Foundations in Ethical Research
UCSF GRAD214 Winter/Spring 2024

Tori Sharma, PhD

Housekeeping

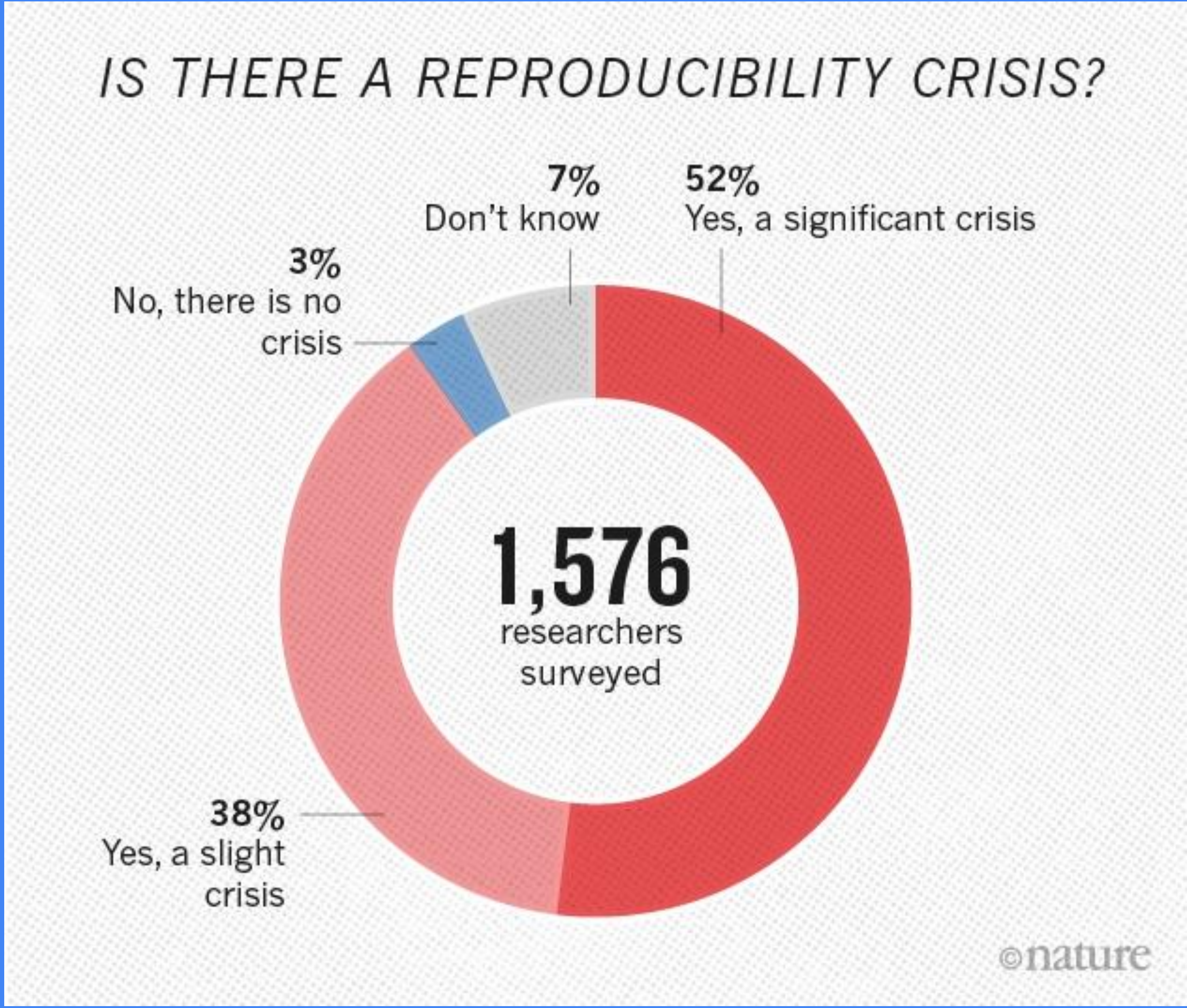
- PollEverywhere
- IMPORTANT- your screen name **MUST** match your registered name for the course or you will not be given credit for attendance or participation

How do you feel about graduate school so far?

