NSW GOVERNMENT | Department of Primary Industries

ANIMAL RESEARCH REVIEW PANEL

Statistics Animal Rehoming Webinar - Q&A

9 August 2022

A core challenge for researchers and reviewers alike is ensuring high quality study design with as few animals used as possible. The use of power calculations and A Priori Statistical Plans is, however, rare in the world of animal research. There is a need to engage with fundamental statistics in the design and reporting of studies. How prepared are we, as researchers and animal ethics committee members, to embrace this? During this webinar we heard from statisticians and researchers about some of the evidence for the need for change and easy first steps we can adopt during committee and investigator conversations.

How can we overcome the omission of null results and publish them when many journals do not accept them?
One way we can overcome the omission of null results is by using a publishing format called registered reports. In registered reports your introduction and study plan are peer reviewed, and then you are given 'in principle acceptance', this means that you can publish your results regardless of whether your results are statistically significant or not (as long as you have not deviated from your study plan without justification). Another benefit of registered reports is that you get input via peer review before you begin your study, meaning that peer review feedback can help improve your study as you can incorporate it into your experimental designs. The registered reports format is offered by over 300 journals. More information can be found here: https://www.cos.io/initiatives/registered-reports There are also several journals that publish robustly designed studies regardless of whether the results are statistically significant, including PLoS journals and F1000 research. It is interesting to note that many journals are publishing negative findings these days, provided that the study design is appropriate. That's why it's critically important to have a robust study question – if you are trying to answer a question that is relevant to the public, then it should be worthy of publishing regardless of the findings.
Has anyone used the EDA in designing wildlife studies?
The EDA can be used to design wildlife studies. We are not sure if anyone has used the EDA for wildlife studies as we do not have access to user accounts, or the experiments users design, for security reasons. If you do want to use the EDA for wildlife studies bear in mind that many of the examples given are for laboratory-based experiments. If your experiment is exploratory or observational the analysis suggestions will not necessarily be relevant to your study. The NC3Rs website has a page on wildlife research that might also be helpful: https://nc3rs.org.uk/3rs-resources/wildlife-research
Should we decide on effect size based on biological reality?
It is recommended that researchers base the effect size they use in power calculations on the minimum difference between groups that is biologically relevant for the scientific question. For example, if you are looking for an improved drug to treat the symptoms of stroke and current gold standard treatments reduce symptoms by 30% your effect size for a power calculation might be 30%, as you are looking for a drug that improves symptoms of stroke by 30% or more.

How do we account for unequal variances across sample groups?

If you are determining sample size for difference in means between 2 independent groups, then software such as G*Power allows the user to input different variances for each group. Sometimes though, even if you have different variance estimates for the groups it might be reasonable to assume that the underlying variances are the same and the estimates are different because of sampling error. In this case just use a pooled estimate for all groups.

What is the Bonferroni adjustment and why is it important?

Conducting multiple hypothesis tests can increase the risk of finding something 'significant' by chance alone (Type 1 error). It is important that this is addressed in scientific manuscripts by publishing study methods in advance, reporting all hypotheses that were tested (not just the significant ones), describing the study as an 'exploratory study', and by not getting too carried away with conclusions about 'significant' findings.

The Bonferroni adjustment is used to keep the Type 1 error rate at the specified level (e.g. 5%) across all the tests in an analysis.
This is what can happen if you allow multiple testing to get out of hand: https://xkcd.com/882/

Note that Bonferroni is only one of several methods for controlling Type 1 error across multiple testing. Others include Tukey, Sidak, and False Discovery Rate.

It is also possible to adjust P-values using Bonferroni/Tukey etc - however it is very common to not do this (see link - https://www.jstor.org/stable/20065622#metadata_info_tab_contents)

What researchers should **not** do is P-hacking – see this video for an explanation - https://www.youtube.com/watch?v=i60wwZDA1CI&ab_channel=TED-Ed

The blanket rule that a statistically significant test result (P ≤ 0.05) means that the test hypothesis is false or should be rejected can lead to potentially wrong conclusions drawn from individual tests and presented in publications. How do we change this mindset prior to publication?

A lot has been written on this topic and of course there is no easy answer. For example. There is a move in epidemiology to do away with P-values and to use alternatives like confidence intervals. The American Statistical Association has been very active in recent years on this topic. See the ASA President's taskforce statement: https://magazine.amstat.org/blog/2021/08/01/task-force-statement-p-value/

Studies that have multiple outcomes, measures and varying assumptions can make simple power calculations, including paired t-tests, an unsuitable analysis method. Is it more appropriate to use more sophisticated techniques that consider the more complex realities and recognize that simple power calculations are questioned on both theoretical and practical grounds? If so, what are some examples?

In cases where power analysis on a simplified version of the intended design is not appropriate then it will often be necessary to do the power calculations by simulation methods. There are software programs that provide the tools to do this, but they still rely on the user providing good estimates of what the data parameters will be. An example is the "superpower" package by Lakens and Caldwell (Lakens being a leading researcher in this area) https://www.researchgate.net/publication/350317460_Simulation-Based_Power_Analysis_for_Factorial_Analysis_of_Variance_Designs