2016 Biomedical Research Integrity Program
Integrity from the Inside Out

Topic for Discussion: Research Misconduct/P-Hacking

Cross-Cutting Themes: data management, scientific integrity

Overview questions to group– What methods do you use for your analysis in publication? Do you make judgment calls about what data to include, or which analyses to use based on possible outcomes? Do you use P-values in your research? Have you encountered the challenges laid out by the "p-hacking" panel?

Getting Started:

- What is a good definition of a P-value and what does a P- value of <0.05 mean with regards to research?
- What follow up steps with your data, if any, would you take if the P-value was <0.05?
- In what ways might the misuse of P-values and data use constitute a lack of research integrity?
- What are the take home points from either the panel discussion on P-values or the reading that caught your interest or you would like to discuss? What else?

Use the 4 R's to structure the discussion through a particular case or issue.

Process for Thinking through Difficult Ethical Dilemmas

Recognition*: What are the issues being raised? What is the underlying ethical concern? How does this issue impact me?*

Reasoning*: What values are at stake? Are there competing points of view? What are the potential benefits and harms of different actions? Are there any rules or guidelines that can help?*

Responsibility*: What are my responsibilities? Do others have responsibilities also?*

Response*: What should I do – and why?*

Assigned Reading:

RP Galr, M-J Zhang, What is the P-value anyway? *Bone Marrow Transplantation* advance online publication 11 July 2016; doi: 10.1038/bmt.2016.184;

R. Nunzio, Scientific method: Statistical errors, *Nature* 2014; 506: 150-2

Highly Recommended:

John Oliver on Scientific Credibility: https://www.youtube.com/watch?v=0Rnq1NpHdmw

Topic for Discussion: Research Misconduct/P-Hacking

Case Study 1: **Promising New Drug or Data Dredging?**

A biotech company discovered a promising new drug, interferon gamma-1b, for those with fatal lung disease. 330 patients were randomly assigned to get either interferon gamma-1b or placebo injections. Disease progression or death occurred in 46 percent of those on the drug and 52 percent of those on placebo. That was not a significant difference, statistically speaking. When only survival was considered, however, the drug looked better: 10 percent of people getting the drug died, compared with 17 percent of those on placebo. However, that difference wasn't "statistically significant," either.

Specifically, the so-called P value — a mathematical measure of the strength of the evidence that there's a true difference between a treatment and placebo — was 0.08. Technically, the study was a bust, although the results leaned toward a benefit from interferon gamma-1b.

The PI on the study asked his stat people to see if there was a group of patients in which the results were better. They found that people with mild to moderate cases of the disease had a dramatic difference in survival. Only 5 percent of those taking the drug died, compared with 16 percent of those on placebo. The P value was 0.004 — highly significant.

When the study was published in the New England Journal of Medicine in January 2004, the authors wrote that "a clinically significant survival benefit could not be ruled out."

Later, another trial— 826 patients at 81 hospitals — was run in order to maximize the chance of getting clear-cut results. It enrolled only people with mild to moderate lung damage, the subgroup whose success was touted in the press release. A little more than a year into the study, more people on the drug had died (15 percent) than people on placebo (13 percent).

Questions

Is it a problem that the mild-to-moderate subgroup wasn't one the researchers said they would analyze when they set up the study? Why or why not?

If the P value did not prove "a clinically significant survival rate" in the mild to moderate cases, what other factors might be causing the effect?

2016 Biomedical Research Integrity Program
Integrity from the Inside Out

Topic for Discussion: Research Misconduct/P-Hacking

Case Study 2: Of Mice and P Values

Alice is a postdoc studying the effect of a specific environmental toxin on levels of oxygen in mouse blood. Her study is to run for three months and several hundred mice will be used. After one month she notices that the oxygen levels from one group of mice getting the toxin are quite different from the others who are also getting the toxin. She has lots of mice in the experiment so decides not to include data from this group thinking that they may be diseased in some way. One of the animal techs agrees with her.

After another month of comparing blood in the mice subjected to the toxin and to those not subjected to it, she is sure she sees some important changes that have never been noticed before. One of the other postdocs suggest she run a statistical test that will give her a P value, just to be sure[Office1].

When she plugs in her data, she's excited to find that the P value is <0.04. The other postdoc assures her this really means something.

She talks to her PI who suggests she stop the study. After all, the mice are expensive and he's not sure the animal use committee would like it if the study continued when they already have enough data to maybe move forward on a paper and next stages of the experiment. He also hints that this will probably be a very important paper for her. With a P value of <0.04, the journals will be really interested in her work.

Alice stops the study and begins writing up the results.

Questions

Given that ending the experiment will save money and also possibly save Alice from having a conflict with the animal use committee what are some arguments for continuing the experiment?

Is leaving out the data from one set of mice research misconduct?

What decisions did Alice make along the way that compromised the integrity of her research?

Resources and Additional References

- American Statistical Association statement on p-values: https://www.amstat.org/newsroom/pressreleases/P-ValueStatement.pdf
- A Dirty Dozen: Twelve P-Value Misconceptions, http://www.perfendo.org/docs/BayesProbability/twelvePvaluemisconceptions.pdf