Power Analysis for Multilevel Modeling

Note: The following materials are from a manuscript under review. If you would like to view the entire manuscript, please contact Kate Thorson at katherine.thorson@gmail.com. Please cite the work as follows:

Thorson, K. R., West, T. V., & Mendes, W. B. (2017). Measuring physiological influence in dyads: A guide to designing, implementing, and analyzing dyadic physiological studies. Manuscript under review.

From the main text:

**Power analysis.** One approach to conducting power analyses when using the stability and influence model is a simulation method (see Bolger et al., 2012, and Lane & Hennes, 2016). This technique has the advantage of being flexible and able to accommodate many types of models. Because a substantial amount of information is needed to conduct such analyses, we strongly recommend pilot testing before conducting these analyses to gain a sense of values of the stability and influence paths and the random effects in the model.

Here, we outline four steps for a researcher who has collected pilot data and wants to conduct a power analysis using a simulation method to plan the final sample size. We present annotated SAS syntax for each of these steps in our Supplemental Materials. The first step would be to run the model one intends to use on a full sample of dyads on a pilot sample of dyads. The pilot data can then be analyzed to get the fixed and random effects estimates to simulate data for 1000 hypothetical studies with the same number of dyads and time points as the pilot study. When the data have been simulated, the third step is to analyze each of those 1000 samples individually, using the model that was run on the pilot data. Initially, we suggest using the same number of dyads and time points in one’s pilot study to check the estimates, standard errors, and degrees of freedom obtained in the power analysis against the results obtained from the pilot study. This can identify mistakes in one’s syntax at this stage of the power analysis. Next, one can then document the number of times a hypothesized effect is significant, using the percentage of significant effects across all 1000 studies as an estimate of power. One can then go back to the second step of simulating data and change the number of dyads and time points per sample and repeat the 3rd and 4th steps, discovering the final sample size needed for a hypothesized effect to have sufficient (typically 80%) power. We fully outline how to implement each of these four steps, as well as tips and potential pitfalls, for the stability and influence model in our Supplemental Materials.

From the Supplemental Materials:

**Power Analysis**

We walk through a power analysis in SAS using data collected from 34 dyads composed of one man and one woman each across ten time points. This sample of 34 dyads is used as a pilot study to plan the sample size for the full study. The dependent variable in the following model is receiver pre-ejection period (PEP) reactivity (“pepreact\_R”). Receivers’ own PEP reactivity at the prior time point (referred to as receiver lag) provides the stability path (“Zpep\_lag\_R”). Senders’ PEP reactivity at the prior time point (referred to as sender lag) provides the influence path (“Zpep\_lag\_S”).

The first step of the power analysis is to run the model one intends to use on the full sample of dyads on this pilot sample of 34 dyads. In Figure S8, we present this model. Note that for ease of presentation, we include only the stability and influence paths as fixed effects and estimate only a random stability path, but additional effects can be added to the model. “Dyad” is a unique identification number for each dyad, which is the same for each member of a dyad. “Obs\_id” is a unique identification number for each pair of observations that occur at the same time point for the same dyad. It is calculated as “time + nt(dyad-1)” where “time” represents the time point of the observation, “nt” is the number of time points, and “dyad” is the unique identification number for each dyad. “Gender\_class” is a variable representing gender, which is coded as -1 for females and 1 for males. “Male” is coded as 0 for females and 1 for males and “female” is coded as 0 for males and 1 for females. The receiver lag and sender lag variables have been standardized before conducting the analysis. The reason for this will become clear in the following step.



**Figure S8.** Syntax for Step 1 of the power analysis: conducting a two-level crossed model on a pilot sample of data.

 Once the pilot data are analyzed, that information can be used to simulate data for 1000 hypothetical studies. In Figure S9, we show syntax for how to do this. In line 3, we chose an arbitrary number for the random number generator so that the simulation will output the same values every time. In lines 4 through 6, we specified the value for the intercept (line 4), the coefficient for receiver lag (line 5), and the coefficient for sender lag (line 6) that we obtained from the model in Figure S9. Note that the values in lines 4 through 6 are unstandardized. We specified the value for the random receiver lag slope in line 7 for women and in line 8 for men. We specified the residual variance for women in line 9 and for men in line 10. The common covariance is listed in line 11.

