Playing with Independence in (Some) Basic Science Experiments:

Independence, Replicates, and the Experimental Unit plus Clustered data

Kathleen C. Torkko

September 23, 2019

Objectives

Revisit independence and paired/dependent/correlated data

Biological vs. technical replicates

Learn what an experimental unit is

Learn what an observational unit is

Learn what a cluster is and how to deal with it

Remember independence?

One of the common and important assumptions of statistical tests

INDEPENDENCE

- a study subject/sample is independent from other subject/samples
  different people, different tumors

- measurements within a subject/sample are independent
  foot size and blood pressure

When observations are not independent, the dependency between them has the potential to decrease variance within that group and can lead to false conclusions

Dependent/paired/correlated samples

Tricks: The words like dependent, repeated, before and after, matched pairs, paired and so on are hints for dependent samples

- pre-test/post-test samples in which a factor is measured before and after an intervention
- cross-over trials in which individuals are randomized to two treatments and then the same individuals are crossed-over to the alternative treatment
- matched samples, in which individuals/samples are matched on personal characteristics such as age and sex
- duplicate measurements on the same individual/samples
- individuals/samples matched by location and/or time

Dependent or Independent?

Do people catch more fish by fishing from a boat or from the shore? Let Row B represent hours fishing per fish from the shore, and let Row A represent hours fishing per fish using a boat. The following data are paired by month from April through October.

	Apr	May	June	July	Aug	Sept	Oct
B: Shore	3.3	3.6	3.9	3.2	2.0	1.8	1.6
A: Boat	3.8	3.0	3.3	2.2	1.6	1.4	1.5

Dependent or Independent?

A researcher conducted an experiment to see if specific eye exercises can improve peripheral vision. A random sample of 5 people were rated for peripheral vision on a scale from 1 to 20 where 9 is considered average and higher scores indicate better peripheral vision. Then they followed the prescribed eye exercise program and were rated again. The results follow:

Subject	1	2	3	4	5
Before	9	8	7	10	6
After	10	9	11	12	9

Dependent or Independent?

Is there a significant difference in the mean number of physical therapy sessions attended between male and female patients?

Gender	Number of Patients	Mean Number of Sessions Attended	Standard Deviation
Male	20	14.6	6.3
Female	15	18.1	5.9

How Many Patients? How Many Limbs? Analysis of Patients or Limbs in the Orthopaedic Literature: A Systematic Review

By Dianne Bryant, MSc, PhD, Thomas C. Havey, BHSc, Robin Roberts, MSc, and Gordon Guyatt, MD, MSc

Investigation performed at McMaster University, Hamilton, Ontario, Canada

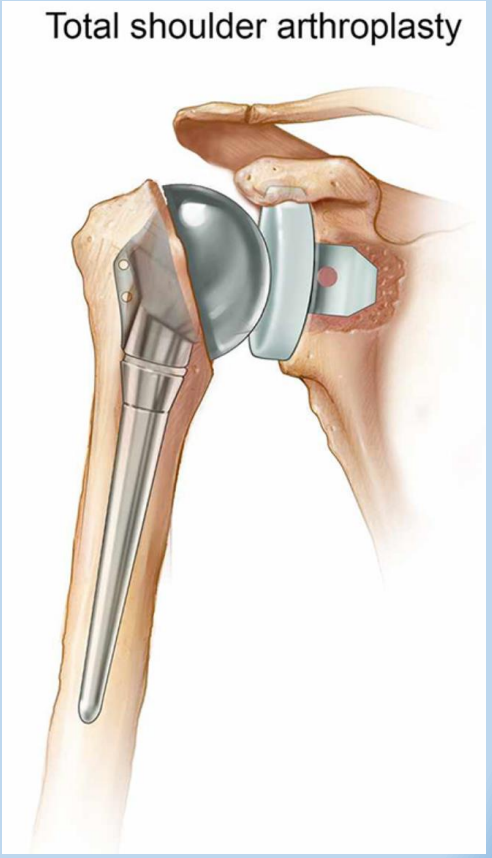
The Journal of Bone & Joint Surgery · jbjs.org
Volume 88-A · Number 1 · January 2006

When two limbs or multiple joints from a single patient are counted as independent observations, the precision of the estimate may be falsely improved and the potential for a biased estimate increases.

Dependency between Observations

A patient undergoes bilateral shoulder arthroplasty, with the shoulder replacements occurring within months of each other.

No matter how the first shoulder responds to the intervention, the outcome for the second shoulder is likely to be affected, either positively or negatively, by the outcome for the first shoulder, creating an association between the outcomes for each shoulder.



THE JOURNAL OF BONE & JOINT SURGERY · JBJS.ORG
VOLUME 88-A · NUMBER 1 · JANUARY 2006

Our findings suggest that a high proportion (42%) of clinical studies in high-impact-factor orthopedic journals involve the inappropriate use of multiple observations from single individuals, potentially biasing results.

We suggest that before initiating any orthopedic investigation involving an intervention on multiple limbs or joints in individual patients, investigators should consult with an experienced biostatistician.

THE JOURNAL OF BONE & JOINT SURGERY · JBJS.ORG
VOLUME 88-A · NUMBER 1 · JANUARY 2006

Dependent or Independent?

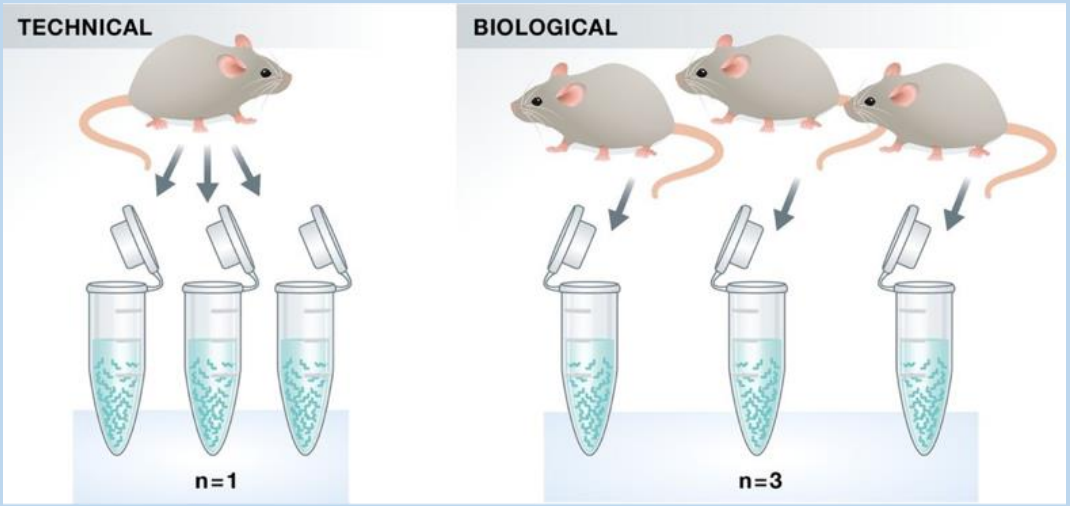
Heights of fathers and sons

Heights of women and their best friends

Fraternal twins
- Genetic studies
- Dietary studies on twins who grew up in the same household
- Dietary studies on twins who grew up in different homes

Identical twins
- Genetic studies
- Dietary studies on twins who grew up in the same household
- Dietary studies on twins who grew up in different homes

Independence and Replicates



Dependent samples *Independent samples*

- **Technical replicates:** use the same biological sample to repeat the technical or experimental steps in order to accurately measure technical variation and to control it
  - Not meant to measure biological variability

- **Biological replicates** use different biological samples of the same condition to measure the biological variation between samples
  - goal of research and usually determines sample size

hbctraining.github.io/GCC-BOSC-2018/lessons/experimental_planning_considerations.html

Replicates

Consider a colony assay using bone marrow cells cultured in soft agar

Set up triplicate plates for each of the conditions

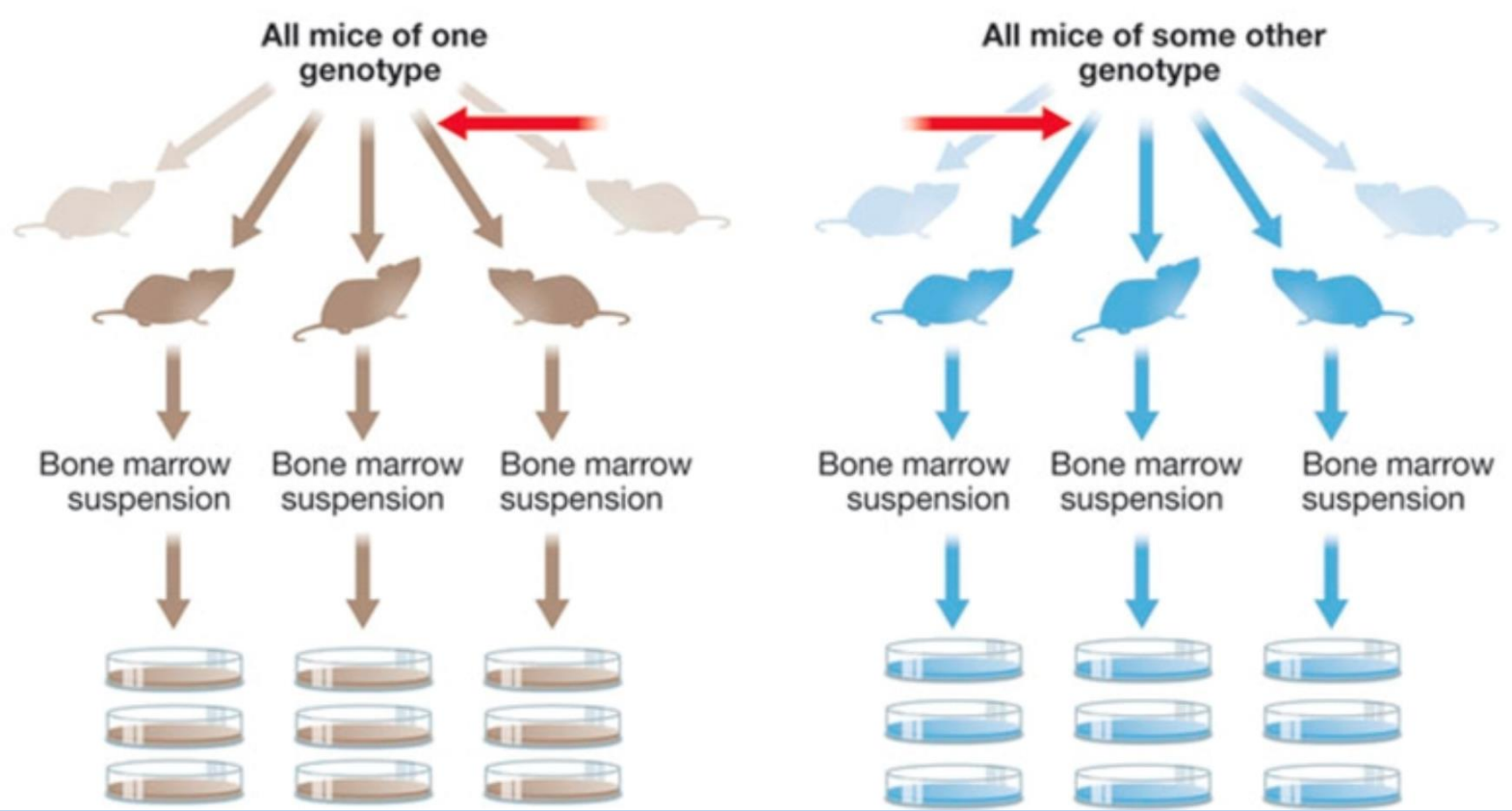
- Triplicates are intended to be as similar as possible to each other
- Expectation that each of the triplicate plates to grow a similar number of colonies

