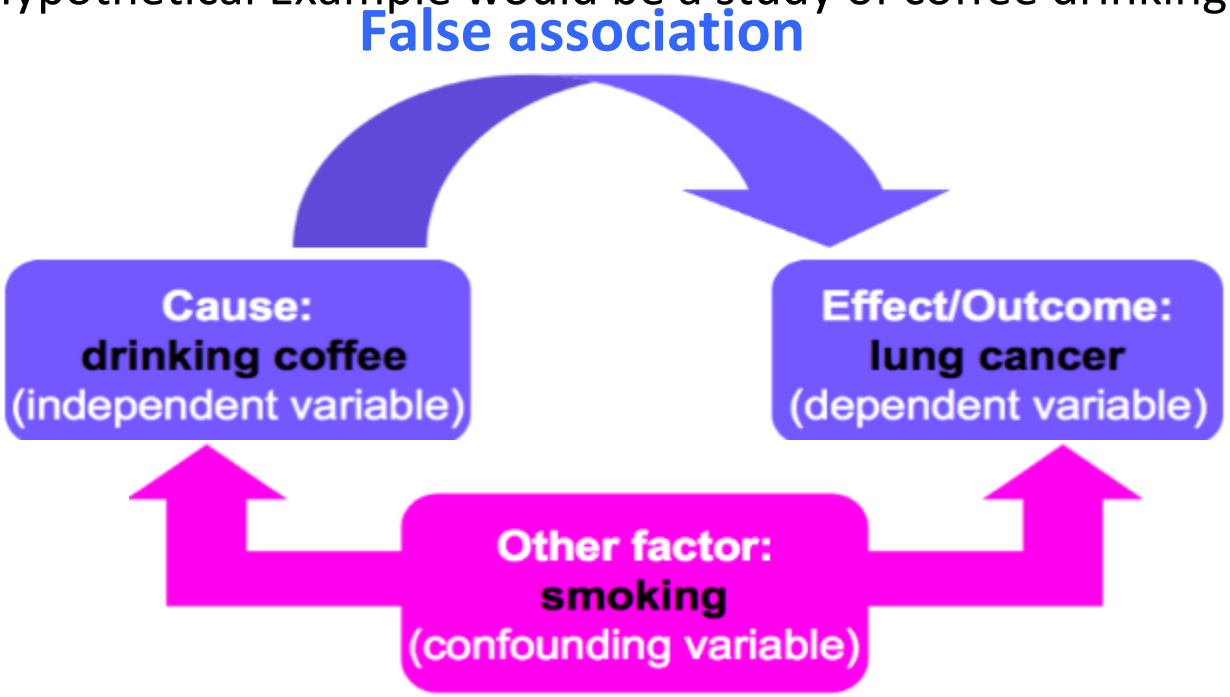


Confounding Factors

- 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.



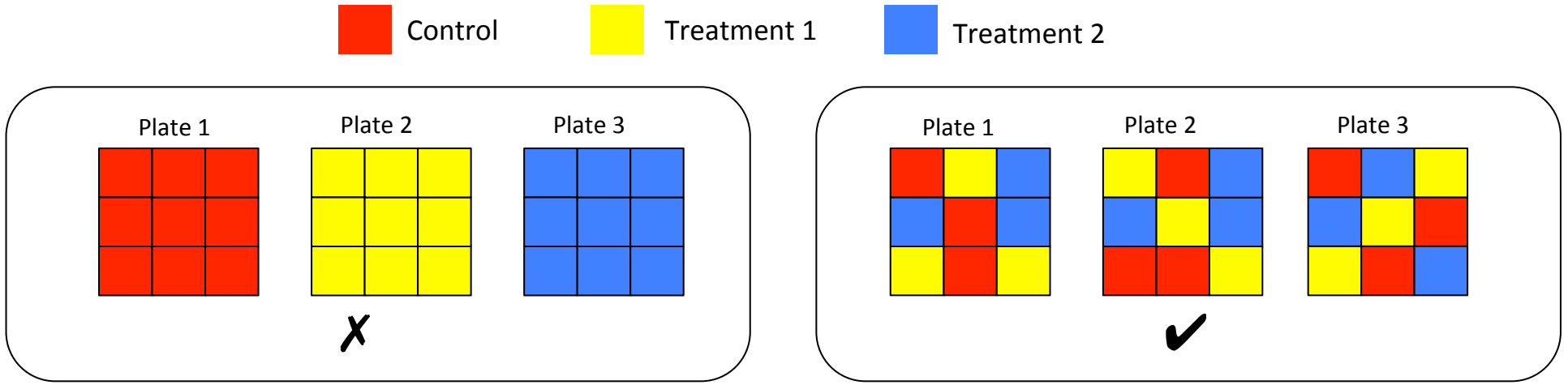
Solutions

- 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)