Rigor and Reproducibility

“Irakli Loladze, a mathematical biologist at Bryan College of Health Sciences in Lincoln, Nebraska, estimates that ***efforts to ensure reproducibility can increase the time spent on a project by 30%***, even for his theoretical work. He checks that all steps from raw data to the final figure can be retraced. But those tasks quickly become just part of the job. ‘**Reproducibility is like brushing your teeth,” he says. “It is good for you, but it takes time and effort. Once you learn it, it becomes a habit.’”**

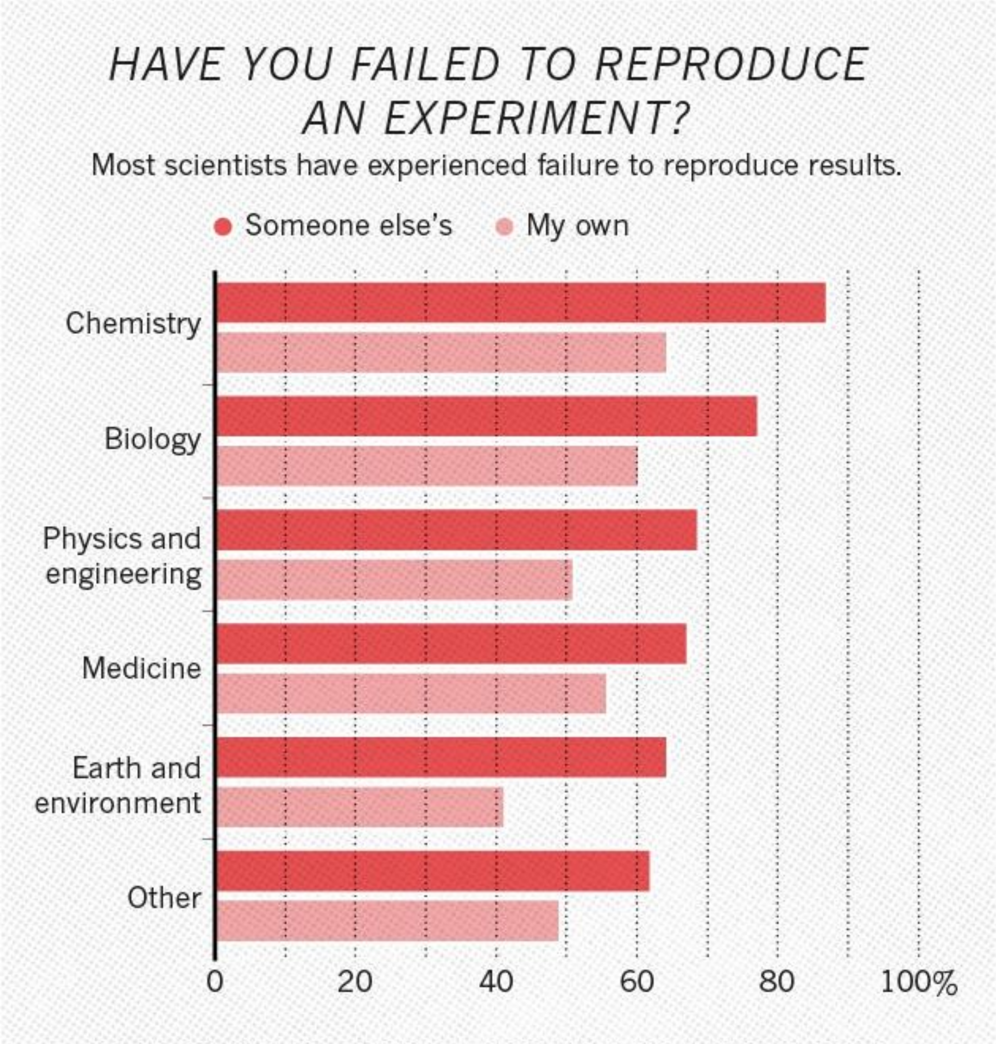
http://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970



Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* 533, 452–454 (2016). https://doi.org/10.1038/533452a