If one of the cultures in a triplicate has more or less colonies than the other two, need to find out why

- Maybe cells not dispersed properly, and one of the triplicate cultures got a clump of cells
- Maybe the colonies are not bone marrow–derived but contaminating yeast colonies
- Maybe there was a blockage in the pipette when the cells were dispensed

Biological replicates

Technical replicates



Three independent mice (biological replicates) were chosen from each genotype, so we can make inferences about all mice of that genotype. Note that in the experiments, *n* = 3, no matter how many technical replicate plates are created.

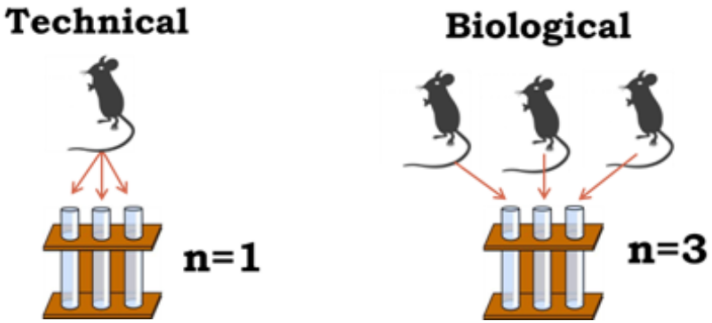
David L Vaux et al. EMBO Rep. 2012;13:291-296

Technical and biological replicates

Always easy to tell the difference?

- Definition of **technical** and **biological** depends on the model and the question.

- The model: mouse, rat … mammals in general.
  - Easy: one value per individual
    - e.g. weight, neutrophils counts …



- What to do? Mean of technical replicates = 1 biological replicate

- Some biological replicate cases are obvious: e.g. tissue samples from different mice, blood samples from different people, cell cultures from different people.

- Others are less clear cut:
  - e.g. Experiments using a single cell line

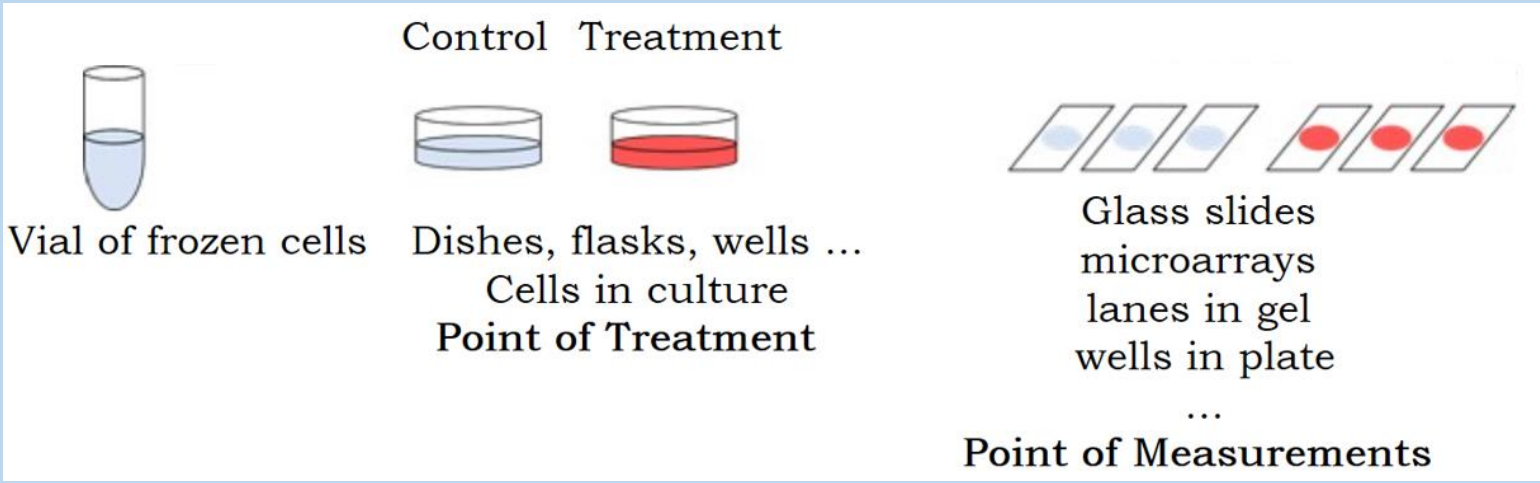
slideplayer.com/slide/11256121/

Technical and biological replicates

Always easy to tell the difference?

Using material from a single cell line:

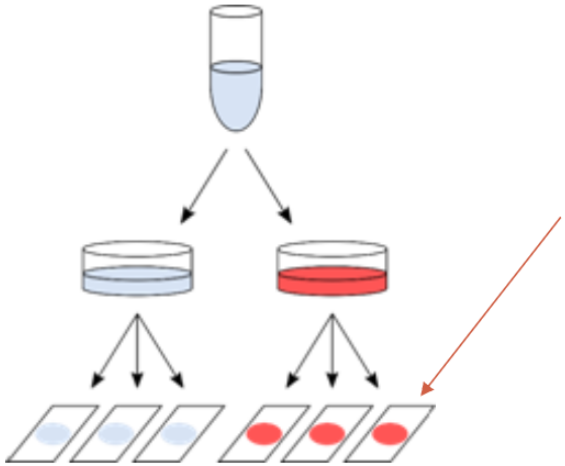
Are these true biological replicates?



slideplayer.com/slide/11256121/

Technical and biological replicates
Cell cultures

- Design 1: As bad as it can get



One value per glass slide
e.g. cell count

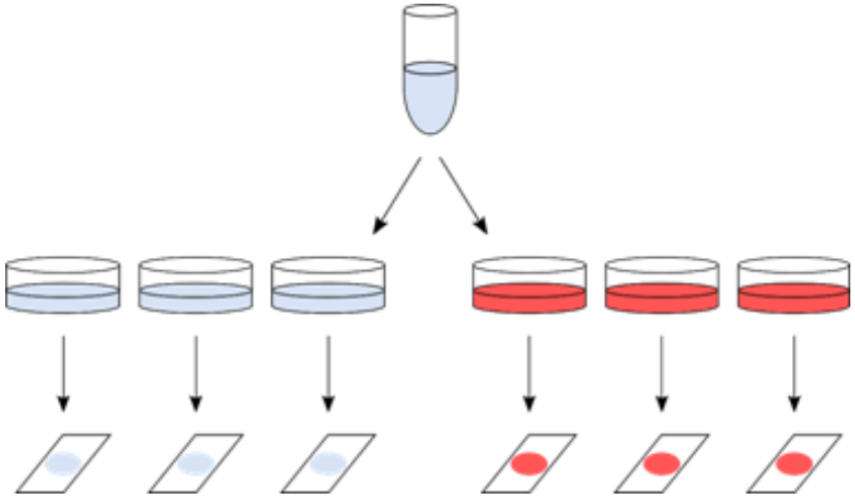
- After quantification: 6 values
  - But what is the sample size?
    - n = 1
      - no independence between the slides
      - variability = pipetting error

there is no independence among different groups

slideplayer.com/slide/11256121/

Technical and biological replicates
Cell cultures

- Design 2: Marginally better, but still not good enough



Everything processed on the same day

After quantification: 6 values

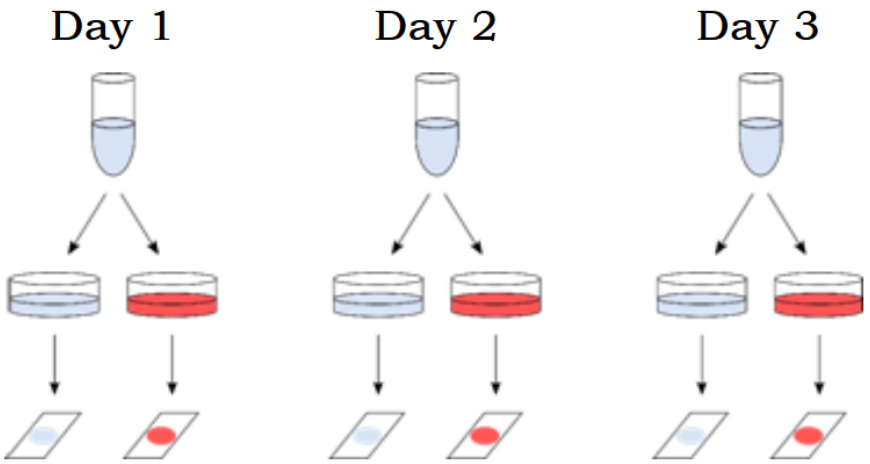
- But what is the sample size?
  - n = 1
    - no independence between the plates
    - variability = a bit better as sample split higher up in the hierarchy

