Experimental design workshop
George Savva (QIB) - January 2024

Reading ahead of experimental design workshop

This workshop will introduce some key ideas in experimental design and analysis. For much of the workshop we will discuss questions motivated by examples from NBI research over the past few years.

To get the most from the session you should read through the eight experimental design dilemmas below and think about how you might answer the questions. Decide on which you would like to discuss further during the workshop. We will spend more time on those that are of most interest to most people, but I will provide a written discussion of each issue following the session.

As a professional scientist you should be prepared to study the science of experimental design as well as your substantive area of interest. The specific issues you will encounter will vary a lot depending on the kind of research you are doing, but there are some common themes that we will discuss during the workshop.

I will also direct you to field specific literature, but throughout your PhD and beyond you should be prepared to seek out literature on study design relevant to your area of work.

There has been a lot written over the past 10 years about problems with accuracy and reproducibility of published biological research results. In 2015 our funders (BBSRC, MRC, Wellcome Trust and the Academy of Medical Sciences) prepared a report into causes of this and potential solutions. Much remains relevant today. While I do not expect you to have read this ahead of the workshop, it is an important document for all involved with biological scientists to be familiar with, with the need for better experimental design and statistical analysis emerging as a central theme.

You can find this report and related documents at the Academy of Medical Sciences website here: https://acmedsci.ac.uk/policy/policy-projects/reproducibility-and-reliability-of-biomedical-research

Reading on 'statistical significance'

An important recent article calls on scientists to reconsider the use of 'statistical significance' as a goal of scientific experiments, which has important implications for design and analysis of experiments. While I don't advocate all of the arguments here it does include some important ideas and reflects one of the main debates and sources of confusion in statistics today. We will cover this material in the workshop, but if you get time, please read the short article by Amrhein and others before our session.

Amrhein, Greenland and McShane (2019) Retire Statistical Significance. *Nature* 567:305-7 https://www.nature.com/articles/d41586-019-00857-9

Experimental design workshop
George Savva (QIB) - January 2024

Experimental design and analysis dilemmas!

I have encountered all of these 'experimental design' related dilemmas while working with NBI staff and students.

During the workshop we will break out into groups where you will discuss different options for solving each dilemma, before you report back to the rest of the cohort.

Before the session! **Read all of the dilemmas, think about how you might answer the questions, particularly if they seem relevant to your own work, and decide which you are most interested in discussing during the workshop.**

1. Early diagnosis *(selection)*

You want to study the effect of gut microbial composition on Parkinson's disease, using an observational design. Your plan is to compare the microbiomes of people with and without Parkinson's disease to look for associations.

How should you select your cases and controls?

2. Trust me, I'm a doctor *(feasibility / human studies)*

You worked with a statistician to calculate that you need 40 participants for your endoscopy study. Your second supervisor (a senior clinician) says this isn't a problem as their clinic is always busy with patients, so you shouldn't worry about it. However your inclusion criteria are quite strict and you aren't sure how easily you can recruit people for what is quite an invasive study.

What should you do?

3. Optional stopping *(p-hacking / QPRs)*

You have completed an experiment into the effect of a new growth medium on growth in wheat plants. You get a p-value for the effect of 0.07! Your supervisor suggests that if you grow a few more plants then analyse all the data together then you might get down to $p<0.05$ so that you can publish. Alternatively, you could switch the focus of the paper to a different outcome measure where you did see a significant difference.

Are these strategies acceptable? What should you do? What could you have done at the outset to avoid the issue?

4. Inadequate sample size *(power)*

You want to know whether a difference in bread starch content can reduce stomach bloating. The only feasible way to measure this is stomach volume measured by MRI scanner. You can afford to test 20 patients using this technology, but a sample size calculation suggests you need at least 80 to have a good chance of detecting the effect you are interested in.

What should you do?

5. **Replications give different results *(replication / significance)***

You run a colon model experiment into the effect of different substrates on short chain fatty acid production using stool samples from human donors. You see a significant improvement with your experimental substrate compared to a control.

However when the experiment is repeated (with exactly the same conditions), the difference was not statistically significant. Your supervisor suggests running a third replication to see which was correct, again this is not statistically significant.

What could have happened here? What would you conclude? What should you do? What could you have done differently at the outset?

6. **Mouse microbiome study with limited resources (*confounding/pseudo-replication/feasibility*)**

You are conducting an experiment on germ-free mice, to test the effect of an experimental bacteria on stress hormones (compared to untreated control). You only have access to two isolators.

You know that you cannot house treated and untreated mice in the same isolator because there will be contamination between the groups. But if you house all of the treated mice together and all of the untreated mice together you will have completely confounded any isolator effect with the treatment effect, so the treatment effect cannot be identified.

What should you do?

7. **Sample size calculation when no data is available (*sample size*)**

You are planning an experiment and need a sample size calculation. Your sample size calculator requires you to enter the variance of the outcome measure, and what is the smallest target effect size you want to be able to detect. Unfortunately, you have no prior information with which to calculate either of these things.

What should you do?

8. **Choosing between two analyses that give you different answers (*p-hacking, pre-registration*)**

You have conducted an experiment to check whether mice have better blood sugar on an experimental diet compared to their usual diet.

When using the conventional t-test that you had planned, there is no evidence of a difference. But a post-doc in your lab shows you a new statistical test that suggests the groups are significantly different.

What should you do? Which should you report?