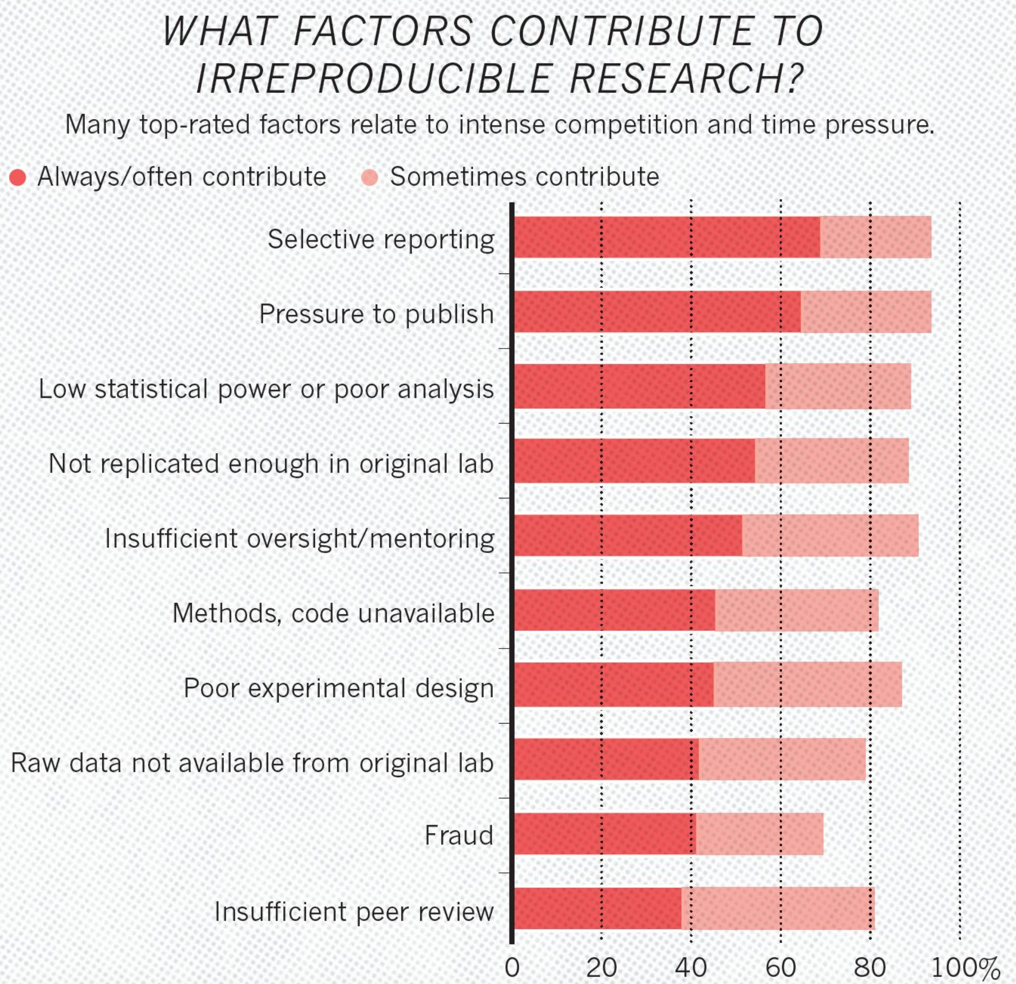
Nature Survey

- Survey of 1,576 researchers
- 70% of researchers failed to reproduce another scientist's experiments
- More than half have failed to reproduce their own experiments



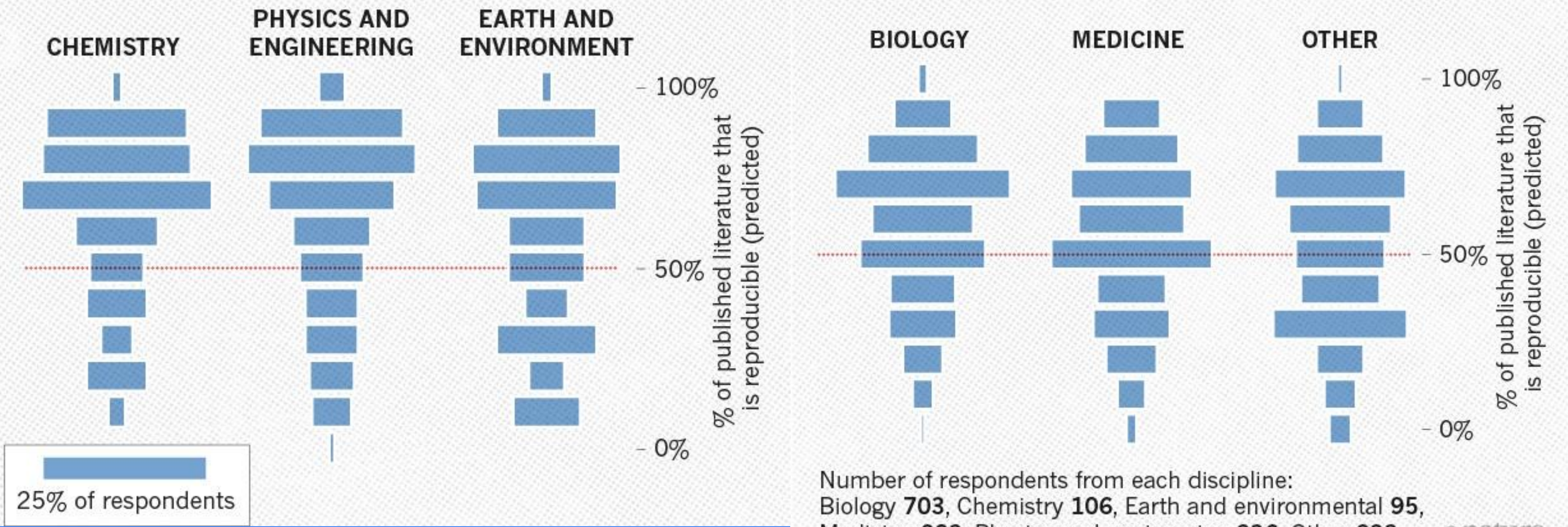
Nature Survey

- Selective reporting
- Pressure to publish
- Low statistical power
- Low replications
- Poor oversight/Mentori
- Raw data not available
- Fraud
- Insufficient Peer Review



HOW MUCH PUBLISHED WORK IN YOUR FIELD IS REPRODUCIBLE?