repeats are not independent replicates

slideplayer.com/slide/11256121/

Technical and biological replicates
Cell cultures

- Design 3: Often, as good as it can get



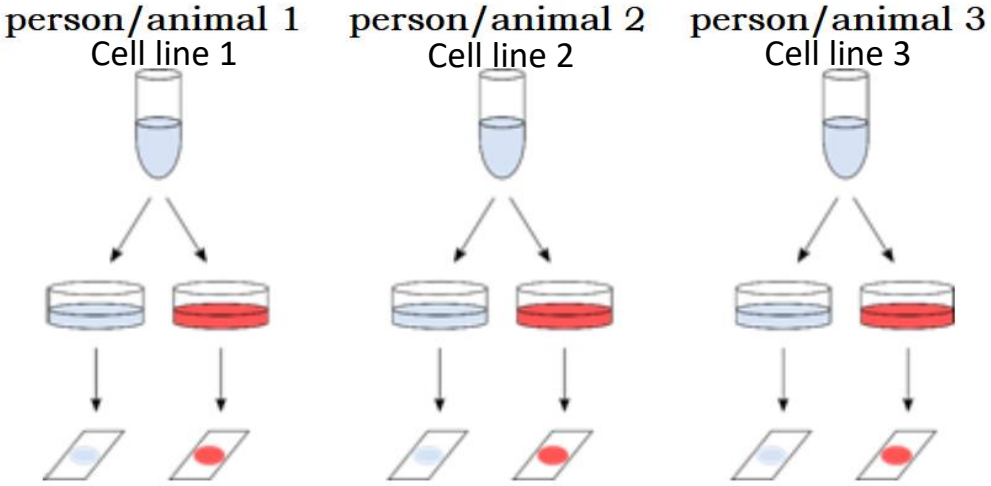
After quantification: 6 values

- But what is the sample size?
  - **n = 3**
    - Key difference: the whole procedure is repeated 3 separate times
    - Still technical variability but done at the highest hierarchical level
    - Results from 3 days are (mostly) independent
    - Values from 2 glass slides: paired observations

slideplayer.com/slide/11256121/

Technical and biological replicates
Cell cultures

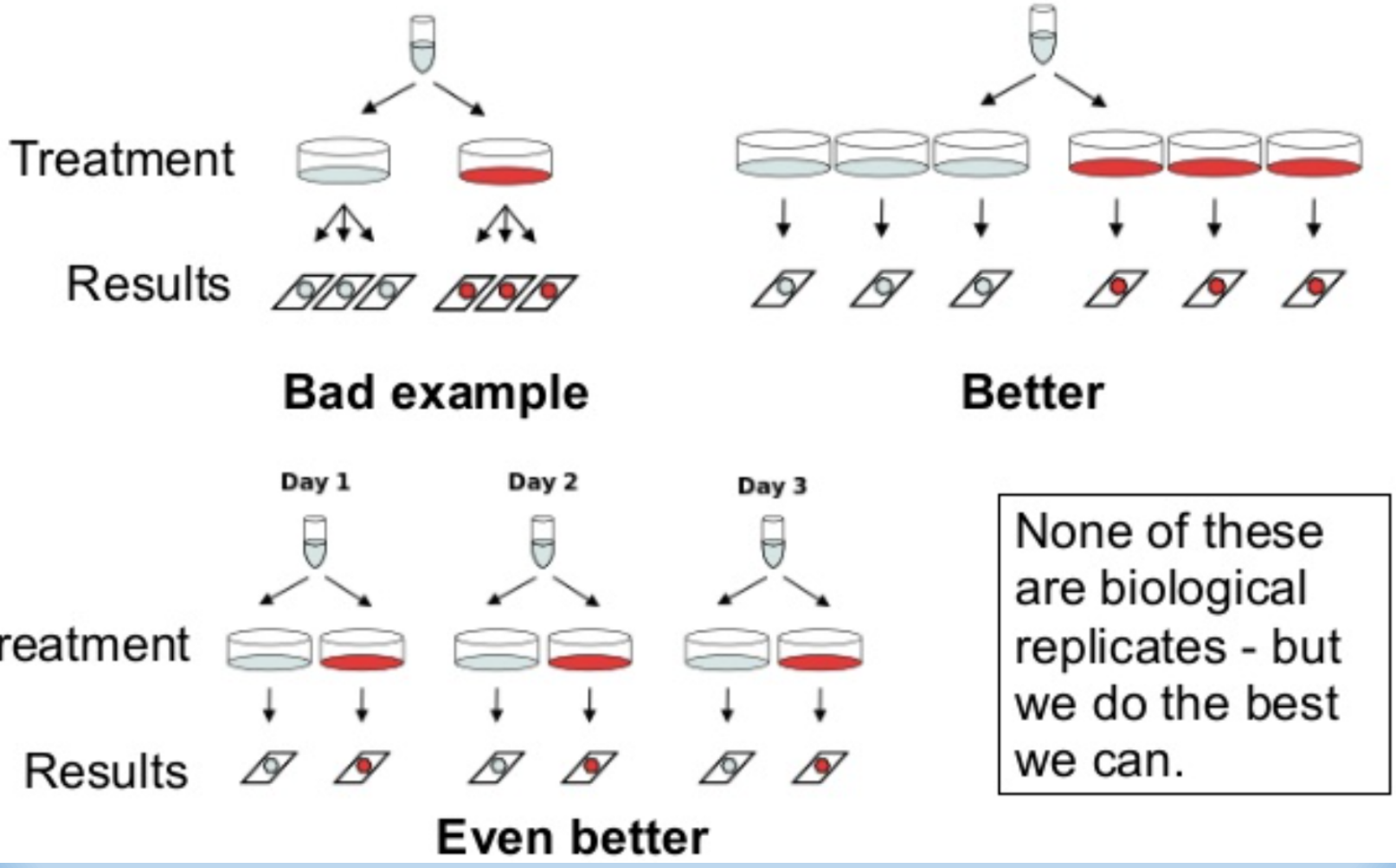
- Design 4: The ideal design



- After quantification: 6 values
  - But what is the sample size?
    - n = 3
      - Real biological replicates

slideplayer.com/slide/11256121/

Cell line replicates



www.slideshare.net/jelena121/introduction-to-ngs-experimental-design

What is the minimum number of replicates*?

Experiments with only one replicate **break statistics and make puppies cry.**

*Biological not technical
(The author said n=3)

In truth, I prefer the term “biological samples” to “biological replicates”

www.slideshare.net/jelena121/introduction-to-ngs-experimental-design

The Experimental Unit of Biological Samples

The experimental unit is the physical entity which can be assigned, at random, to a treatment.

Commonly it is an individual human, animal

cage is the sample size, but not individuals

The experimental unit is also usually the unit of statistical analysis.

The experimental unit may be:

An individual animal (human, mouse, cell line, etc.)
A family or a mouse and her litter
A cage of animals or a classroom
A part of an animal

The individual animal:

This is the most common instance.

Individual animals are assigned to the treatments

NOTE: If the treatment is given to several animals in the same cage in the diet or water then individuals within a cage can not receive different treatments, so the individual animal is not the experimental unit.

We will focus on non-human animal studies for examples

The breeding female and her litter:

In some experiments the pregnant mouse is treated with the test substance, but measurements are made on individual pups after birth.

n of 1 for all the pups from the same cage, the same mother.

Because individual pups within a litter can not have received different treatments the experimental unit is the whole litter and the unit of statistical analysis is the litter mean or proportion affected, possibly weighted by litter size.

If there are, say, 10 treated and 10 control females, then the statistical tests will be based on a group size of 10, not on the numbers of individual pups

unless you can treat the animals individuals there some cage clustered factors

The cage of animals:

If the treatment is incorporated into the diet or water then any two or more animals within the same cage can not be assigned to different diets so the experimental unit will be the cage of animals.

Cage effects (differences between cages on the same treatment) can occur as a result of fighting, sub-clinical infection or differences in environment due to cage position.

The statistical analysis will then be based on the mean of all animals within a cage, and "N" will be the number of cages, not the number of animals.

NOTE: if the treatment can be given to individual animals within a cage, say by injection, then the individual animal may be the experimental unit.

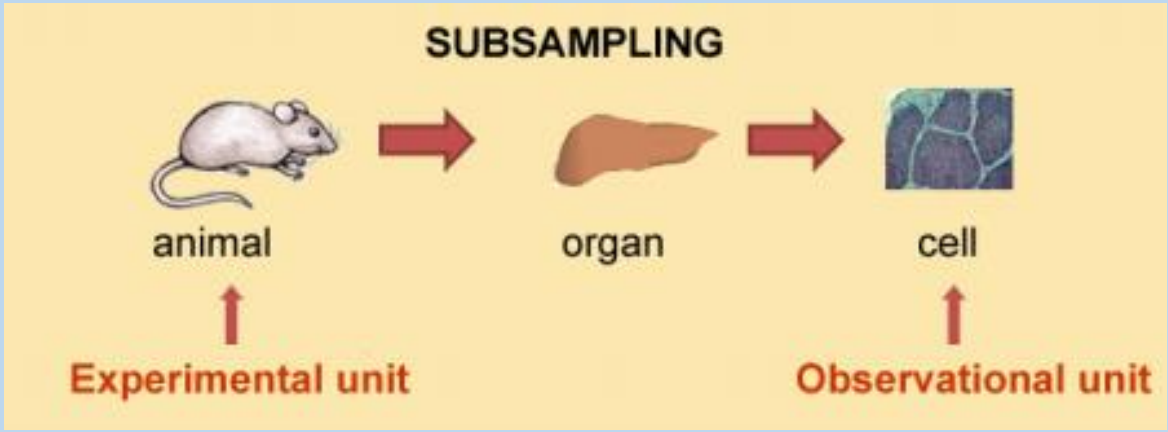
A part of an animal:

If the experimental treatment is the application of a substance to the shaved skin of an animal, then it may be possible to divide an area of skin into a number of patches which can receive different treatments. In this case it is the patch on the back of an animal that is the experimental unit.

