

SCIENCE FACT OR SCIENCE FICTION: IF I INCLUDE FEMALE ANIMALS, DO I NEED TO DOUBLE MY SAMPLE SIZE?

ACCOUNTING FOR SEX IN ANIMAL RESEARCH

Ensuring Rigour and Reproducibility

Reproducibility is a key element required to advance health research. Study design is a primary factor driving poor reproducibility, yet animal research does not always use efficient designs.¹ The consequences include economic loss² and potentially wasting millions of animal lives on low quality work.³ When studies are well-designed and properly analyzed, they contribute meaningfully toward improving human health and thus do not waste animal lives. Considering sex as a component of rigour and reproducibility in animal research is one of many factors that must be addressed.

Sample Size Does not Need to be Doubled to Incorporate Both Sexes

On the surface it may appear that including male and female animals in a study requires doubling the number of animals to conduct the experiment in each sex. However, this is not always the case, as more efficient and powerful experimental designs that incorporate both sexes, while maintaining powerful control over variance, can be used. While sample sizes may need to be slightly increased to account for the extra parameter being estimated, they do not need to be doubled.⁴

Analyzing Experiments

The information generated by an experiment can be represented by the sum of all sources of variation based on the sample size (N) -1 degrees of freedom.⁵ Note that the degrees of freedom for each factor are always equal to the number of levels in that factor minus 1. Thus, any experiment can be represented by a version of the following equation (factors and interactions can be added or removed as necessary):

$$\text{Outcome} = \text{Factor 1} + \text{Factor 2} + \text{Factor 1} * \text{Factor 2} + \text{Error}$$

For our case, let's say we have two factors: 'Treatment' and 'Sex'. 'Treatment' is the factor we are applying to the animals. 'Sex' is whether the animals are male or female. 'Treatment*Sex' is the interaction between 'Sex' and

'Treatment' (i.e. testing to see if the treatment differentially affects males vs. females). Testing the interaction between two factors allows us to assess how each level of a factor varies in relation to the levels of the other factor. The 'Error' component represents all of the unexplained variation left over after accounting for variation due to the other factors and is used to calculate the standard error for treatment comparisons.⁵ As a rule of thumb, higher degrees of freedom in the 'Error' term mean that there is more statistical power to detect an effect, though this has diminishing returns as sample sizes increases. It has been suggested that error degrees of freedom in the 'Error' term can reasonably be between 10 and 20.⁵

A SIMPLIFIED COMPARISON: 'One-sex-at-a-time' Design vs. Factorial Design

'One-sex-at-a-time' Design: 10 treated animals vs. 10 controls

Assuming individual animals are experimental units, the total sample size is 20 and the total number of treatments is 2. Thus, 'Treatment' is a factor with 2 levels, and the following equation can be used to represent the experiment, with the degrees of freedom noted below each element of the equation:

$$\text{Outcome} = \text{Treatment} + \text{Error}$$

(20-1) (2-1) (18)

Note that because we only have one sex, there is no term for 'Sex' or for an interaction between treatment and sex ('Treatment*Sex') in this experiment. Thus, this experimental design answers the question: does the outcome differ between treated and control animals? In order to answer a question about the effects of treatment on the opposite sex, the same experiment has to be run again, effectively doubling the number of animals, to 40. However, this design would not provide a valid way to examine if the treatment differentially affects the sexes, because there is no direct comparison of the 'Sex' factor.

2x2 Factorial Design:

Each sex is represented within the two treatment groups

A factorial design is a simple, yet powerful way to incorporate both sexes into a single experiment. Factorial designs incorporate at least two factors, with at least two levels each, arranged such that the experimental units incorporate all combinations.⁶

Using the same example as above, the total sample size is 20 animals and the number of treatments is 2. However, there are now 2 sexes of animals (i.e. 'Sex' is also a factor with 2 levels). In this example, the 'Treatment' can be analyzed such that the variation due to sex is accounted for when testing for an effect. This reduces the amount of unexplained variation in the outcome (i.e. the degrees of freedom in the 'Error' element of the equation), and allows the researcher to look at the effect of both sex and treatment on any desired outcome variable. Most importantly however, the researcher can test this design with a 2-way ANOVA to assess interactions between these two factors.⁷

Thus, the following equation can be used to represent the experiment, with the degrees of freedom noted below each part:

Outcome = Treatment + Sex + Treatment*Sex + Error

(20-1) (2-1) (2-1) (1*1) (16)

	Treatment (n=10)	Control (n=10)
Female (n=10)	n=5	n=5
Male (n=10)	n=5	n=5

Compare the latter equation with the equation from the first design. First, the degrees of freedom in the 'Error' term have diminished slightly (from 18 to 16), meaning there will be less power to detect effects of the treatment. However, the decrease is slight, and the value for the 'Error' degrees of freedom is still well within the acceptable 10-20 range. Note that if a researcher wanted to maintain the error degrees of freedom of at least 18 because they needed that much power to detect effects of treatment, only four additional animals would be required (this value ensures a balanced design—adding one male and one female to each group).

Second, this design allows us to uncover much more information overall. **Specifically, we can now answer the following three questions:**

- 1) Does the outcome variable differ between treated and control animals?
- 2) Does the outcome variable differ between males and females? And most importantly,
- 3) Does the treatment have the same effect on males that it does on females?

Using a factorial design allows the researcher to statistically investigate these differences: something not possible when the assessment of sex effects is spread out over two experiments. Valid inferences can only be made if the two sexes are compared directly, not via a series of independent tests.⁴ Additionally, because factorial designs essentially combine experiments (e.g. one to look at drug treatment in males, and another in females), they require fewer animals than would be needed for running two separate studies.^{1,7} Thus, in this example, all else being equal, only 20-24 animals are required in a factorial design, rather than 40 in a one-sex-at-a-time designs.

Factorial designs can test for a sex difference using the same number of animals as classical single sex studies.

CONCLUSION

The example above provides a proof-of-concept, demonstrating how both sexes can be included efficiently and effectively into a simple experimental design. While many experiments cannot be arranged in such a neat manner, the principles explained here will still hold. A factorial design will always be more efficient compared to running separate experiments, both in terms of number of animals used and information gained. In reality the number of animals saved using the factorial design will depend on several factors, including the initial sample size, the number of treatment levels, the estimated effect size, the expected inter-subject variance, and the complexity of the analysis (i.e. the number of other factors included in the model and their relationship to each other).

REFERENCES

1. Kilkenny, C., et al. (2009). Survey of the quality of experimental design, statistical analysis and reporting of research using animals. PLoS ONE
2. Freedman L.P., Cockburn I.M. & Simcoe T.S. (2015). The Economics of Reproducibility in Preclinical Research. PLoS Biol
3. Taylor K., Gordon N., Langley G. & Higgins W. (2008). Estimates for worldwide laboratory animal use in 2005. Atla-Alternatives to Laboratory Animals
4. Lazic, S. (2016). Experimental design for laboratory biologists: maximising information and improving reproducibility. Cambridge University Press.
5. Mead, R. (1988). The Design of Experiments. Cambridge University Press.
6. Grafen, A. & Hails, R. (2002). Modern statistics for the life sciences: learn how to analyse your experiments. Oxford University Press.
7. Miller, L., et al. (2016). Considering sex as a biological variable in preclinical research. The FASEB Journal