Physicists and chemists were most confident in the literature.

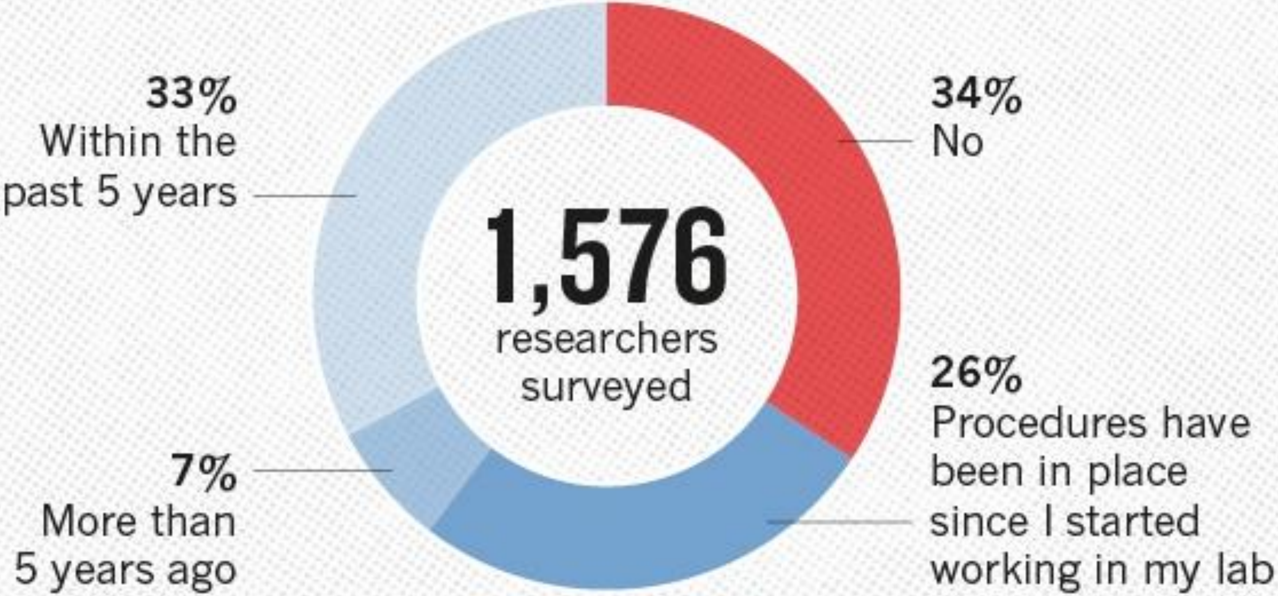


Number of respondents from each discipline:
Biology **703**, Chemistry **106**, Earth and environmental **95**, Medicine **203**, Physics and engineering **236**, Other **233**

©nature

HAVE YOU ESTABLISHED PROCEDURES FOR REPRODUCIBILITY?

Among the most popular strategies was having different lab members redo experiments.

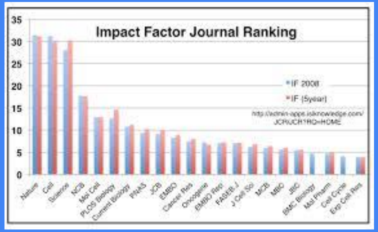


Give one example of a practice that could lead to poor reproducibility in research.

Nobody has responded yet.

Hang tight! Responses are coming in.