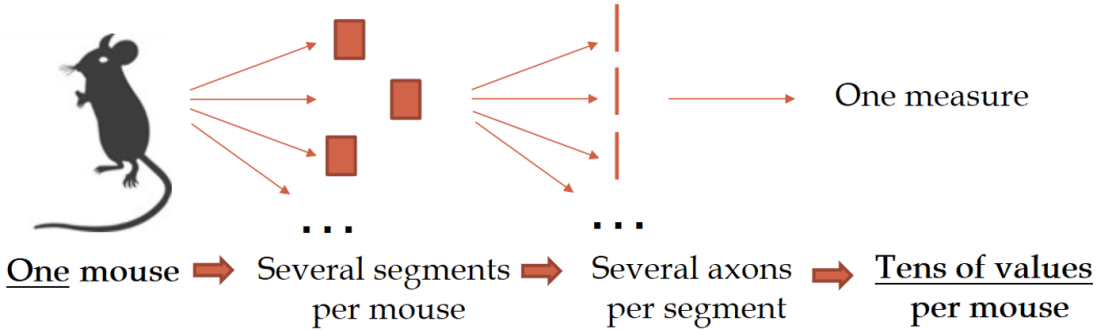
Similarly, some paired organs may be considered as the experimental unit if they can be assigned to different treatments.

If individual cells, say in the brain, can be given different stimuli and recordings of response are made, then individual cells are the experimental unit because different cells can receive different treatments.

In this case it may be possible to get a lot of experimental units out of a single individual. However, usually it is wise to use three or four animals in case individuals behave or react differently.



- The model is still: mouse, rat … mammals in general.
  - Less easy: more than one value per individual
    - e.g. axon degeneration



- What to do? Not one good answer.
  - In this case: mouse = experiment unit

EXAMPLE *Carcinogenic substances*

20 rats are randomly assigned to each of 4 doses of a potential carcinogen: none, low, medium, and high. The rats are kept in individual cages under the same environmental conditions in the same room. Each rat has its assigned dose stirred into its daily meal for 4 weeks. The number of tumors found in each rat is recorded at the end of the 4 week period. What is the experimental unit?

N=20

http://www.biostat.umn.edu/~lynn/ph7406/notes/basicprinciples.pdf

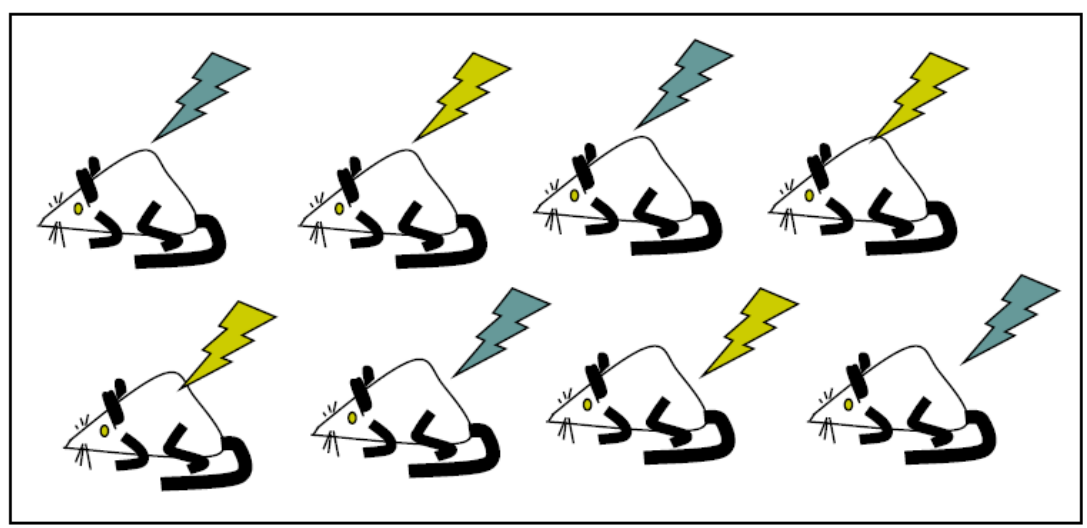
EXAMPLE *Carcinogenic substances*

20 rats are randomly assigned to 4 cages (5 rats in each). Each cage is then randomly assigned to one dose of a potential carcinogen: none, low, medium, and high. The rats are kept in their assigned cages under the same environmental conditions in the same room. Rats are not fed individually; food is placed in each cage in a common dish out of which all rats eat. Each cage has its assigned dose stirred into the food for 4 weeks. The number of tumors found in each rat is recorded at the end of the 4 week period.

