Experimental Design Considerations - OBE – Colorado College.

Experimental design involves judgement and trade-off's - no one can provide ironclad rules for every situation. However, keeping these considerations in mind will strengthen your study tremendously.

Several definitions necessary to understand the material below:

1) Population--the set of things about which you want to make a generalization, e.g., all healthy flat-topped acacia individuals > 4 m tall growing on middle slopes in the northern portion of Tarangire National Park;
2) Unit--one member of the population. In biology, the unit is often one genetic individual, but could be a herd or nest, depending on your study,
3) Sample--the subset of the population that you measure or interview.

In comparative studies, minimize the effect of confounding factors

You often want to isolate the effects of one variable (the independent variable) on an outcome (the dependent variable). In comparing groups, a confounding factor is an undesired, additional independent variable that differs systematically among the groups you want to compare.

> Example: Let's say that you would like to compare growth rates of an herbaceous plant species at high and low elevations. At low elevations, you plant the seedlings from a greenhouse at valley bottom sites which happen to have easy access from roads. The high-elevation roads provide easy access to dry ridgetops, and you plant the seedlings there. Since most plants grow faster in moister areas, your study design confounds elevation and moisture, i.e., if there are differences between drier high elevation sites and moister low elevation sites, you cannot know if elevation or moisture or some combination of the two cause the differences. In order to understand the effects of elevation, you must keep moisture and other factors that might affect growth rate constant.

Sometimes people confuse confounding factors with factors that produce random variation within groups. Within each elevation in the first example, genetic differences among seedlings and microsite differences would cause variation among growth rates in the plants. However, genetic and microsite differences are *not* confounding because they do not consistently cause higher or lower growth rates at one of the two elevations; they simply create random variation of growth rates within each site.

To minimize effects of confounding factors, write down all the factors (independent variables) that might affect your outcome (dependent variable). Design your study to keep all factors except the one independent variable you want to study as constant as possible.

Ensure adequate sample size

How many animals, plants, rates, etc., should you measure so that other researchers or policy makers will have confidence in your results? You don't want your results to lack statistical significance when you are pretty certain that a larger sample size would have shown a difference, but you also don't want to waste time measuring many more than you need in order to show differences among treatments. In order to determine how many units to measure, always consider the variation in the phenomenon that you are measuring. More variation usually indicates that you will need to take more measurements to observe a significant effect. Generally, the greater the replication we have in an experiment, the better is our estimate of the random variation effects within sample plots in a particular treatment.

Although replication generally increases the precision of an estimate, in deciding on your sample size, first consider your time and logistical constraints. Within these limitations, you might look at published literature for sample sizes used in similar studies. *If* these studies found significant differences and *if* you feel that your system has similar variability, you might want to propose using similar sample sizes. You can quantitatively estimate necessary sample size if you have 1) quantitative measures (s, s^2, or MS_E from an ANOVA table [MS_E measures s^2]) from the published literature regarding the variability of your response variable or 2) measures of s^2 from your own preliminary data.

You should generally try to avoid pseudoreplication, in which your experiment might lack replication at certain levels (like a treatment level, for example). If you want to compare some parameter, say brood size of a bird species within burned and unburned patches of forest, pseudoreplication can occur if you have only one study site in burned and one study site in unburned forests. If your results indicate that brood size is impacted by burn site, pseudoreplication may be an issue because you have data from only one patch in each type (burned vs. unburned). It can be hard to know if that single patch is representative of the rest of that burn type. Pseudoreplication will not impact your ability to draw conclusions about the differences between those two specific sites but you can be limited in your ability to describe general effects of burning across the habitat you are studying. Sometimes replication is not possible, due to factors such as the limited size, number, or quality of sites (e.g., burn areas) available for study, or due to logistical limitations. If this is the case, you should acknowledge such constraints in your study design, and in your conclusions exercise caution in the degree to which you assign causation to treatment effects, and frame interpretations of your data as new hypotheses to test.

