This video series was made possible through a generous grant from the National Institute of General Medicines (Grant Number: 5 R25 GM133017-03), awarded to Cohen Veterans Bioscience (Principal Investigator: Chantelle Ferland-Beckham, PhD)

Module 2, Video 12: Study Design Toolbox and Reporting Considerations for Incorporating Sex as a Biological Variable (Part I)

Concept introduction: Preclinical research should recapitulate constructs of human diseases in order to study underlying processes and develop potential therapeutic strategies. But despite documented wide-ranging differences between males and females, few preclinical studies use both sexes, and subgroup analyses by sex are rarely reported even if both sexes are included [1]. Thus, incorporating both sexes remains a significant barrier to improving the utility of preclinical research. But designing sex inclusive studies for the first time can be a daunting task. In these two videos, we will discuss how to appropriately power a study, how to best design studies that incorporate males and females into new fields of study or how to add females to existing work in males (or vice versa), and how to talk about/interpret your results.

The incorporation of sex as a biological variable DOES NOT mean that you have to study sex differences. Nor does it mean you must drastically increase your samples sizes [2]. This is important as they are two of the most common arguments against incorporating females into preclinical studies. First, pursuing any observed sex differences should be made on a study-by-study basis depending on your lab's interests and the outcome measures. Secondly, sample size determinations should ALWAYS be based on a power analysis and ideally on preliminary pilot data or published effect sizes. If no data is available, power calculations should be based on a best estimate of the minimum desired effect size that makes sense. There are a number of publicly available programs for conducting power calculations [3, 4].

In most cases, the feared "doubled n" is not necessary to incorporate both males and females into the same study design. When a sex difference is highly likely, a larger n may be required to sufficiently power the experiment to detect even small differences between males and females with high probability. We will discuss in more detail how to design studies for both scenarios later on in this video.

However, what if you want to simply begin incorporating both males and females into your studies? Perhaps this is a new field of study for your lab or an area in which you don't know if a sex difference exists? In this case, you would simply design a study in a similar manner as you would for only males. The only difference would be that you would now include an equal number of females into your study groups. This would likely require a larger n number, but as explained in Annaliese Beery's 2018 paper, not on the order in which you might expect [2]. For example, if you previously compared two groups—control and experimental—each with a group size of 8, you would now increase your n size slightly to, for example, 12 animals per group. The exact group size is based on a power analysis. Once you determine your group size, then you would evenly split this number between males and females in both groups.

This video series was made possible through a generous grant from the National Institute of General Medicines (Grant Number: 5 R25 GM133017-03), awarded to Cohen Veterans Bioscience (Principal Investigator: Chantelle Ferland-Beckham, PhD)

In this experiment, females and males would be included without consideration of their hormone levels. The results would be analyzed by 2-way ANOVA, factoring for both the treatment AND sex.

If your study is statistically powered to detect a sex difference and no statistical effect of sex is observed in your results, then you can reasonably conclude that sex does not seem to influence your outcome measure under the experimental conditions. It also means that there is **no scientific justification** for excluding females from further similar studies. You should continue to include males and females equally as a blocking factor. This can be particularly advantageous for some genetic models where female animals have been traditionally discarded. If you are including sex but are not interested in studying sex differences or powering the study to detect an interaction between treatment and sex, then you can include sex as a blocking factor rather than a main effect during analysis. This information should be reported but does not have to be focused on.

What happens if your results indicate that there **IS** a potential statistical effect of sex on your outcome measure? At minimum, you should report this sex difference even if you don't plan to pursue these scientific directions. But p values aren't the only benchmark by which to judge the presence of sex differences in your results. The effect size should also be reported. If the effect size is large for a between sex comparison, but the statistical test does not reach significance, the sample size you chose was not large enough. It is also important to closely examine the data separately with other statistical tests to see if a bimodal distribution or increased variability by sex is observed [5]. Even if the result is not statistically significant, this trend should also be reported.

While studies are ideally designed to look specifically for sex differences from the start and run both sexes simultaneously [3, 6], what happens if your lab has an existing body of work in only males and you now want to expand the findings to females? There are some important guidelines for how to add females or males and relate it back to your existing work in a single sex. As covered in Video 7, many environmental factors can affect experimental outcomes. Because potential confounding factors might differ between two separate experiments, you can't simply run an experiment with the missing sex and compare it back to your previous work in the other sex. This is because you can't statistically account for environmental factors that might have changed between the two experiments. A right approach is to run a "validation" subgroup. For example, say you have historical data of an effect between control and experimental groups in males that you want to explore in females. Your new experiment would include experimental and control groups of females that mimics the experimental conditions of your previous work in males PLUS a validation subgroup of males. You would then determine whether the data from the validation group matches your historical data. If the two groups of

This video series was made possible through a generous grant from the National Institute of General Medicines (Grant Number: 5 R25 GM133017-03), awarded to Cohen Veterans Bioscience (Principal Investigator: Chantelle Ferland-Beckham, PhD)

males are the same statistically, you can then compare your new work in females back to your previous work in males with confidence despite potential environmental differences.

