Innovating Preclinical Statistics Education

Jacob Fiksel

June 17, 2025

Overview of our approach to statistics education at Vertex

Why Invest in Internal Statistics Education?

Education helps us scale

- Leverages the resources we already have
- Identifies where we should work

Decision Makers need to understand results

- Statisticians are rarely the final decision maker in research
- Statistics is only a part of decision making
- 3Rs: Unethical to conduct animal studies if researchers do not know how to properly design & interpret the study (Parker & Browne, 2014)

The Place of Experimental Design and Statistics in the 3Rs (Parker & Browne, 2014)

Creating effective statistics education is challenging

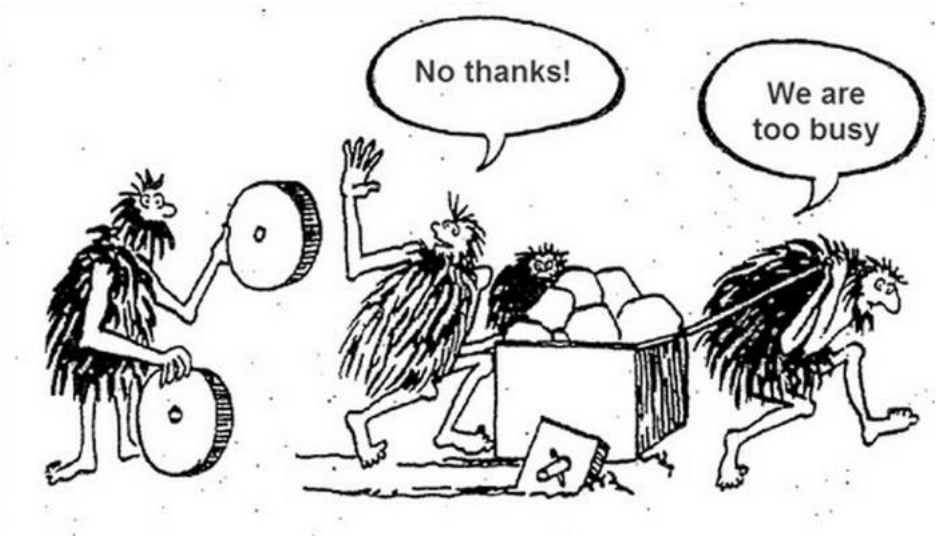
"The correct use and interpretation of statistics requires an attention to detail which seems to tax the patience of working scientist"

If we want content to be consumed:

Engaging: Focused on real problems encountered by research scientists

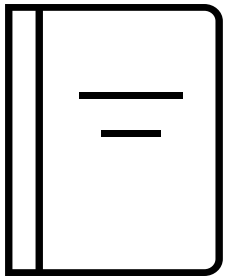
Available on demand: exactly when and how it is needed

Bite-Sized: Short enough to fit into busy schedules: Max 15-20 minutes



We have two modes of internal statistics education at Vertex

Statistics Handbook



- Gives specific recommendations for statistical tools for a range of common questions
- More focused on the what rather than the how & why
- Can be constantly updated & expanded

Online Mini Courses

Vertex U.

- Deep dives into important topics
  - Data visualization
  - Interpretating statistical hypothesis tests
- Developed with internal educational experts at Vertex U
- 1 course per year

Applying these principles to a mini course on statistical hypothesis testing

The goal of the course is to equip pre-clinical researchers with the knowledge and skills to critically evaluate and interpret results from statistical tests.

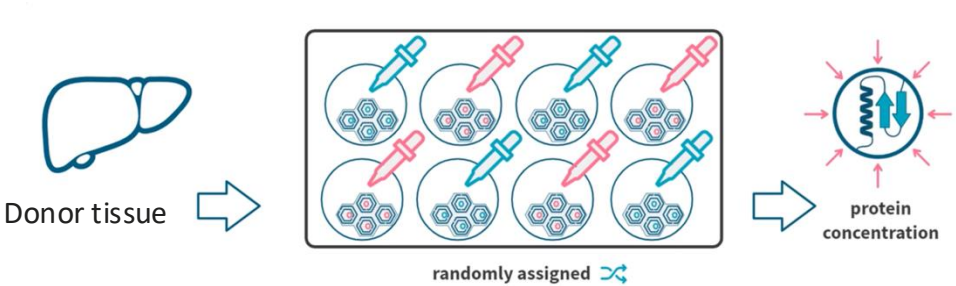
- A low *P* value means there is likely some difference between the two groups
- A high *P* value does not mean there is no difference between the groups you are comparing
- *P* values do not measure importance of effect size