To alleviate issues with pseudoreplication, ideally you should seek to increase the number of study sites or sampled plots within each patch of forest type (i.e. multiple sampling sites in burned vs. unburned areas); you can also replicate your experiment in time by repeating your measurements in each of your forest types for multiple seasons or years. It would also be a good idea to go back to the definitions at the beginning of this document and write down the statistical population to which you would like to extrapolate your findings. If it is possible to sample multiple units of that population, you will be able to draw stronger conclusions about the

particular population you sampled. Time and resources always constrain the number of sites you can study; simultaneously, we often want to suggest that our findings apply beyond our study area. Most researchers solve this conundrum by drawing conclusions about their study area and then suggesting, with a rationale, the other areas to which their findings might apply – replicating your experiment in time helps to makes these arguments even stronger.

Consider independence of observations

Statistical tests (with some exceptions in advanced techniques) assume independence of observations, i.e., that no measures are more related to some other data points than to the remaining observations. For example, if you record behavioral observations of more than one animal in a herd, measurements of the animals within that herd are not independent. They are more likely to be similar to other measurements within that group than to measurements outside of the group [note: this is assuming that animals within a herd experience similar conditions].

Thinking about this in another way, consider the measurements in each group of an ANOVA; if you can cluster any of these data points because they are more likely to be similar, you do not meet the independence assumption. If observations within a group are not independent, the increased similarity of their measurements will unfairly increase the probability of showing a statistical difference. (For those who know some statistics: In an ANOVA, lack of independence tends to decrease MS_E, which leads to a larger F value and lower P value.) To avoid this problem, you might measure one leaf from a plant rather than several or analyze the data with a nested ANOVA, which accounts for the relatedness. Sometimes you can average responses in a group, e.g., in a recent ACM project, a student had quantitative observations of either one or two animals per herd. Considering each animal an independent observation would have been inappropriate; she solved the problem by averaging results with two observations per herd.

Lack of independence can also occur with correlation and regression. If you can group some points in a scatterplot with other points, you violate the independence assumption. For example, you violate this assumption if you want to determine whether needle length of a pine species changes with precipitation across many sites in a region and you measure several needles from each tree. Inappropriately using regression to analyze these data will likely yield more significance than it should. Relatedness can also occur through time; e.g., in looking for a trend over time in population size, the size at any time is likely closer to the previous size than to a random population size. Special regression techniques can handle this situation, but standard regression will yield spurious results that are more likely to show significance.

Think about the spatial and temporal variation of the responses you gather

Every response in a research project varies, either modestly or enormously. Growth rate of one plant species certainly changes along a hillslope due to changes in water availability and soil texture. Grazing animals might spend more time vigilant (heads up and looking around) when they are near thick brush that can hide predators. Sometimes you explicitly want to study these

spatial and temporal differences. In other situations, these variations will make finding patterns in your information very difficult. In either case, if you consider these variations, you will design a better study.

One approach to thinking about spatial and temporal variation is to visualize a map that shows variation in what you will measure. This map might have different intensities of a color with dark colors representing high values of a parameter and light colors representing low values. Identify the patterns in this mental map and ask yourself what factors might be responsible for the patterns. If looking at growth rate of a wide-ranging plant, the colors may be darkest near streams and lightest on dry ridgetops with some less obvious, finer-grained patterns in areas shaded or unshaded by trees.

To add temporal change, imagine the map changing through time. For example, plants in unshaded areas senescing earlier in the season than shaded plants, or, at a shorter time scale, plants in the sun getting water stressed earlier in the day than shaded plants.

Randomly select units to measure

If you will use statistics to fully analyze your data or just report means, standard errors, or confidence limits, you must sample randomly. If you do not, your statistical results and conclusions cannot be trusted. When attempting to sample randomly, you should use a mechanical means to generate random numbers (like a random number generator function in a calculator).

In a random sample, each unit in the population has an equal probability of being chosen. The best (but often unreasonable) way to do this involves numbering each unit in the population and using a random number table, random number generator on a calculator, or hundredths digit on a digital stopwatch to choose which units you measure. Use this approach when reasonable, but usually you will use a more time-efficient approach. For example, you might tell the driver to stop the vehicle when the tenths digit on the odometer is a certain digit. After stopping, you might walk a set number of paces forward and sample the nearest individual of a tree species on the right side of the road. This might work well, but if you are constantly thinking critically, you may realize that if the trees vary in density, trees more distant from others are more likely to be chosen. Think about the definition of random sampling (each unit having an equal probability of being chosen) in evaluating your proposed methods and consider how your proposed technique might bias your results.

