A Guide to Power Analysis for Hypothesis Tests with One Categorical Independent Variable with Two Groups

Anna-Marie Ortloff
University of Bonn

Christian Tiefenau
University of Bonn

Matthew Smith
University of Bonn, Fraunhofer FKIE

Four parameters are relevant to power analysis: Power, the significance criterion (i.e. the α error level), the reliability of the sample results or sensitivity of the test, and the effect size [2]. The power of a statistical test is the probability of the test correctly rejecting the null hypothesis, i.e. that a statistical test yields a significant result, when the alternative hypothesis is true [4]. Power can also be represented as $1-\beta$, wherein β is the Type II error, i.e. wrongly rejecting the null hypothesis. This means that if a test has a statistical power of 0.8, as is an often used, acceptable value [2, 3], an actual effect will be detected 80% of the time. The significance criterion or significance level represents the threshold of maximum accepted probability of making a Type I error, i.e. wrongly assuming the alternative hypothesis, detecting an effect, when there actually is none [2]. Using the widely accepted threshold of 0.05 for statistical significance means that only in 5% of cases, an effect is detected in the sample, even though in the population, it does not exist. Reliability refers to how well a sample estimate represents the corresponding population parameter [2]. Reliability is influenced by different factors, depending on the type of estimated parameter, such as the quality of the measurement instrument, and controlling sources of variance in the data, which might distract from the effect you are trying to measure [4]. The largest and invariably present influencing factor, however, is sample size [2], such that larger samples produce more consistent and reliable estimates than smaller ones. Finally, the effect size measures the amount of impact of an independent variable on dependent variables, rather than only judging the presence or absence of an effect [4]. There are generally two types of effect sizes: Non-standardized, or simple effect sizes, which represent the size of effect in the units of the outcome variable, and standardized effect sizes which represent the effect size relative to the variability in the sample or population [1]. When comparing two means, e.g. with a t-test, the difference in mean completion time between two different interface variants represents a simple effect size, measured in units of time, e.g. minutes, while a standardized effect size for this scenario, such as Cohen's d, takes into account the standard deviation in the two groups. Standardized effect sizes are commonly classified as either belonging to the d-family, such as Cohen's d in the example above, or as belonging to the r-family, such as the correlation coefficient Pearson's r [6].

These four parameters are interdependent, such that when three of them are available, it is possible to calculate the fourth. Such calculations are referred to as power analysis. In general, there are four different kinds of power analysis, each used to determine one of the parameters from the other three, although it is also possible to determine both α and power if a ratio for α and β is given together with the other two parameters - this is termed compromise power analysis [5]. The other four flavors are summarized, e.g. by Cohen [2] in Chapter 1.5.

The following short tutorials provide an overview of the data necessary to conduct power analysis for basic hypothesis tests, where to find this data in our database, and how to use it to conduct apriori power analysis using G*Power or R.

USENIX Symposium on Usable Privacy and Security (SOUPS) 2023. August 6–8, 2023, Anaheim, CA, USA

1 Fisher's Exact Test

1.1 Terminology

		DV	
		success	failure
IV	group 1	a	b
	group 2	c	d

Table 1: Contingency table of data used in Fisher's Exact test

odds ratio effect size, commonly abbreviated OR, OR=$\frac{\frac{a}{1-a}}{\frac{c}{1-c}}$ given a is the proportion of desired outcome for group 1 and c is the proportion of desired outcome for group 2, as in Table 1.

proportion of desired outcome propability of a positive outcome for one of the compared groups. The outcome does not necessarily need to be good in a general sense of the word, but it is important that the same outcome is considered for both groups.

1.2 What data do you need and where can you find it?

What data do you need?		**Where can you find it?**
Option 1	**Option 2**	
allocation ratio		Calculate from group sizes: $\frac{group\_2\_size}{group\_1\_size}$, both are fields in the database
	odds ratio	odds_ratio in the database; commonly abbreviated as OR, can also be reported as Exp(coef) or exponential of the coefficient, especially in logistic regression
expected success proportion for group 1	expected success proportion for group 1	group_1_p_desired_outcome in the database; p(a) (see Table 1)
expected success proportion for group 2	expected success proportion for group 2	group_2_p_desired_outcome in the database; p(c) (see Table 1)

= all needed = only one needed

1.3 How to do power analysis in G*Power

1. Choose the *Exact* test family
2. Select *Proportions: Inequality, two independent groups (Fisher's exact test)*
3. Select *A priori* for the type of power analysis
4. (a) **Using Data from Option 1**
   Enter the data into the following fields:
   Allocation ration N2/N1 allocation ratio
   Proportion p1 expected success proportion for group 1
   Proportion p2 expected success proportion for group 2

   (b) **Using Data from Option 2**
   Enter the data into the following fields:

Allocation ratio N2/N1 Adjust your allocation ratio so that the size N2 is the group size for the group where you do have data on the expected success proportion and N1 is the group size for the group where you do not

Proportion p2 Enter the expected success proportion for the group you have it available

Proportion p1 Determine the expected success proportion for the group where you do not have data available

- Hit the *Determine* => button
- Select *Calc P1 from ... odds ratio* and enter the odds ratio
- If not already given, enter the expected success proportion for the group you have it available in the field *P2*
- Hit *Calculate and transfer to main window*

5. Calculate your power analysis

1.4 How to do power analysis in R

1. Use the `ss2x2` function from the `exact2x2` package

2. **Using Data from Option 1**
   Enter the data as the following parameters:

   n1.over.n0 allocation ratio

   p0 expected success proportion for group 1

   p1 expected success proportion for group 2

3. **Using Data from Option 2**
   Enter the data as the following parameters:

   n1.over.n0 allocation ratio

   p0 expected success proportion for whatever groups' data is available

   p1 calculate the expected success proportion from the odds ratio and whatever groups' data is available by transposing the equation for the odds ratio in the terminology section.