“The Reproducibility Crisis” Timeline
(Hilda Bastian)



1955 Eugene Garfield- Proposes The Journal Impact Factor

1958-59 Senator Estes Kefauver- “Sorry state of science supporting drug effectiveness”

1962 Kefauver-Harris Act- FDA requires 2 studies

1966 Stanley Gartler- Reports contamination of cancer cell lines in *Nature*

1976 Development of Meta analysis

1977 Michael Mahoney- Reports confirmation bias in peer reviews

http://blogs.plos.org/absolutely-maybe/2016/12/05/reproducibility-crisis-timeline-milestones-in-tackling-research-reliability/

1983 Doug Altman- Offers statistical guidelines for medical journals

1986 John Simes- Proposes international registry for clinical trials

1992 John Claerbout- Coins term ‘Reproducible Research’

1996 CONSORT- guidelines for reporting results of clinical trials

2004 ICMJE- refuses to publish non-registered clinical trials

2005 John Ioannidis- “Why Most Research Findings are False” *Science*

http://blogs.plos.org/absolutely-maybe/2016/12/05/reproducibility-crisis-timeline-milestones-in-tackling-research-reliability/

2007 FDA Amendment Act- public reporting of trial results (2017)

2012 Begley and Ellis- Amgen preclinical reproducibility study 11%

2013 Cortex- Requires registered studies

2013 Katharine Button- 20% power for neuroscience studies

2014 NIH- released reporting guidelines for preclinical research

2015 Open Science Collaboration- 30-50% Reproducibility

2016 American Statistical Association (ASA)- p value guidance

2016 *Nature* Survey

http://blogs.plos.org/absolutely-maybe/2016/12/05/reproducibility-crisis-timeline-milestones-in-tackling-research-reliability/

Factors Affecting Reproducibility

- Experimental Design
- Statistical Analysis
- Study Pre-registration
- Data Management and Sharing
- Incentives for Best Practices

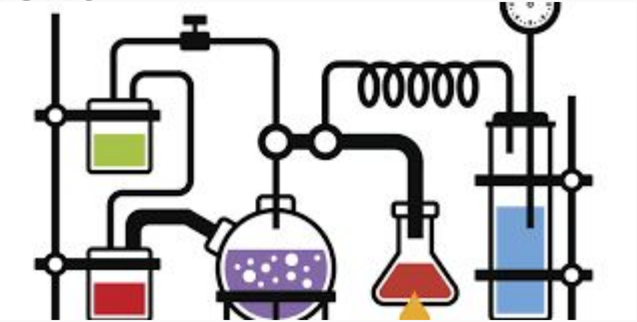
What is the most important aspect of experimental design?

Nobody has responded yet.

Hang tight! Responses are coming in.

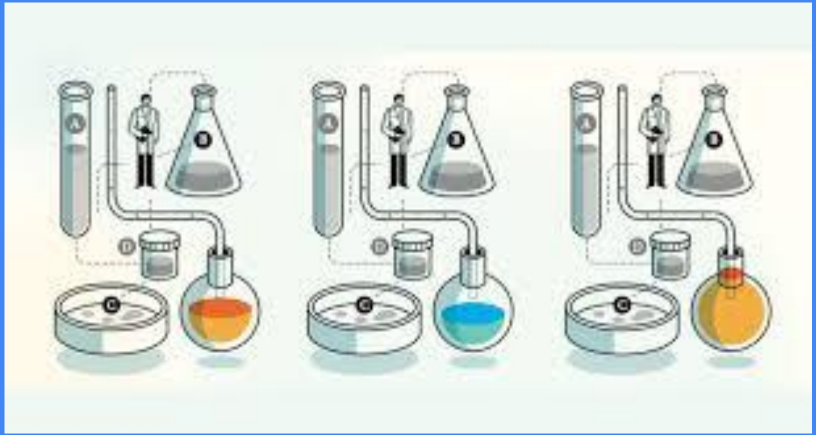
Experimental Design

- Replication
- Randomization
- Blinding
- Hypothesis
- Controls
- Calibration of tools
- Precision and Accuracy
- Observer Variability
- Baseline Assessment
- Recording Data



https://www.google.com/search?q=experimental+design&source=lnms&tbm=isch&sa=X&ved=0ahUKEwiorvrzherZAhWGj1QKHWNDDjUQ_AUICigB&biw=866&bih=615

Replication



- Confounding factors
  - Random variation
  - Unseen bias
- Replicate experiments (inter replication)
  - Can reveal bias
  - Tests ability to replicate conditions
- Increase sample size within an experiment (intra replication)
  - Decreases error due to variation
  - Improves precision
  - Can increase statistical power

https://www.sciencenews.org/article/redoing-scientific-research-best-way-find-truth

Figure 1. Technical replicates help identify variation in technique. For example, lysate derived from a mouse that is loaded three times (A1, A2, A3) on each membrane then run and measured independently will help identify variation in technique.