Once you have data from both sexes, the most important final consideration is how to interpret and report the results in both sexes. Data should always be reported in accordance with the ARRIVE 2.0 Guidelines [7]. When possible, data should also be considered by all 4 points: male and female, control and experimental. This is in contrast to a side-by-side comparison of control and experimental animals separately in males and females.

When you statistically compare males and females directly, and see a difference, this difference can be attributed to the sex of the animal. Another method to determine the presence of sex effects is to normalize all your data to one group (e.g., control males), allowing you to easily see sex differences, particularly in the control groups.

When the data must be analyzed separately in males and females, then careful language should be used to report the results. Because males and females are not directly compared statistically, we can no longer say this is a sex effect. Instead you can say "we saw X effect in one sex and not the other" or "X is a male-/female-specific effect." Put another way, discuss each sex as if the other sex was in a different paper. This is particularly relevant when you run an experiment in one sex and then try to relate it back to previous work in the other sex. Again, this is best done with a "validation" subgroup so previous results can be replicated to strengthen these comparisons (for example, to control for environmental changes). Even when no sex effect is found—whether statistically or even a bimodal distribution or trend—the data should be presented disaggregated by sex somewhere in the paper. This way data and outcomes can be best judged and used by the scientific community.

In this video, we introduced how to begin to incorporate both males and females into your experiments, whether this is a new area of study or whether you want to compare new data in females to historical data in males. In the second half of this video, we will take a more in depth look at how to design an experiment when studying sex differences is your primary intention.

Combined Part I & II References

1. Zakiniaeiz, Y., et al., *Balance of the Sexes: Addressing Sex Differences in Preclinical Research.* Yale J Biol Med, 2016. **89**(2): p. 255-9. DOI
2. Beery, A.K., *Inclusion of females does not increase variability in rodent research studies.* Current opinion in behavioral sciences, 2018. **23**: p. 143-149. DOI
3. Percie du Sert, N., et al., *The Experimental Design Assistant.* PLoS biology, 2017. **15**(9): p. e2003779-e2003779. DOI

This video series was made possible through a generous grant from the National Institute of General Medicines (Grant Number: 5 R25 GM133017-03), awarded to Cohen Veterans Bioscience (Principal Investigator: Chantelle Ferland-Beckham, PhD)

4. van Wilgenburg, H., P.G. van Schaick Zillesen, and I. Krulichova, *Sample Power and ExpDesign: tools for improving design of animal experiments.* Lab Anim (NY), 2003. **32**(3): p. 39-43. DOI
5. Pfister, R., et al., *Good things peak in pairs: a note on the bimodality coefficient.* Frontiers in psychology, 2013. **4**: p. 700-700. DOI
6. Becker, J.B., et al., *Strategies and methods for research on sex differences in brain and behavior.* Endocrinology, 2005. **146**(4): p. 1650-73. DOI
Developmental effects of sex hormones.
7. Percie du Sert, N., et al., *The ARRIVE guidelines 2.0: Updated guidelines for reporting animal research.* PLOS Biology, 2020. **18**(7): p. e3000410. DOI
8. McCarthy, M.M., et al., *Sex differences in the brain: the not so inconvenient truth.* J Neurosci, 2012. **32**(7): p. 2241-7. DOI
9. Arnold, A.P. and X. Chen, *What does the "four core genotypes" mouse model tell us about sex differences in the brain and other tissues?* Front Neuroendocrinol, 2009. **30**(1): p. 1-9. DOI
Review of the Four-Core Genotype model.
10. Büdefeld, T., et al., *Sex differences in brain developing in the presence or absence of gonads.* Developmental neurobiology, 2008. **68**(7): p. 981-995. DOI
Review of the Four-Core Genotype model.
11. Becker, J.B. and G.F. Koob, *Sex Differences in Animal Models: Focus on Addiction.* Pharmacol Rev, 2016. **68**(2): p. 242-63. DOI
12. Perry, A.N., et al., *The Roles of Dopamine and α1-Adrenergic Receptors in Cocaine Preferences in Female and Male Rats.* Neuropsychopharmacology, 2015. **40**(12): p. 2696-2704. DOI
13. Perry, A.N., C. Westenbroek, and J.B. Becker, *The development of a preference for cocaine over food identifies individual rats with addiction-like behaviors.* PLoS One, 2013. **8**(11): p. e79465. DOI