- Statistical significance is an objective procedure that uses P values for decision making
- Useful for reducing the chance of continuing to invest in a treatment with no effect.
- Also not a measure of whether the effect size is biologically relevant

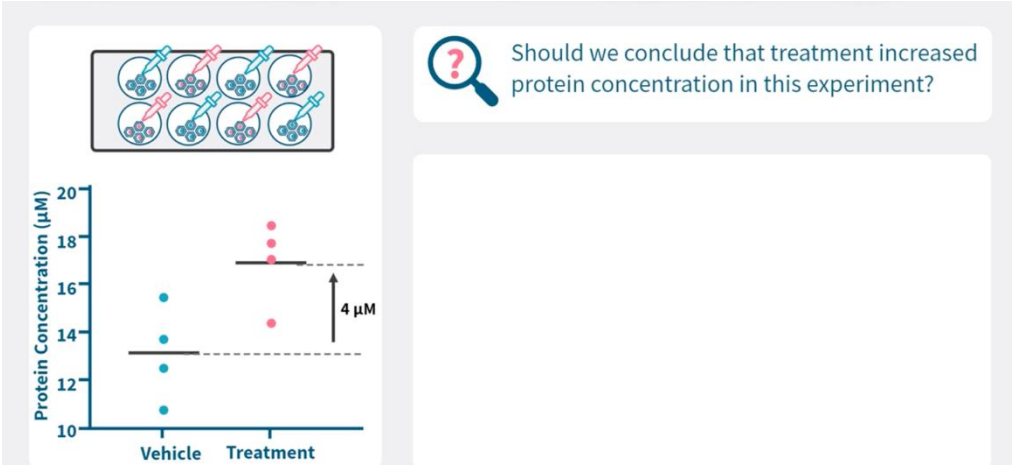
- Confidence intervals inform on the magnitude of the true treatment effect and uncertainty of the observed treatment effect
- Confidence intervals can be used to inform if the difference between the groups is meaningful or important.

Correctly explaining why we use statistical tests is the most important (and challenging) part of the course

- Course is motivated by a **randomized** experiment

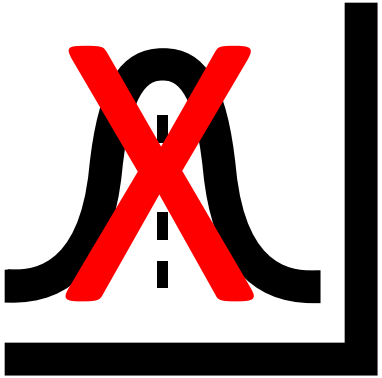


- Tissue slices in the treatment group had a better average outcome than those in the vehicle group, so why do we have to use statistics to be sure of this?
  - Many scientists can't say how *P* values can actually help with decision making

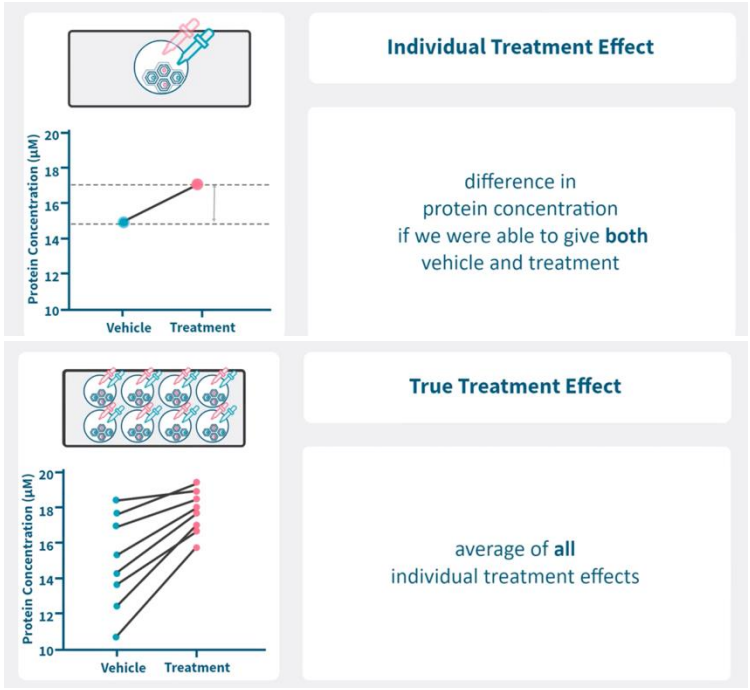


Our biggest innovation was using *potential outcomes* to explain why we use statistics

