Protocol Guidelines

There are four important steps to take in order to prepare for your experiment. These steps are not necessarily sequential, but they are all important to complete before you run the experiment.

1. The Dry Run

Create an initial protocol, gather/create your materials, etc. then perform a dry run of the experiment on each other. This should not replicate the entire experiment. You should have as your goal to make sure that you test each level of each factor. Here are some guidelines of how this might look, but you should adjust this to match your experiment:

- If you have a CB experiment, do a CB with only the three members of your group as subjects.
- If you have a factorial experiment that involves a lot of treatments and no real blocking, then do enough runs so that each level of each main effect is tried (which will be in combination with levels of the other main effects).

After the dry run, reflect on improvements you could make to get a finalized protocol.

2. Estimate σ^2 for Power Calculations

A key ingredient for your power calculation is an estimate of σ^2. Ideally you can use your estimate from your initial proposal, if not too much has changed (and you actually did it for your initial proposal). If you use your dry run data, be sure that the nature of your treatments is reflective of the treatments in the real experiment.

3. Power Calculations

See the lab assignment on power for details about how to calculate power for a more complicated design. Note that the effect sizes (i.e. parameter values) should be set to the smallest value that would be scientifically meaningful to you. Ideally, you would obtain power of roughly 0.7-0.95 for all of your tests. If your power curve suggests you a sample size that is not feasible for this experiment, you can pick the largest feasible sample size even if that means you will have lower power.

If you find that the number of samples is surprisingly small then bump it up to a larger number that is still feasible. For example, if it suggests only 3 blocks for a blocking design, that is probably surprisingly small, particularly if blocks are subjects. However, if you are running a factorial with a large number of treatment

combinations (e.g. 12+ combinations), then $n = 3$ replications would imply a lot of individual observations (36+), so that's not surprisingly small.

If you get surprisingly small n, there are two common problems that might be causing it, so look more closely before you finish: 1) You defined the effects you want to be able to detect as very far apart. While not always an error, if the effect size you defined was not realistic for what you could reasonably hope to see if the alternative was true, then your power analysis is not going to provide relevant guidance for obtaining a practically interesting answer to your research question. 2) You have done the power calculations incorrectly.

4. Final assignments of treatments

After you have n from your power calculations, you should make your final assignments of treatments to units so that you are completely ready to run the experiment. This should take the form of a data frame with one row per measurment unit and columns for the unit identifier and the treatment(s) assigned to that unit. You should also have a column for the block (if applicable).

Experimental Proctocol (your writeup)

Your protocol should be in paragraph format, except where noted below. The protocol writeup should contain the following information:

1. Start with a short description of the basic experiment to remind me, like the design proposal. This should briefly (in 1 paragraph) describe the core parts of the experiment without getting into the detail described below. This can be copied from your earlier submissions and updated as relevant.
2. Describe your power calculation, and the results of your calculation (i.e. what does it tell you about how many observations you are going to run). In describing your power calculation, you should include:
    - a definition of how you have defined sample size (is it # subjects, # runs, # LS squares, etc?). It should be be a reasonable value for your experiment.
    - your estimate of σ^2 and a description of how you estimated it, i.e. how you collected the data, how many observations it's based on, what estimate you used. You can use the initial estimate of σ^2 from your initial proposal, if that is still relevant to your current design, but you still need to describe how you got it.
    - A description of why you chose your alternative hypothesis. This should be based on your understanding of the problem, so your description is a justification of why it made sense.
    - Power curves. You should have one for each experimental factor and interaction term if you have them.
    - The final sample size you chose, and what specific power that gives you.

If your power curve gives you a sample size that is not feasible for this experiment, you should pick the largest feasible sample size, and then describe both:

- what power you will have with this reasonable sample size for your original choice of effect size
- what effect size you could actually find with good power (i.e. 0.7-0.9) with this reasonable sample size

3. Describe the finalized protocol of the experiment. This can be in list or bullet point format to indicate the steps to be followed, as appropriate. It should be written so that you could hand this off to someone else and they could run the experiment correctly. Think of it as a training guide for someone you've recruited to run the experiment for you.
4. "Preregister" your plan of analysis. This is a best practice in the scientific community, and helps prevent "p-hacking" and other questionable research practices. You should give a complete list of all of the inferential procedures you intend to run (i.e. hypothesis tests and confidence intervals), and you should describe how you will interpret the results of those procedures. For example, if you are going to run an ANOVA, you should say something like "If the p-value for the main effect of factor A is less than 0.05, then we will conclude that there is evidence that factor A has an effect on the response variable." You should do this for each test or confidence interval that you intend to run.

You should also include as appendices to your protocol the following information:

- The data frame with the final assignments of treatments to units.
- Code to perform the analysis outined in your preregistration so that it's ready to go when you have the data. This should be well organized and commented so that it is clear which code corresponds to which part of the analysis. You can also include code for any data formatting or cleaning steps that you intend to perform before the analysis. Note that this code can be directly copied from the code you used to perform the power calculations.
- Describe your dry run and what changes were made from your first initial draft of the protocol based on your experiences (there should always be some changes).
- (if applicable) Copies of any materials you had to create (e.g. if you gave subjects a test, attach the questions that they answer; if they have to read a passage, attach the passage(s), etc.). If you used some existing tool (e.g. an online quiz or typing program) please provide screen snapshot(s) of it so it is clear how it works for someone trying to implement your procedure. If the materials are not feasible to attach (e.g. a 3D game that they played) please provide a photo of them `in action'.