N=4, cage number is the replicate

The animal as the experimental unit

Animals individually treated. May be individually housed or grouped

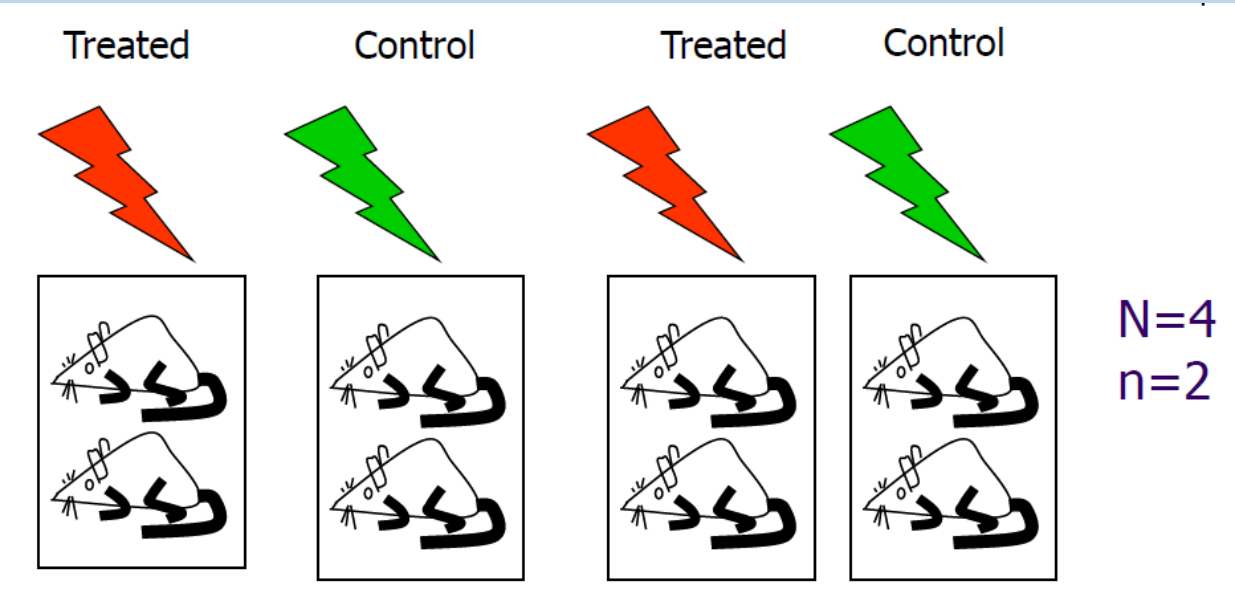


N=8 Total n for analyses

n=4 n for each group

Cage as the experimental Unit

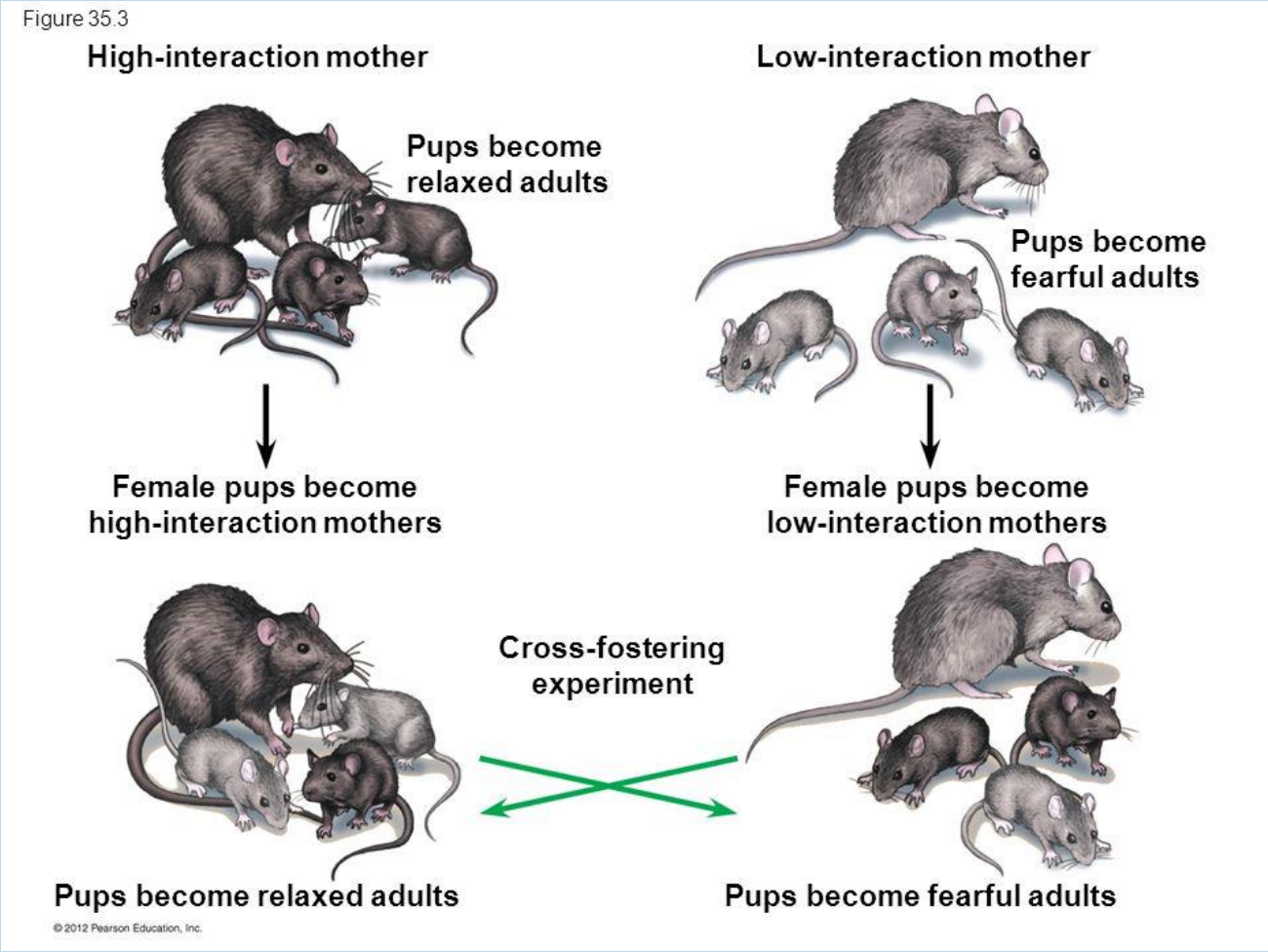
Treatment in water/diet



N=4
n=2

Total n for analyses
Observational units for each cage

What is the experimental unit?



Example – Scientists want to study the effect of an anti-bacterial drug in fish lungs. The drug is administered at 3 dose-levels (0, 20, and 40 mg/L). Each dose is administered to a large controlled tank through the filtration system. Each tank has 100 fish. At the end of the experiment, the fish are sacrificed, and the amount of bacteria in each fish is measured.

Treatment:

Dose levels of antibacterial drug

Response:

Amount of bacteria

Experimental unit:

Tanks

Observational unit:

Fish

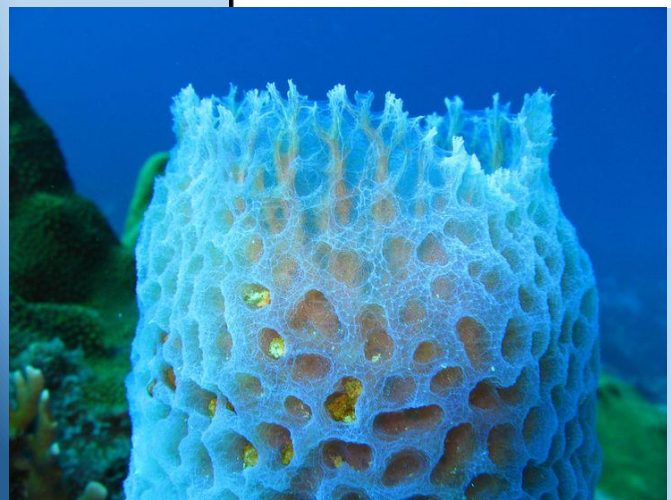
What is the n for data analysis?

Experimental units and observational units

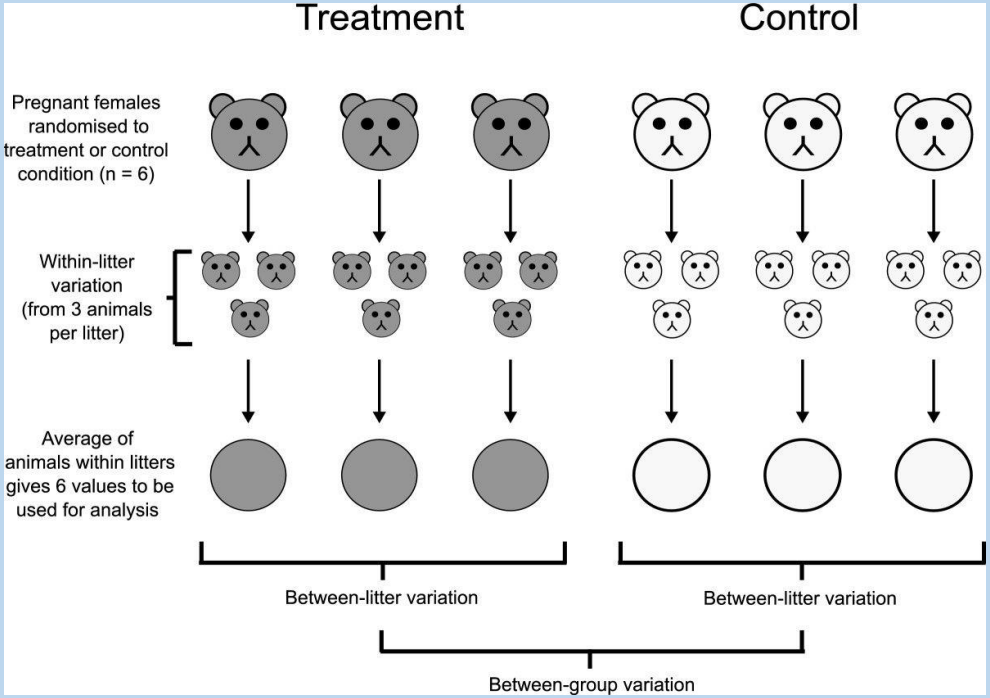
Example adapted from George W. Cobb *Introduction to Design and Analysis of Experiments* (1998, p. 126):

A graduate student was studying the shape of sponge cells to see whether the shapes of the cells depended on the colour of the sponge (white or green) and the location of the cells (tip or base). He painstakingly measured hundreds of cells under a microscope, but, unfortunately, he had based his experimental plan on bad advice: all his hundreds of cells came from just two sponges: one white and one green.

The goal of his study was to compare two hypothetical populations: all the white sponges and all the green sponges of a particular type. Thus the experimental unit was a sponge, not a cell, and he only had two experimental units (sponges) even though he had almost one thousand observational units (cell measurements).



Defining the experimental unit

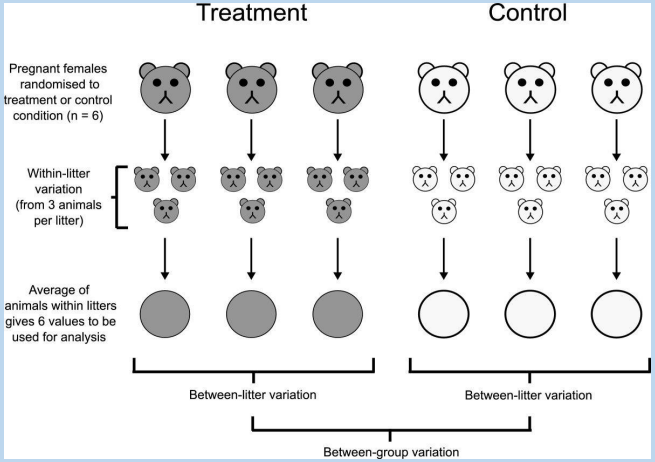


Pregnant females are the experimental units because they are randomized to the treatment or control conditions and therefore n = 6 in this example.

The three offspring within a litter will often be more alike than offspring from different litters and multiple offspring within a litter are observational units.

Values for each pup in a litter will be averaged and serve as the value for the experimental unit

Defining the experimental unit



Using all of the offspring without averaging will result in an inflated sample size (pseudoreplication)

Instead of averaging, one could randomly select only one animal from each litter, or use a mixed-effects model to appropriately partition the different sources of variation.

The only way to increase sample size, and thus power, is to increase the number of litters used.

False Positives

Pseudoreplication can lead to inflated Type-1 error rates

Too many false positives occur when dependent samples are analyzed as independent

- Artificially larger sample sizes leads to easier rejection of the null hypothesis

- Reduced variance from correlated samples leads to easier rejection of the null hypothesis

Addressing the Unit of Analysis in Medical Care Studies

A Systematic Review

Aaron W. Calhoun, MD, Gordon H. Guyatt, MD, MSc,†‡ Michael D. Cabana, MD, MPH,§¶ Downing Lu, MD,**‖ David A. Turner, MD,**‖ Stacey Valentine, MD,**‖ and Adrienne G. Randolph, MD, MSc**‖*

Medical Care • Volume 46, Number 6, June 2008

A systematic review of medical journal articles that could be subject to misidentification of the experimental/analysis unit discovered a misidentification error in 44% of articles

INTERNATIONAL JOURNAL OF ONCOLOGY 47: 1655-1662, 2015

Effectiveness of plasma treatment on pancreatic cancer cells

NORIFUMI HATTORI[1], SUGURU YAMADA[1], KOJI TORII[1], SHIGEOMI TAKEDA[1], KAE NAKAMURA[2], HIROMASA TANAKA[3], HIROAKI KAJIYAMA[2], MITSURO KANDA[1], TSUTOMU FUJII[1], GORO NAKAYAMA[1], HIROYUKI SUGIMOTO[1], MASAHIKO KOIKE[1], SHUJI NOMOTO[1], MICHITAKA FUJIWARA[1], MASAAKI MIZUNO[3], MASARU HORI[4] and YASUHIRO KODERA[1]

Departments of [1]Gastroenterological Surgery (Surgery II) and [2]Obstetrics and Gynecology, [3]Center for Advanced Medicine and Clinical Research, Nagoya University Graduate School of Medicine, Nagoya 466-8550; [4]Plasma Nanotechnology Research Center, Nagoya University, Nagoya 464-8603, Japan