Before considering the bias in your technique, you should also carefully define your population. You would not want to compare growth rates of seedlings to mature trees, and you might be interested in how many mature trees respond to elephant browsing. Based on previous research and your good judgement, you decide that trees greater than a certain size and with a crown at least 80% alive best represent the healthy, mature trees you want to study. While considering other factors that might affect growth, you realize that damage to roots from the road, road dust, browsing damage on the trunks, position on the hillslope, etc., can affect growth. You might decide that your population of interest includes trees of the appropriate size a certain

distance from the road and on mid-slope (not in a swale and not near a hilltop). Generally, you want to choose these characteristics so that your data will apply to most individuals of the species but use your judgement. If you are comparing three groups, e.g., three habitats, choose random samples from within each group.

You might also judge that you want to sample several trees/rocks/etc. at a randomly chosen point. If you do this, your sample size is the number of points, not the number of trees (see section on sampling independently). Each situation is different, but if getting to the next sample does not take too much time, your study may well have more statistical power (be more likely to show significant differences) with one tree per sampling point. The best approach will vary according to your situation.

Avoid bias

Humans are extremely suggestible, and scientists go through many transformations to bend their results to fit their preconceived ideas while still convincing themselves they behaved objectively. (Stephen J. Gould's *The Mismeasure of Man* provides disturbing descriptions of this phenomenon.) Due to this well-documented tendency, medical studies use double-blind designs where neither the patient nor provider knows if the patient is receiving the test drug or placebo. In lab biology, especially when rating on a subjective scale as in rating disease in microscopic analysis of tissue, specimens are given to the evaluator with only a number and not the name of the treatment. In the field, similar blinding often poses more challenges since you know, e.g., whether you are in a burned or unburned forest. In these situations, explicitly state your expectations/desires/hypotheses and constantly ask yourself if you are biasing your results. If you use subjective scales, write them out and have pictures of each level.

Plan ahead for statistical analysis

Beginning researchers often make the mistake of collecting data and only afterwards thinking how they will analyze their data. All too often, this leads to some or all their data being useless. Don't waste your valuable research time and effort; always think through your statistical analyses before you collect data. You may consult previous studies in your field for how to analyze data but think critically as standards have changed substantially through time and even otherwise strong papers have used incorrect statistical methods.

As part of planning ahead for analysis, consider the types of variables you will gather so that their level of measurement will fit the analyses. In organismal biology and ecology, you will often collect continuous data or counts but may also use other levels of measurement. To help think about statistical analyses appropriate for your data, decide which category below fits each of your variables; then review your statistics notes, see a statistical text, or consult an experienced data analyst for assistance.

1) Nominal data. Levels have names but no order, e.g., species name, sex. You often end up with counts or continuous measurements for various levels of a nominal variable. See

those data types for analysis.

2) Ordinal data. The levels of a variable have an order, but the levels are not necessarily the same distance from each other, e.g., small / medium / large, very strong / strong / moderate / light / very light disease impacts. When possible, e.g., small/medium/large, you will often obtain better results with continuous variables, but sometimes the time efficiency of ordinal variable outweighs this consideration. When analyzing ordinal data, you will often use the Kruskal-Wallis test to compare groups and Spearman's rank correlation to look for relationships between two variables.

3) Continuous data. This very common type of variable can take an infinite number of values between any two values, e.g., density, phosphorus content, length of twig, speed, and percent (however, if percent was calculated from counts, you often should use the techniques under count data).

    a. To determine if two variables are related, use regression (if cause and effect are obvious and the independent variable is measured with minimal error) or correlation (for other situations).
    b. If you want to compare means of groups, use a t-test (paired or random groups, depending on your data collection design) for two groups and one-way ANOVA (analysis of variance) for two or more groups.
    c. Several other types of ANOVA exist for more complex data collection designs, Including nested ANOVA (e.g., several twigs measured for each tree, several trees per site, and several sites per habitat) and two-way ANOVA (e.g., mass of an animal according to two factors, such as sex and habitat), and repeated measures ANOVA (if you have measured the same unit at two or more points in time).

4) Count data. Counts within nominal categories usually lead to chi square analysis, e.g., number of times impala occur with baboons vs. without baboons. If you have an expected ratio of counts, use a chi square goodness of fit test; if you do not have an expected ratio and are looking to see if variables are associated, use a chi square test of independence.