- Traditional intro to statistics classes use **population based** inference. This doesn't make sense when explaining statistics to preclinical scientists because **preclinical experiments are not done using random sampling from populations**

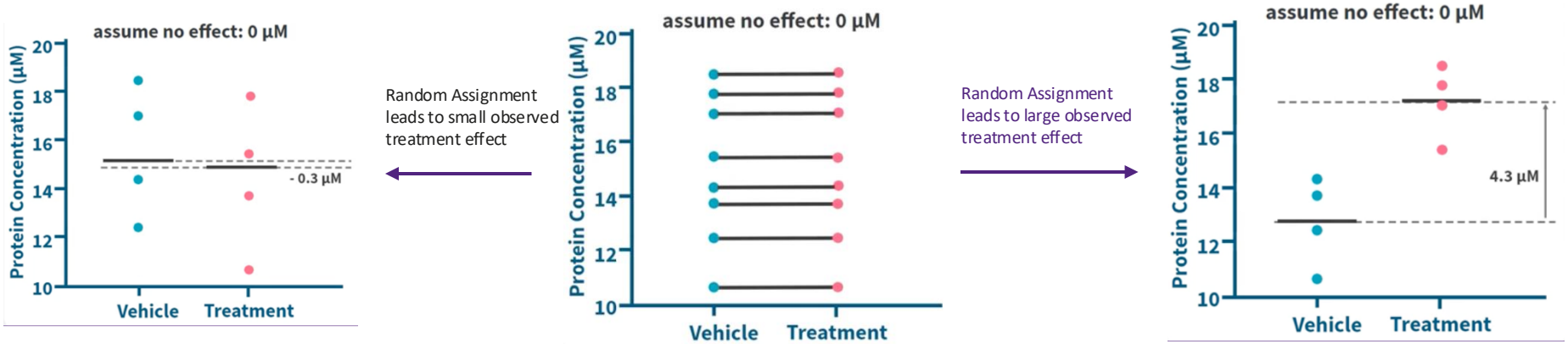


- In preclinical experiments, we need to think about what **could have** happened if we were able to give each sample both the control and treatment. Statistical tests can inform how well a treatment worked **in a specific experiment**, rather than in a population



Potential outcomes allows us to show how observed data can be misleading due to biological variability and random chance, which is why we need statistical tests

Collaborators at Vertex University created animations that enabled us to demonstrate complicated statistical concepts visually, without any formulas or mentioning specific hypothesis tests



Scientists need to know how to interpret *P* values, not define them

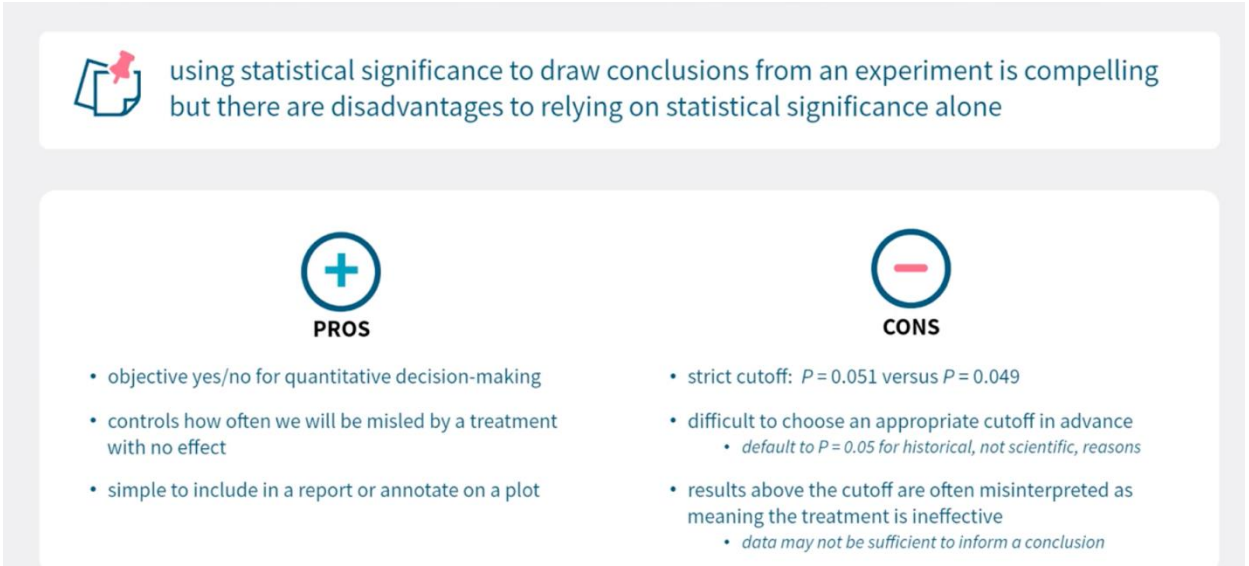
- By starting with showing how observed results can be misleading, we give learners the two concepts they need to understand *P* values:
  1. There is a true treatment effect that we can never know, and we don't want to advance treatments with a true treatment effect of 0
  2. Randomness coming from random assignment and biological variability can change the observed treatment effect