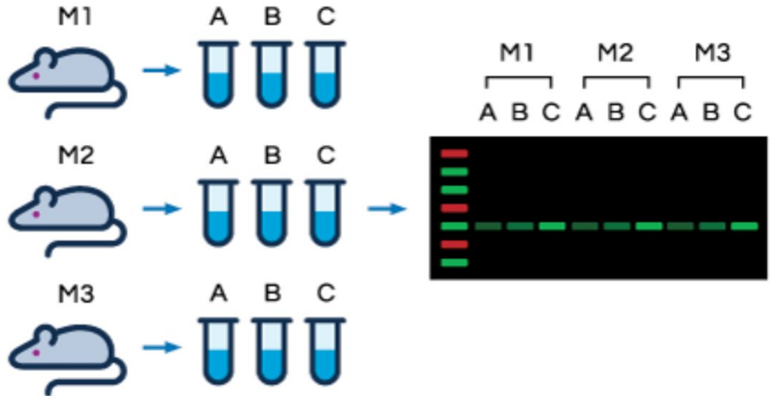


Figure 2. Biological replicates derived from independent samples help capture random biological variation. For example, three biological replicates (A, B, and C) are collected from three independent mice. Each of these biological replicates was run in three technical replicates (A1, A2, A3; B1, B2, B3; C1, C2, C3) in one Western blot assay.

https://www.licor.com/bio/blog/technical-and-biological-replicates

Replication vs Pseudo Replication

- When we think we are replicating but really are not
- Genetic similarities
    - Mice all from same colony
    - Sex of organism
    - Cell lines
- Hard to identify but worth it to consult experts in advance

Replication vs Pseudo Replication cont.

1. The biological entities of interest must be randomly and independently assigned to treatment groups.
2. The treatment(s) should be applied independently to each experimental unit.

   Example: Injecting animals with a drug is an independent application of a treatment, whereas putting the drug in the drinking water shared by all animals in a cage is not.

3. The experimental units should not influence each other, especially on the measured outcome variables.

   Often impossible to verify – how do you prove that the aggressive behaviour of one rat in a cage is not influencing the behaviour of the other rats?

https://blogs.bmj.com/openscience/2019/09/16/genuine-replication-and-pseudoreplication-whats-the-difference/

Randomization

- Powerful tool to overcome random variation
- Simplest method “coin flip”
- Block randomization can help prevent imbalance in study groups
- Stratification can help control confounding factors
- Best practice: blinded study

http://www.shiva3dstore.com/betterRNG

Blinding

- Powerful tool to alleviate bias (which is ubiquitous)
- Conscious or unconscious introduction of confounding factors by the researchers (or subjects)
- Clinical trials
  - Placebo controlled trials (blinds subjects)
  - Double blind also blinds researchers
  - Looking at preliminary data can bias analysis
- Slower adoption of this practice in animal and preclinical research
  - Animals and cell lines can respond to stimuli
  - Researchers always want their stuff to work...

Developing a Hypothesis

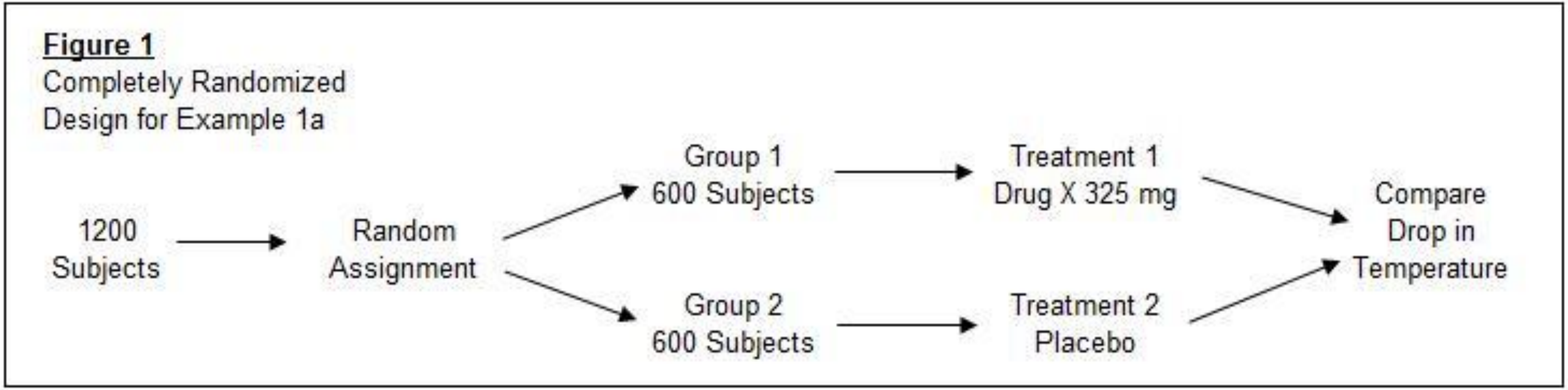
- A clear description of a reason for a given set of observations
- Robust experiments are designed to test the hypothesis (often the null hypothesis)
- Must differentiate between exploratory studies and robust hypothesis driven experiments
  - Pilot studies
  - Hypothesis generating studies