Received July 1, 2015; Accepted August 12, 2015

Same material injected bilaterally

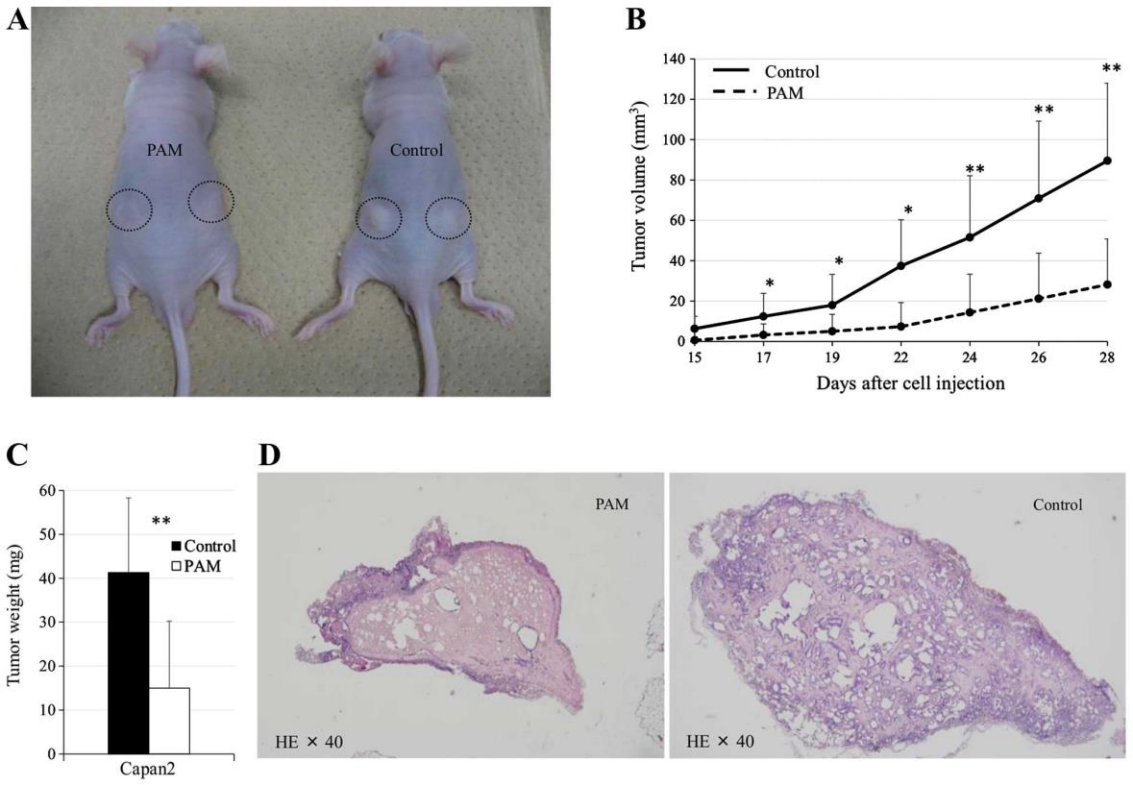
Animal studies

Six-week-old male nude mice (BALB/C) (N=10) were used. A total of 5×10^3 Capan2 cells were suspended in 50 µl of RPMI-1640 and 150 µl of Matrigel, and subcutaneously injected into the bilateral flank of mice. Then, mice were divided into a control group and a PAM-treated group. Mice in the control group received 200 µl of RPMI-1640, whereas the PAM-treated group received 200 µl of PAM by subcutaneous injection.

To evaluate antitumor effects, the tumor volume was calculated using the formula: $\pi/6 \times (\text{largest diameter}) \times (\text{smallest diameter})^2$. At 29 days after cell injection, the mice were sacrificed and tumors were harvested and weighed.

Statistical analysis

All data are presented as means ± SD. Statistical analysis of the data was performed using a Student's t-test. p-values <0.05 were considered statistically significant.



Even though the authors did not state it, what do you think the experimental unit is? What is the n in the picture in the upper left?

Figure 6 - Antitumor effect of PAM on Capan2 tumor xenografts in a mouse model. Macroscopic tumor formation in each group. (B) Tumor growth curves of xenografts of Capan2 subcutaneous injections in each group are shown. Each point on the line represents the mean tumor volume and the bars represent the SD. *p<0.05, **p<0.01. (C) The bar graph shows the mean tumor weight in each group. **p<0.01. (D) H&E staining of tumor tissues in each group.

www.impactjournals.com/oncotarget/ Oncotarget, Vol. 6, No. 30

ELK1 is up-regulated by androgen in bladder cancer cells and promotes tumor progression

Takashi Kawahara[1,2,3,*], Hasanain Khaleel Shareef[1,4,*], Ali Kadhim Aljarah[1,5], Hiroki Ide[1], Yi Li[2,6], Eiji Kashiwagi[1], George J. Netto[1], Yichun Zheng[1,2,6], Hiroshi Miyamoto[1,2]

[1]Departments of Pathology and Urology, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[2]Department of Pathology and Laboratory Medicine, University of Rochester Medical Center, Rochester, NY, USA

[3]Department of Urology, Yokohama City University School of Medicine, Yokohama, Japan

[4]Department of Biology, University of Babylon College of Science for Women, Babylon, Iraq

[5]Department of Biology, University of Baghdad College of Science, Baghdad, Iraq

[6]Department of Urology, 2nd Affiliated Hospital of Zhejiang University School of Medicine, Hangzhou, China

[*]These authors have contributed equally to this work

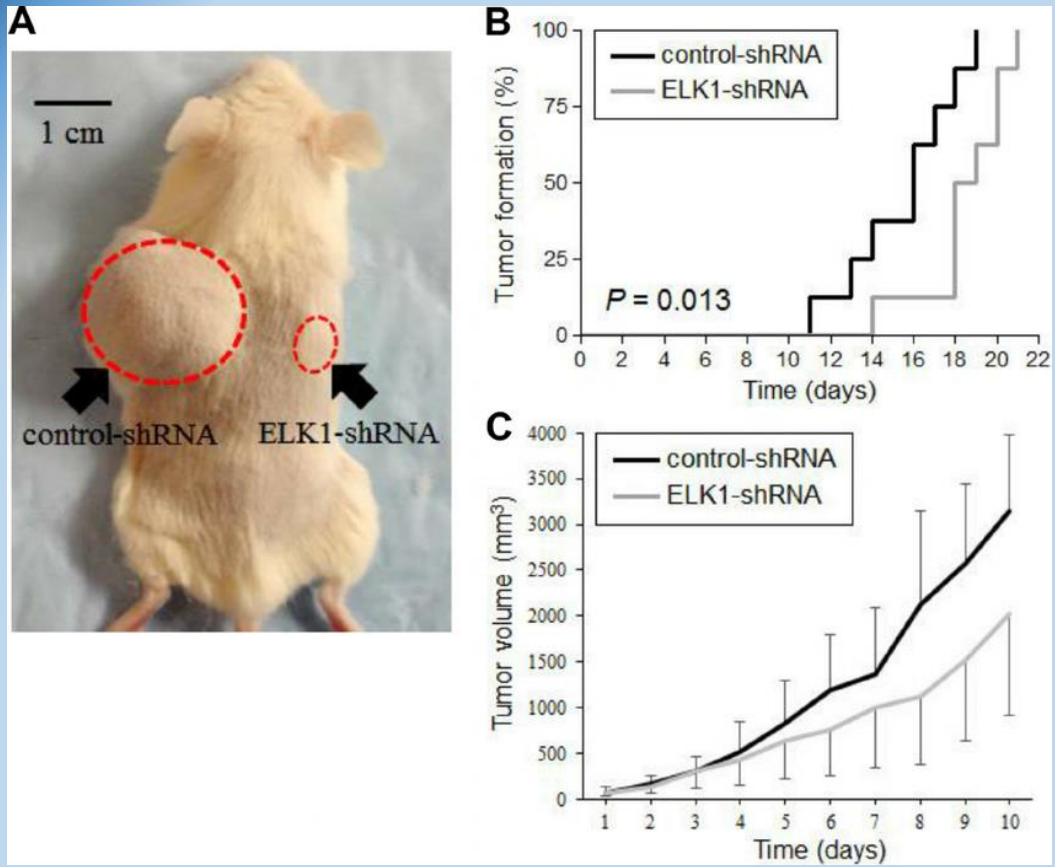
Correspondence to:
Hiroshi Miyamoto, **e-mail:** hmiyamo1@jhmi.edu

Keywords: *androgen, bladder cancer, ELK1, immunohistochemistry, tumor progression*

Received: *June 19, 2015* ***Accepted:*** *August 12, 2015* ***Published:*** *August 22, 2015*

Different material injected bilaterally

Mouse xenograft model. Animal protocols in accordance with the National Institutes of Health Guidelines for the Care and Use of Experimental Animals were approved at our institution. Cells (5 × 105/100 µL/site) resuspended in Matrigel were subcutaneously injected into the flank of 6-week-old male immunocompromised NOD-SCID mice, as described previously [40, 43]. Serial caliper measurements of perpendicular diameters were used to calculate tumor volume by the following formula: (short diameter)2 × (longest diameter) × 0.5.



Even though the authors did not state it, what do you think the experimental unit is? What is the n in the picture?

Figure 8: Effects of ELK1 inactivation on tumor growth in mouse xenograft models for bladder cancer. A. UMUC3- control-shRNA/ELK1-shRNA cells were implanted subcutaneously into the left/right flanks of NOD-SCID mice, respectively, and tumor formation and its growth were monitored. B. Kaplan-Meier curves and log-rank test according to the endpoint set as tumor volume exceeding 40 mm 3 . C. Tumor size (estimated volume of each tumor exceeded 40 mm 3 at day 0) was subsequently monitored every day. Each value represents the mean (+SD or –SD).

Are tumors in the same mouse correlated or independent?

