

## **NIH Rigor and Reproducibility Training Module 4: Sample Size, Outliers, and Exclusion Criteria**

### **Potential Discussion Points and Questions:**

#### Starting Points:

- Sample: here, a sample is defined as a single value or observation from the larger set of values
- Sample size: the optimal number of samples that should be used to reach sufficient statistical power; also referred to as 'n'
- Outliers: an observation that lies an abnormal distance, typically +/- 3 standard deviations, from other values in a random sample from a group of results<sup>1</sup>
- Exclusion criteria: standards set out before a study or review to determine whether a sample should be included or excluded from the study or analysis<sup>2</sup>
- Characterization of "normal" for a specific experiment is an important component to identifying outliers and determining exclusion criteria

#### Lead-in Questions:

- Do you have a standard approach to determining the appropriate sample size and setting criteria for outliers – how you determine the numbers that go into your power analysis?
- How do you know what "normal" is if you don't know the result? Can you do this initially? Will determination of the best statistical method and approach be useful in defining normal?

#### Follow-up Questions:

##### *Interpersonal Dynamics*

- Did power dynamics in the lab play a role and/or contribute to the situation? Do you think that there is a bias to believe an experienced postdoc who contributed a lot to the lab over a graduate student?
- Do you think Harry was exhibiting racial bias when he assumed that Robin was struggling and suggested that "some students come in and can't hack it"?
- Do you think Harry would have taken a similar approach if Robin was a male graduate student?
- Was there a more appropriate and effective approach that Harry could have taken when Robin was attempting to replicate Donna's results?
- Do you think most PIs would take the time to review the lab notebooks themselves to determine what may be causing the discrepancy in the results?
- Is it realistic to think that most PIs would admit they provided inadequate guidance?
- While accepting some responsibility for the situation, were you frustrated that Harry did not apologize to Robin for his behavior?

##### *Lab Management*

- Can you relate to this situation – not being able to generate similar results, whether from unpublished data in your own lab or a published paper?
- Have you ever tried to replicate someone's experimental approach and discovered that information was missing in their lab notebook? Did you feel as though you needed a "Rosetta Stone" to decrypt their handwriting/abbreviations?
- Do you maintain a thorough laboratory record? If so, what methods do you follow to ensure

that your lab notebook is comprehensive?

- Do you think an electronic lab notebook would have helped identify the issue(s) faster? What characteristics would the electronic lab notebook need to have?

#### *Statistical Methods and Issues*

- Have you ever had data that was “close” to significance? If so, what did you do? How did you interpret these results?
- Would Dr. Fielding (Harry) have suggested adding a few more samples and trying a different statistical test if they had initially defined their sample size and exclusion criteria, and identified the most appropriate statistical approach?
- Jamal told Robin to drop outliers above a certain value, as it is outside the physiologic range. Do you think this should have been considered further when they established their exclusion criteria? Do you think they actually developed exclusion criteria, or just considered that point as valid (potentially, without confirming) and made it their sole criteria for determining outliers?

#### *Sex as a Biological Variable*

- One of the fundamental variables in preclinical biomedical research is sex: whether a cell, tissue, or animal is female or male<sup>3</sup>. Do you generally consider sex as a variable when designing experiments?
- Have you or someone you know only used male mice in an experiment as a way of avoiding the “sex issue?” Do you think this is appropriate? Does it depend on the type of experiment being done?
- Can an experiment be considered rigorous if sex is not considered?
- A commonly used example advocating for the consideration of sex as a biological variable in research is the zolpidem (Ambien) dosage that was amended in 2013. The drug was found to affect men and women differently, which resulted in a decrease in the recommended dosage for women<sup>4</sup>. Would this have occurred if sex was considered in the preclinical and clinical experiments?

---

<sup>1</sup>National Institute of Standards and Technology, Engineering Statistics Handbook (7.1.6)

<http://www.itl.nist.gov/div898/handbook/prc/section1/prc16.htm>

<sup>2</sup> Modified from the Agency for Healthcare Research and Quality Glossary of Terms

<http://effectivehealthcare.ahrq.gov/index.cfm/glossary-of-terms/?pageaction=showterm&termid=105>

<sup>3</sup> NIH Office of Research on Women’s Health <http://orwh.od.nih.gov/news/scientificseminars.asp>

<sup>4</sup> <http://www.fda.gov/Drugs/DrugSafety/ucm352085.htm>; <http://abcnews.go.com/Health/fda-recommends-slashing-sleeping-pill-dosage-half-women/story?id=18182165>