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

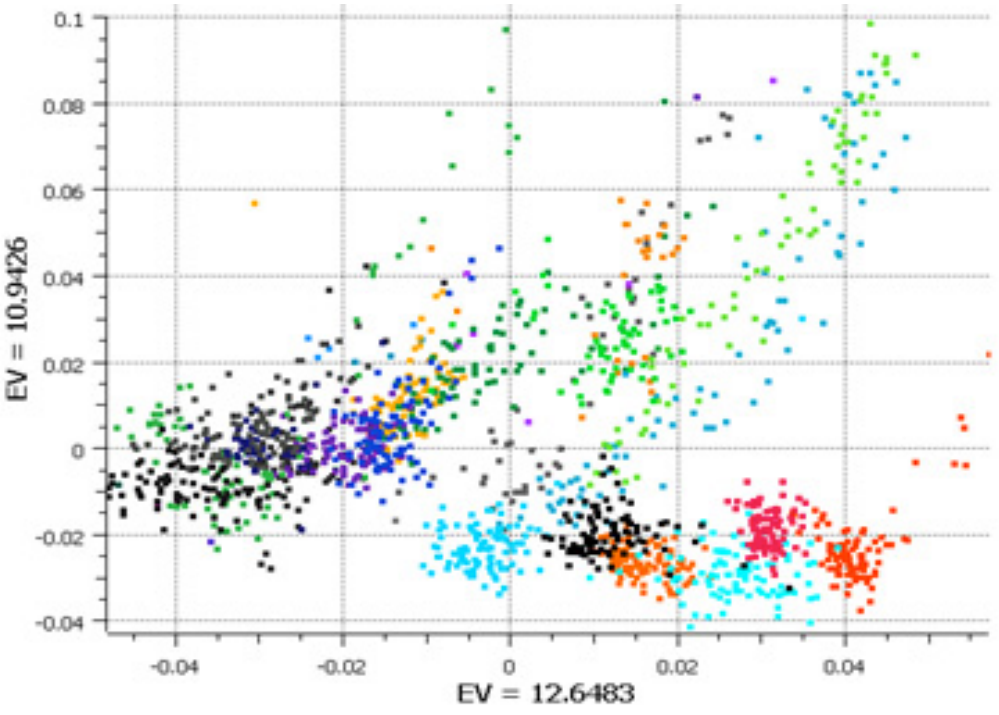
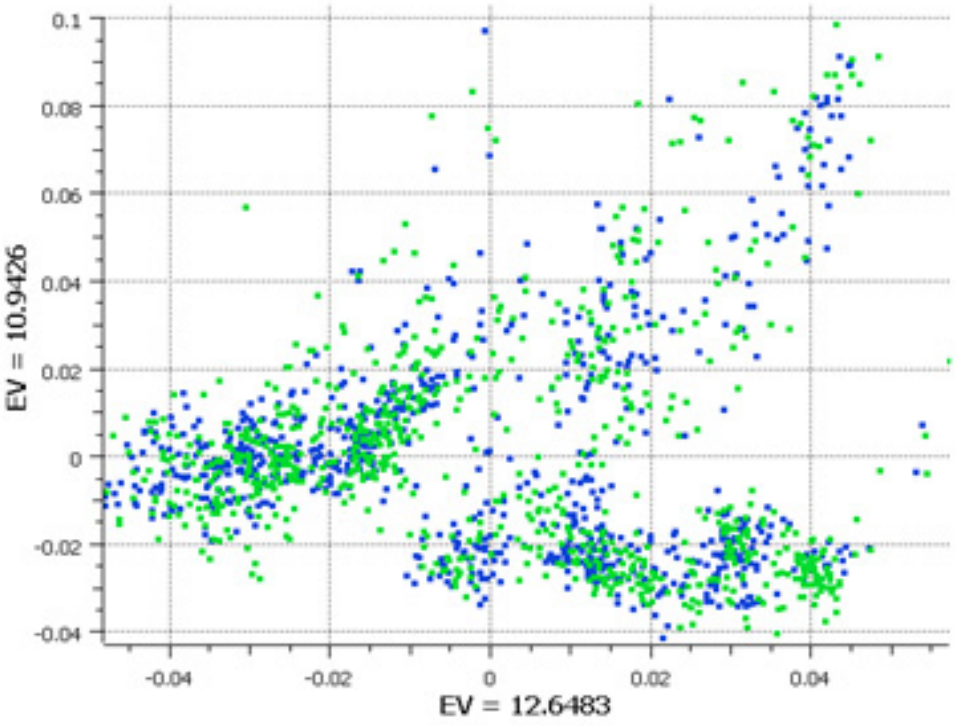


Plate effects by ***case/control***

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



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

Experimental Controls

- 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)
- Spike-ins (quantification/normalisation)
- “Gold standard” datapoints
- Multi-level controls
  - e.g. contrast Vehicle/Input vs. Treatment/Input

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

Practicals

RNA-seq: Effects of mutant vs wildtype HHEX in liver and brain development

People will be divided into groups and will be allocated to breakout rooms.