Block Design

- Break up a large experiment into smaller mini experiments
- Assists with reducing the amount of variation within each block- a statistical advantage

Implementing Block Design

- Assess any sources of variability in the experiment
  - Network with other researchers
  - Read the literature
- Block by the variables that could bias the experiment
  - Reagent batch
  - Researcher/observer
  - Cage of mice
  - Sex of organism
  - Strain of cells, passage of cells
- Attempt to distribute subjects/samples equally among the blocks



https://introductorystats.wordpress.com/2011/03/09/design-of-experiments/

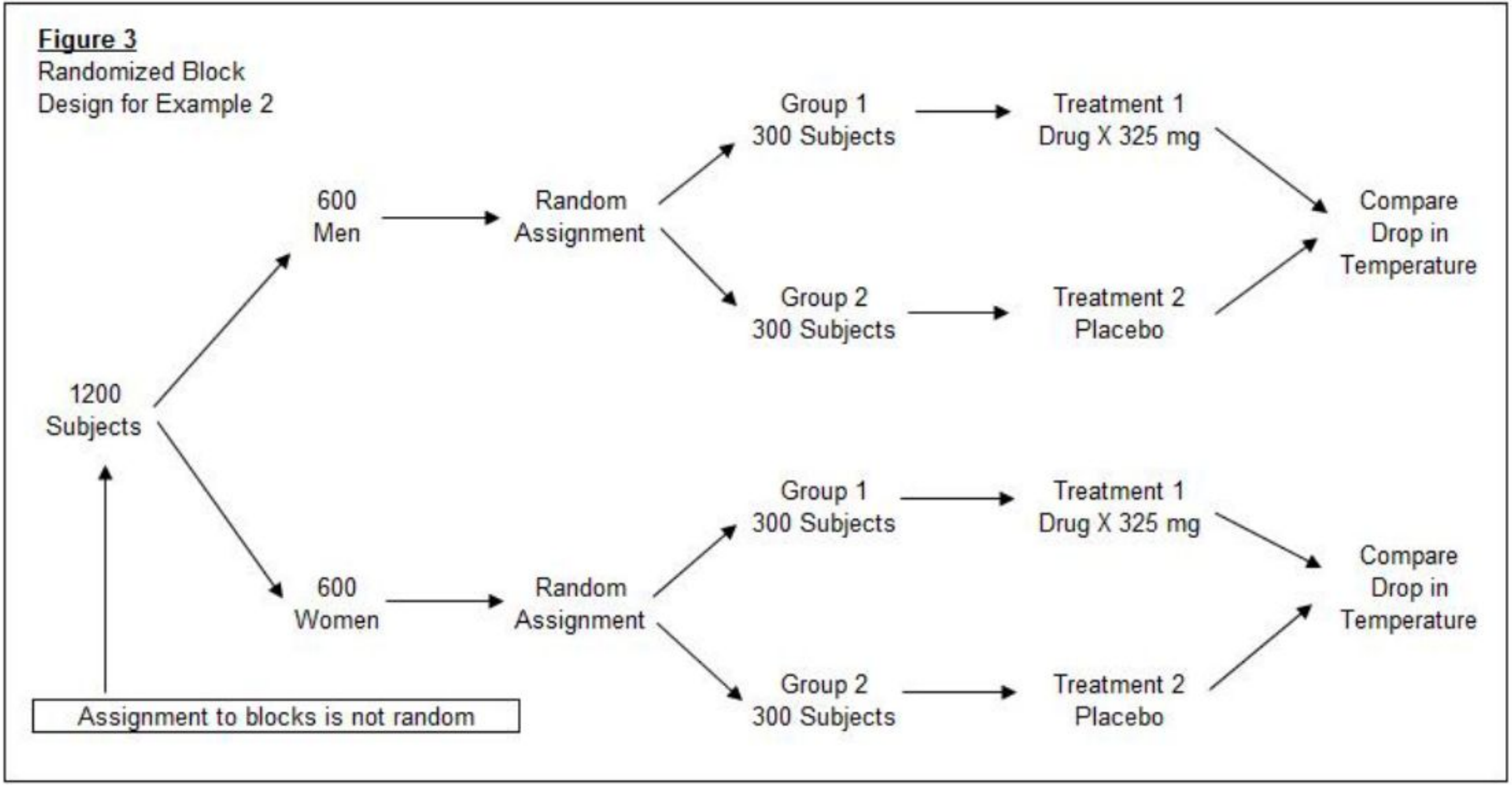
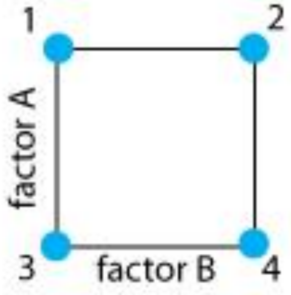