Experiment: implant small pieces of tumor from human donor into each flank of a mouse. Each individual mouse was treated through oral lavage with either placebo or drug. Tumor growth was observed over 4 weeks. Size of tumor was calculated each week using caliper measurements.

Tumors from each flank of the mouse treated as independent tumors

n=2 for each mouse

If treated as paired (correlated) tumors, use mean of tumors

n=1 mouse

Argument for independent tumors:

Each implanted tumor piece varied in size and possibly amount of tumor

Reduces number of mice needed

Argument for dependent tumors:

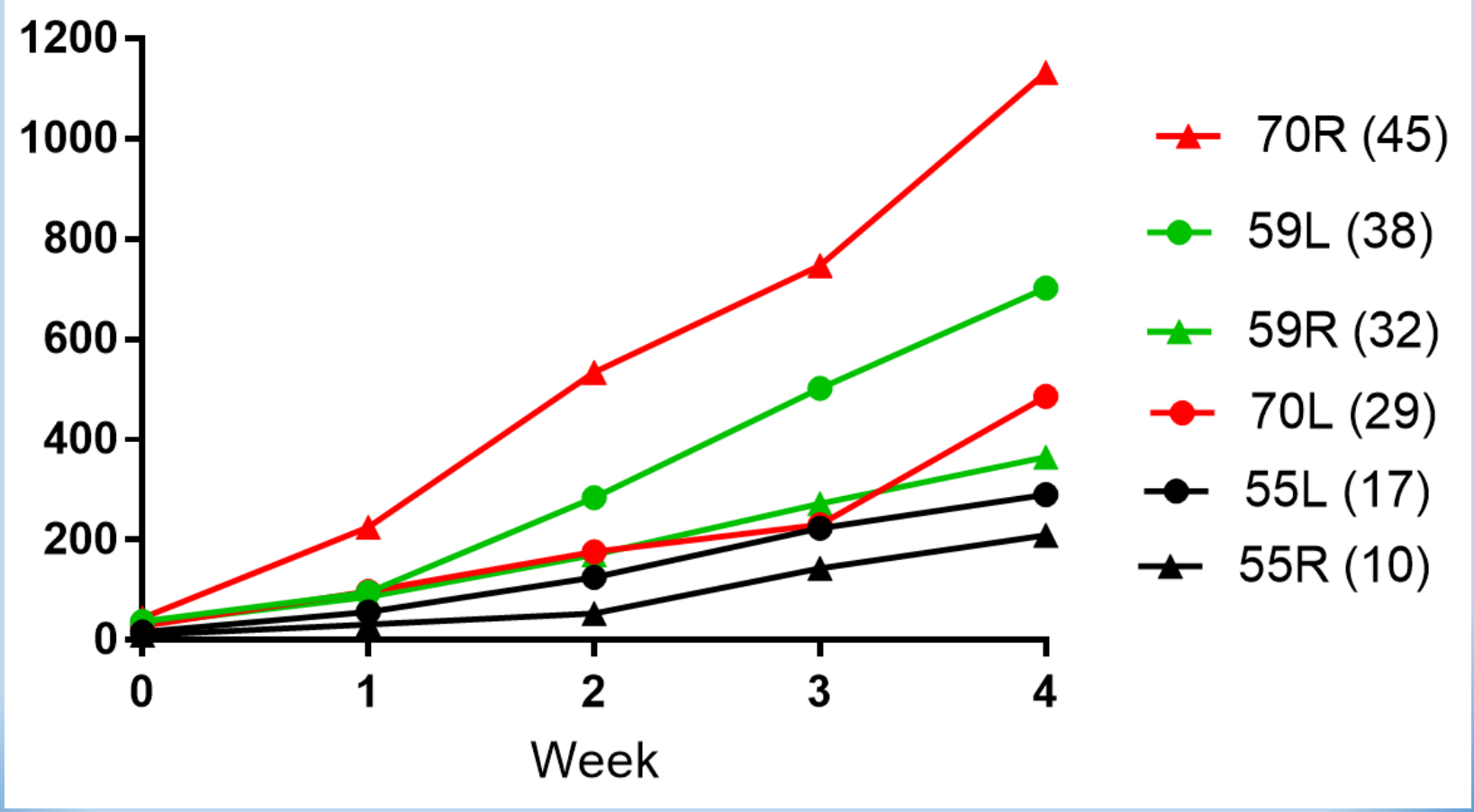
The host mouse may influence the growth of each tumor

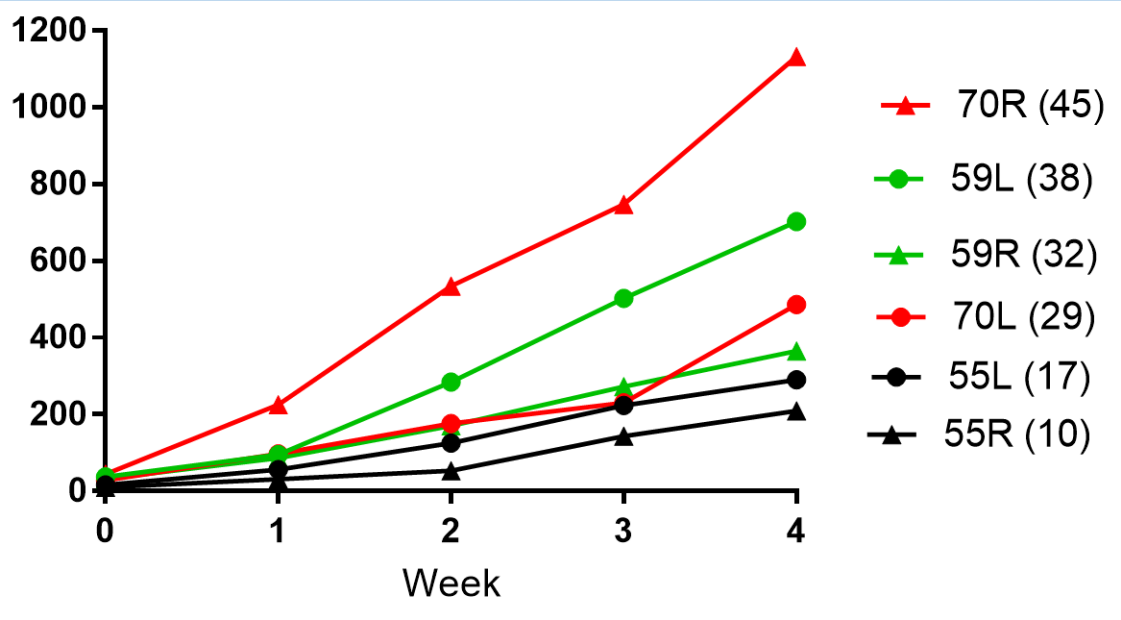
Tumor growth is thus correlated in each mouse

To test independence, I looked to see if the tumors in the same mouse have similar growth over time.

If the growth plots for left and right tumors for an individual mouse are the same, then the tumors should be treated as correlated

Tumor growth for paired tumors. For example, the red lines are for left and right tumors in the same mouse (#70). The values next to the mouse ID number in the legend are the tumor size at week 0.





The growth of the left and right tumors for mouse 55 are almost identical, but the growth response is different for the other two mice.

Is the similarity (or difference) in growth patterns due to the possibility that tumors in the same mouse will grow similarly (so are correlated), or if the initial size of the tumor is the most important factor determining growth.

It appears that the larger starting tumors may grow the fastest and the largest over time, independent of mouse

Experiment

Pipetting cell line onto coverslips

Coverslips get treatment and are put into incubators (except time 0 which is fixed just before treatment). Each coverslip is removed and fixed at specified time point.

Time 0

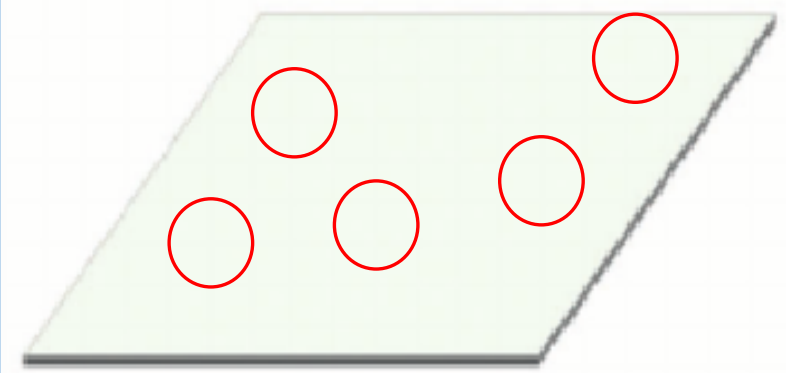
Time 1

Time 2

Time 3

Time 4

For each coverslip, you choose 5 random areas to image



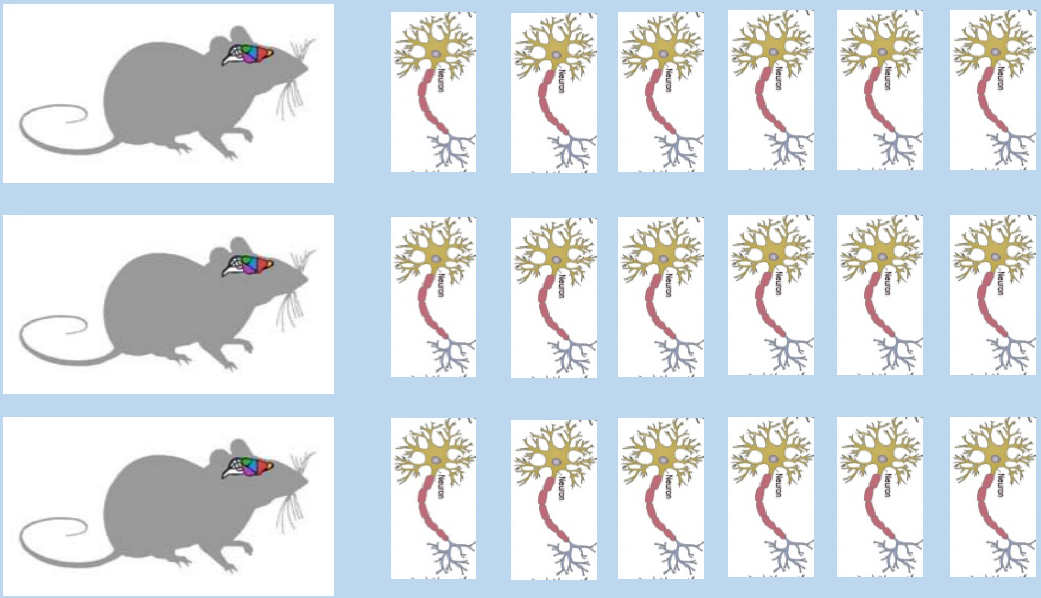
Here are the results

	Time0	Time1	Time2	Time3	Time4
Area1	12.2	22.2	42.1	48.2	41.0
Area2	15.9	14.3	39.6	50.1	60.3
Area3	6.8	19.9	18.5	33.3	51.9
Area4	4.4	11.8	29.8	45.1	40.9
Area5	17.0	16.3	35.7	39.9	51.6
Mean	11.3	16.9	33.1	43.3	49.1

What is the sample size?
What is the unit of analysis?