In line 12, we indicated the number of samples to be generated, and in line 14, we indicated the number of dyads for which we wish to obtain power estimates. Initially, we suggest using the same number of dyads and time points in one’s pilot study to check the estimates, standard errors, and degrees of freedom obtained in the power analysis against the results obtained from the pilot study. This can help to identify mistakes in one’s syntax at this stage of the power analysis. In lines 15 and 16, we created the unique receiver lag slopes for each dyad—once for women and once for men. We indicated that there are two people per dyad in line 18, and in lines 19 through 24, we assigned those individuals to the appropriate value for the variables “gender\_class,” “male,” and “female.”

In line 25, we specified that each person has ten time points of data (as in the pilot study). We chose numbers randomly from a standard normal distribution in lines 26 and 27 to be the receiver and sender lag variables. Because these numbers are selected from a standard normal distribution, we used standardized receiver and sender lag variables in the model conducted on the pilot data to ensure that our simulated data matches our pilot data as closely as possible. In lines 29 and 31, we generated a residual for women and men, and in lines 30 and 32, we specified the model, as used in Step 1 on the pilot data. Once this step has been completed, we recommend opening up the new file that has been generated (here it is called “example1”) and checking that the data appear as you anticipated—for example, that there are 34 dyads per sample, 2 people per dyad, and 10 time points per person, with unique X1 and X2 values at each time point.

**Figure S9.** Syntax for Step 2 of the power analysis: simulating data for 1000 hypothetical studies.

Once the data have been simulated, Step 3 is to analyze each of those 1000 samples individually. We present syntax for this step in Figure S10. Note that the model directly replicates the model specified in the first step in Figure S8. Lines 9 and 10 of this syntax create two new data files—one with the results for the fixed effects for the analysis conducted on each sample and one with the results for the random effects for the analysis conducted on each sample.



**Figure S10.** Syntax for Step 3 of the power analysis: analyzing the simulated data for the 1000 hypothetical studies.

The final step in this process is to go through the two new files created and count the number of times that an effect is significant (out of the 1000 samples). Syntax for this purpose is presented in Figure S11. Lines 1 through 6 provide syntax for looking at the fixed effects, and lines 7 through 16 provide syntax for looking at the random effects. For every two lines of syntax, SAS will output data that indicate the average estimate, standard error, degrees of freedom (for fixed effects), and the average amount of times an effect was significant. Figure S12 presents this output for X2 (produced by lines 5 and 6 of the syntax in Figure S11), which is the sender lag variable. The value in the “sig” line represents the power for the specified effect, given the parameters specified. In Figure S12, the power for detecting a significant effect of X2 is 48.10%, given the parameters specified.



**Figure S11.** Syntax for Step 4 of the power analysis: counting the number of times each effect is significant.



**Figure S12.** SAS output from lines 5 and 6 of Figure S11 in Step 4 of the power analysis.

As noted before, when first conducting a power analysis on a sample of data, we suggest checking the estimates, standard errors, and degrees of freedom obtained in the power analysis against the results obtained from the pilot study to help identify errors in one’s syntax. If the results here match what was found on the sample of data, one can then move ahead and change parameters of the model in Step 2 (for example, number of dyads, time points, size of effects, etc.) to examine how these changes will impact the power obtained for various effects. If researchers have difficulty obtaining results that match what they found with their sample data, we recommend simplifying the model as much as possible, checking that results match, and then adding complexity from there. One reason results may not match is because of missing data. If there are missing data in the pilot study but missing data are not simulated in Step 2 (as we did not do here), researchers will find that the degrees of freedom will be different. Furthermore, if the variables in one’s pilot study are not normally distributed but values in the second step of the power analysis are chosen from a standard normal distribution, the results may also be slightly mismatched.