- This allows us to define the *P* value up front without showing normal distributions or formally defining the sampling distribution using the central limit theorem

The P value is the probability that if the true treatment effect were 0, random chance would lead to an observed treatment effect at least as large as the one in the experiment we performed.

- We don't dwell on this definition or how to calculate a *P* value, but instead emphasize 2 key points:
  1. A low *P* value means we can rule out the treatment having no effect, but it does not guarantee the effect is biologically meaningful
  2. A high *P* value means we cannot rule out the treatment having no effect, especially with small sample sizes

We teach decision makers how statistical significance can help (or hurt) decision making, rather than encouraging mindless use or banning it



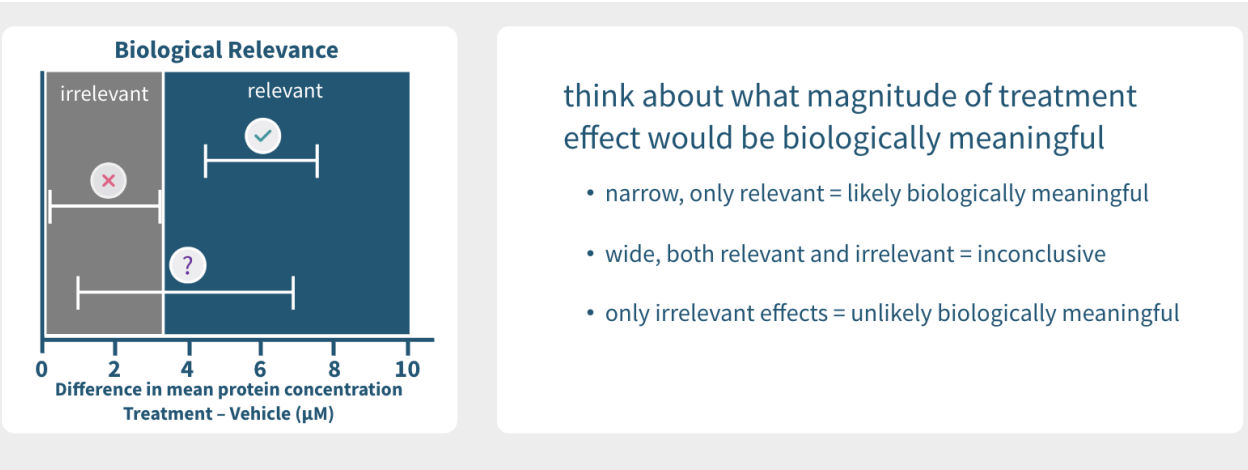
We do not commit the “statisticians fallacy” by insisting that statistical significance answers the wrong question, but instead help learners understand what questions & decisions can be informed with *P* values and statistical significance

https://daniellakens.blogspot.com/2017/11/the-statisticians-fallacy.html

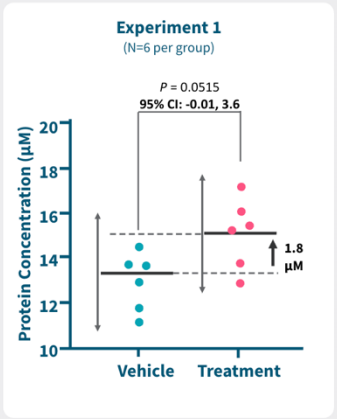
We believe connecting confidence intervals to decision making will increase their use amongst scientists and decision makers