Figure 3
Randomized Block
Design for Example 2
1200
Subjects
600
Men
Random
Assignment
Group 1
300 Subjects
Treatment 1
Drug X 325 mg
Group 2
300 Subjects
Treatment 2
Placebo
Compare
Drop in
Temperature
600
Women
Random
Assignment
Group 1
300 Subjects
Treatment 1
Drug X 325 mg
Group 2
300 Subjects
Treatment 2
Placebo
Compare
Drop in
Temperature
Assignment to blocks is not random

Factorial design

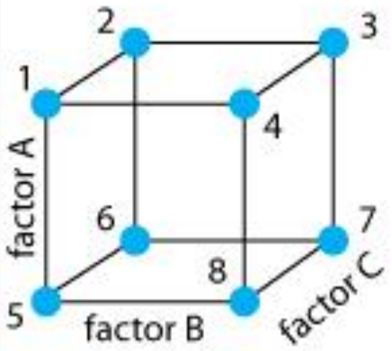
Useful for testing affects between different treatments or interventions

- How does A affect cell line
- How does B affect cell line
- How does A and B together affect cell line



factor levels

trial	A	B
1	+	−
2	+	+
3	−	−
4	−	+



factor levels

trial	A	B	C
1	+	−	−
2	+	−	+
3	+	+	+
4	+	+	−
5	−	−	−
6	−	−	+
7	−	+	+
8	−	+	−

http://community.asdlib.org/imageandvideoexchangeforum/2013/08/05/modeling-response-surfaces-using-factorial-designs/

Experimental Design for Human Studies

- Recruitment
  - Careful consideration for selecting participants that are representative of the population you want to make conclusions about
  - Thoughtful recruitment of groups less likely to participate
  - Changes to recruitment plan need to be approved by IRB and very well documented
- Adherence
  - Are participants following the protocol
  - Can also affect animal studies
  - Deviations need to be meticulously recorded

Negative Controls

- Help ensure the observed effect is due to the intervention and not some other confounding factor
- Examples:
  - Placebo
  - Carrier/excipient of an intervention
- Can be difficult to recognize all negative controls
- Network/read to gain insight

Positive Controls

- Assist with ensuring the negative controls are truly negative
- Ensure we can measure the effect we intend to assess
- Examples:
  - Known active intervention/treatment
  - “gold standard”
- Useful for calibrating instrumentation and determining dynamic range
- Tool to look at dose response or estimation of concentration

Calibration of tools

- All instruments require periodic calibration
  - Social science tools
  - Biological/Chemical instrumentation
  - Even people need calibration
- New instruments/tests/assays should undergo robust experimentation to validate
- Using an established instrument/test/assay in a new way also requires robust experimentation to validate the new process
- Validations can often be published

Precision and Accuracy

- Precision: measurements are tightly clustered
- Accuracy: measurements are close to the "real" value
- Ideally we want to achieve both
- Consider/measure the dynamic range of your instrumentation
- Accuracy and precision depend on calibration and validation

Intra-Observer Variability

- A researcher’s method of determining a measurement may drift over time
- Highly objective measurements that are well defined will have the least variability over time
- Making all observations for a given experiment in a short time frame can help minimize the variability
- Check yourself by replicating measurements at a different time
- Plotting replicates against each other can reveal variability and bias

Inter-Observer Variability

- Different researcher’s may record observations differently
- Having multiple researchers take the same measurement can reduce bias and increase accuracy (sometimes with a loss of precision).
- Defining measurement methods as tightly as possible and training of researchers can alleviate variability
- Conveying accurate and detailed descriptions of measurements in publications can increase reproducibility

Baseline Assessment

- Confirms the status of the samples/subjects prior to intervention
- Assesses the variability of samples/subjects before intervention
- Comparing baseline to end of experiment is more accurate than just comparing endpoints
- Careful thought needs to be given to baseline assessments especially in human and animal studies where stimuli can affect behavior

Which of these factors should be considered during calibration (choose all that apply)

Replications — 0%

Inter-observer variability — 0%

Hypothesis development — 0%

Baseline Assessment — 0%

Intra-observer variability — 0%

Other considerations

- Validation of cell lines
- Image manipulation and analysis

How do you feel about reproducibility in your work?

Poor — 1 0%

2 0%

Fair — 3 0%

4 0%

Excellent — 5 0%

End for Today

- Remember your discussion sections!
- Next at Parnassus (Zoom if Mission Bay is your home campus)