5 imaged areas per coverslip per time point (n= 5 x 5 = 25)
Mean values per coverslip per time point (n = 1 x 5 = 5)

Are observations independent or dependent?



What is the sample size?
What is the unit of analysis?

6 neurons per mouse (n= 6 x 3 = 18) –or-
Mean neuron response per 3 animals (n = 3)

Are observations independent or dependent?

Are the data independent or dependent and what are the samples sizes for data analyses?

Samples of a cell line pipetted into three wells on the same plate

Samples of a cell line pipetted into one well on three different plates

One mouse and her litter of 5 where mothers are given different diets

A litter of 5 mice

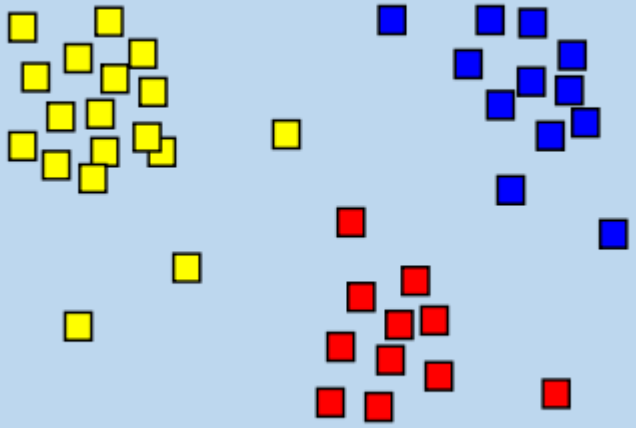
A litter of 5 mice where each pup was exposed to a different diet

Percent staining of marker X measured from 10 random areas imaged and on one coverslip containing a layer of cells from one cell line

10 neurons measured in the brain of one mouse

One neuron measured in each brain of 10 mice

CLUSTERED DATA



Cluster: Different Terms, Different Uses

Clusters: A group of the same or similar elements

Cluster sampling: a method in statistics where clusters of similar elements are sampled

Classrooms

Cluster analysis or clustering: grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters)

Clusters: A group of the same or similar elements

In animal research, it is relatively common to study groups of animals
    Could be considered clusters

Each cluster contains multiple observational units

A key feature of clustered data is that observations within a cluster are usually are more correlated or dependent than observations from different clusters

This must be accounted for in the data analyses

Galbraith S et al. J Neurosci, August 11, 2010; 30(32):10601–10608

A Study of Clustered Data and Approaches to Its Analysis

Sally Galbraith,[1]* James A. Daniel,[2,3]* and Bryce Vissel[2,4]

[1]School of Mathematics and Statistics, University of New South Wales, Sydney 2052, Australia, [2]Neuroscience Program, Garvan Institute of Medical Research, Sydney 2010, Australia, [3]Cell Signalling Unit, Children's Medical Research Institute, Sydney 2145, Australia, and [4]St Vincent's Medical School, University of New South Wales, Sydney 2052, Australia

The Journal of Neuroscience, August 11, 2010 • 30(32):10601–10608 • **10601**

Clustered data are frequently obtained in the neurosciences, but is its analysis is not often discussed explicitly in neuroscience literature

The authors did a study of clustered data and demonstrated that it is critical to take clustering into account in the analysis of data.

Hypothetical experiment to assess whether a drug leads to altered rates of exocytosis from synapses *in vitro*.

Neurons on 10 coverslips are each treated with 10 μg/ml drug

Neurons on 10 coverslips are treated with vehicle

The rate of exocytosis for 10-40 synapses per coverslip is determined

Exocytosis rate data collected on a total of 123 synapses from 10 drug-treated coverslips and on 157 synapses from 10 untreated coverslips.

Galbraith S et al. J Neurosci, August 11, 2010; 30(32):10601–10608

Problem: Correlation in cluster

What happens if conditions within a coverslip, apart from treatment, are more similar than conditions on different coverslips?

This cluster effect must be taken into account to ensure validity of the treatment comparison.

Two questions that researchers should ask are:

(1) Do I have clustered data?

(2) Should I use a method that accounts for clustering?

Galbraith S et al. J Neurosci, August 11, 2010; 30(32):10601–10608

Approaches to deal with clustered data

(A) Ignore clustering

Analyze all data regardless of clustering as independent samples

(B) Reduce clusters to independent observations

Use the mean, median, or proportion of the samples in each cluster

Assumption: separate experiments are independent.

Clusters with more observations could be expected to contribute more information and should be given more weight

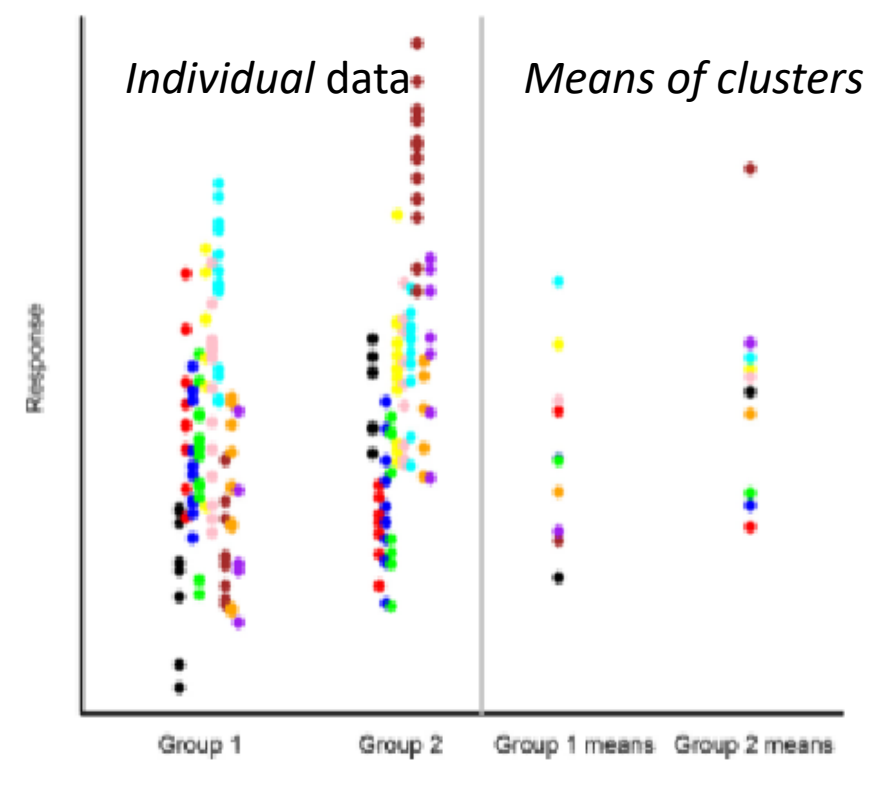
Lose data from individual observations

(C) Use fixed effects regression/ANOVA approaches

(D) Account for clustering in mixed models

Galbraith S et al. J Neurosci, August 11, 2010; 30(32):10601–10608

A sample dataset



This dataset consists of 200 observations in 20 clusters.

The dataset was set up so there is no difference between group 1 and group 2 observations.

Galbraith S et al. J Neurosci, August 11, 2010; 30(32):10601–10608

Results

Table 2. Performance of various tests in analyzing sample data

Method	Estimated difference	95% CI for difference	*p* value for test of no difference
1. Two-sample *t* test (individual observations)	0.517	(0.230, 0.805)	0.0005
2. Wilcoxon (individual observations)	0.486	(0.200, 0.769)	0.0013
3. Two-sample *t* test (means)	0.476	(−0.318, 1.269)	0.2235
4. Wilcoxon (means)	0.425	(−0.387, 1.286)	0.2475
5. LMM	0.476	(−0.322, 1.275)	0.2260
6. GEE	0.476	(−0.228, 1.180)	0.1850
7. Rank-sum test for clustered data (Datta and Satten)			0.2015

LMM: Linear mixed models

GEE: Generalized estimating equations

Galbraith S et al. J Neurosci, August 11, 2010; 30(32):10601–10608

Are there clusters in these data samples?

Samples of a cell line pipetted into three neighboring wells on the same plate
n=1 **(cluster?)**
Samples of a cell line pipetted into one well on three different plates
n=3
One mouse and her litter of 5
n=1 if you experiments on the mother
A litter of 5 mice **(cluster?)**
n=1 or 5 depending on the study
A litter of 5 mice where each mouse was exposed to a different diet
n=5
Percent staining of marker x measures from 10 random areas imaged on one coverslip containing a layer of cells from one cell line **(cluster)**
n=1 (coverslip)
10 neurons measured in the brain of one mouse **(cluster)**
n=1
One neuron measured in each brain of 10 mice
n=10