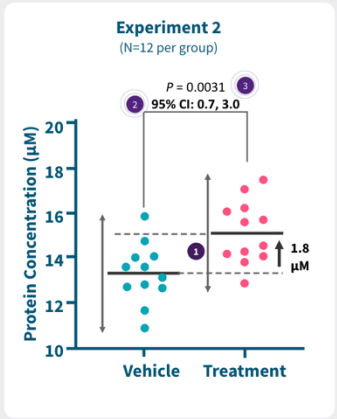
- We set learner up to want to use confidence intervals:
  - We have already emphasized that the observed treatment effect is not the true treatment effect AND
  - The *P* value only helps answer the limited question of whether the true treatment effect is greater than 0

- No need to go into what confidence means. They just need to be aware that the confidence interval gives a **range** of values that **likely** contains the true treatment effect



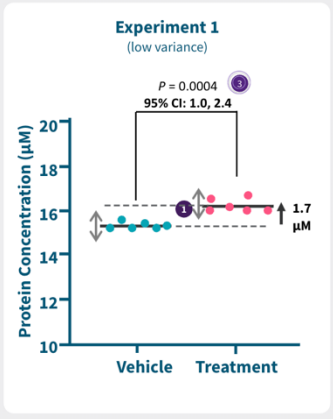
We use visual examples rather than equations to explain the impact of sample size and variance on statistical tests

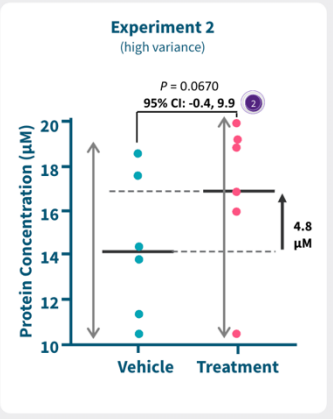




increasing the sample size:

results in a substantially lower *P* value and narrower confidence interval for Experiment 2, even though the difference in means and biological variability is the same in both experiments





low variance:

results in a substantially lower *P* value and narrower confidence interval for Experiment 1, despite the difference in means being smaller than Experiment 2

This course is innovative because:

1. It uses randomization and potential outcomes as the basis of random chance in preclinical experiments

2. We collaborated with educational experts who created visuals that explained statistical concepts more intuitively than formulas
    - Huge contributor to explaining the most difficult concepts in introductory statistics in 15 minutes

3. Maintained focus on connecting statistical concepts to scientific questions and decision making, rather than teaching statistical concepts for the sake of it

Preclinical statistics education moving forward

Statisticians need to educate on the fundamentals to maximize our impact

- Vast majority of preclinical data can be analyzed using simple statistical tests, and the challenge is helping scientists avoid pitfalls and properly interpret the data
  - Proper visualization of data, especially with small sample sizes
  - Ensuring no pseudoreplication
  - Understanding how output from a statistical test connects to the scientific question and decision making process

- In our experience, scientists who exhibit statistical thinking are more likely to know when they need to bring in a statistician

- Even when situations require bespoke statistical tests, scientists and leadership who understand the fundamentals of statistics will more clearly see our impact

Just like science, statistics education will require experimentation and collaboration

- Statisticians should aim for excellence in educating, and this means constantly evaluating our materials & methods
  - Poor/mediocre teaching is worse than doing nothing at all

- Collaboration with colleagues who are education experts greatly improved our course
  - Not only for the animations & professional look of the course, but also for giving feedback as non-statisticians on if the content made sense from a learner perspective. This was a big reason why we had little focus on calculating a *P* value

- Collaboration between statisticians in industry **and** those who focus on education in academia will allow the preclinical statistics community to more quickly find the most effective methods for statistics education on a variety of topics

- While we believe our course on statistical testing is innovative and will be impactful, it will take time to see the results and is just one way of building a course on this topic
  - We are not beholden to the format or content of our material, and will re-evaluate all content every 1-2 years based off interactions with our collaborators and formal feedback from learners after they complete the course

Acknowledgements

- Ben Chittick (head of the preclinical statistics center of excellence)
- Bridget Sullivan, Jacob Zolotovsky, and Carolyn Farricy (Vertex University collaborators)
- All other members of the preclinical statistics center of excellence
- Angela Yen (head of Data & Computational Sciences)
- Mark Bunnage (head of research)
- Sarah Luchansky & Josh Bittker for their feedback