4. Use parameter `print.steps=TRUE` to see progress, as execution can be slow

2 Chi-Squared Test

2.1 Terminology

		DV	
		success	failure
IV	group 1	a	b
	group 2	c	d

Table 2: Contingency table of data used in Chi-Squared test

2.2 What data do you need and where can you find it?

What data do you need?			Where can you find it?
Option 1	**Option 2**	**Option 3**	
phi			phi in the database; can be calculated from $\phi = \sqrt{\frac{\chi^2}{N}}$ for 2x2 tables
	Chi-Squared		chi_square in the database; χ^2; test statistic of the chi-squared test
	Number of participants		total number of participants involved in the test; n_of_participants in the database
		p1, p2	the success probabilities from the contingency table(Table 2), i.e. p(a), p(c), if larger than a 2x2 table, then more probabilities are necessary
degrees of freedom			often abbreviated df; degrees_of_freedom in the database; in a statistical report: $\chi^2(df) =$ value ofχ^2

= all needed = only one needed

2.3 How to do power analysis in G*Power

1. Choose the χ^2 *tests* test family
2. Select *Goodness-of-fit tests: contingency tables*
3. Select *A priori* for the type of power analysis
4. (a) **Using Data from Option 1**
   If your degrees of freedom are 1: Enter the data into the following fields:

   Effect size w phi

   Df degrees of freedom

   (b) **Using Data from Option 2**
   If your degrees of freedom are 1: Enter the data into the following fields:

   Effect size w is equivalent to phi for df=1: $\phi = \sqrt{\frac{\chi^2}{N}}$

   Df degrees of freedom

   (c) **Using Data from Option 3**
   Don't use this with data from Option 3, as *Determine* => calculates an effect size for a (one sample) goodness of fit test, instead of the test of independence.
5. Calculate your power analysis

2.4 How to do power analysis in R

1. Use the `pwr.chisq.test` function from the `pwr` package
2. **Using Data from Option 1**
   If your degrees of freedom are 1: Enter the data as the following parameters:

   w phi

   df degrees of freedom

3. **Using Data from Option 2**
   If your degrees of freedom are 1: Enter the data as the following parameters:

   w phi, calculated from $\phi = \sqrt{\frac{\chi^2}{N}}$ available data

   df degrees of freedom

4. **Using Data from Option 3**
   Enter the data as the following parameters (this also works when degrees of freedom are higher than 1):

   contingency table Use the group sizes and success probabilities to calculate a contingency table. Depending on your choice of calculating w, it should contain either values or probabilities.

   w calculate w using the `ES.w2` function from the `pwr` package. The function requires a two-way contingency table of probabilities. Alternatively, you can calculate w using the `phi` or `cohens_w` functions from the `effectsize` package, which both require a contingency table of values.

   df degrees of freedom

3 McNemar's test

3.1 Terminology

		post-study	
		result 1	result 2
pre-study	result 1	a	c
	result 2	b	d

Table 3: Contingency table of data used in McNemar's test

discordant pairs b and c in the contingency table 3, e.g. where group 1 before and group 2 after the intervention and vice versa

odds ratio OR; $\frac{c}{b}$

proportion of discordant pairs P(c) + P(b)

smaller of discordant pairs here b: $b = \frac{\text{total proportion of discordant pairs}}{OR+1}$

larger of discordant pairs here c: $c = b \times OR$

3.2 What data do you need and where can you find it?

What data do you need?				Where can you find it?
Option 1	**Option 2**	**Option 3**	**Option 4**	
odds ratio				effect size, often abbreviated OR, odds_ratio in the database
proportion of discordant pairs		proportion of discordant pairs		proportion_of_discordant_pairs in the database, can calculate this value if given a full contingency table from P(c) and P(b)
	one of the discordant pairs			either P(c) or P(b) from the contingency table (Table 3), in the database as lower_val_of_discordant_ps and higher_val_of_discordant_ps
			both of the discordant pairs	both P(c) or P(b) from the contingency table (Table 3), in the database as lower_val_of_discordant_ps and higher_val_of_discordant_ps

= all needed = only one needed

3.3 How to do power analysis in G*Power

1. Choose the *Exact* test family
2. Select *Proportions: Inequality, two dependent groups (McNemar)*
3. Select *A priori* for the type of power analysis
4. (a) **Using Data from Option 1**

   Odds ratio odds ratio

   Prop discordant pairs proportion of discordant pairs

   (b) **Using Data from Option 2**

   Odds ratio odds ratio

   Prop discordant pairs Calculate the proportion of discordant pairs from the given discordant pair and the odds ratio, using the formulas above

   (c) **Using Data from Option 3**

   Odds ratio calculate the odds ratio from the given part of the discordant pairs and the total proportion of discordant pairs, using the formulas above

   Prop discordant pairs proportion of discordant pairs

   (d) **Using Data from Option 4**

   Odds ratio calculate the odds ratio from the values of the discordant pairs, using the formula above

   Prop discordant pairs calculate the proportion of discordant pairs from the values of discordant pairs, using the formula above
5. Calculate your power analysis

3.4 How to do power analysis in R

1. Use the `power_mcnemar_test` function from the `MESS` package
2. **Using Data from Option 1**
   Enter the data as the following parameters

   paid calculate one of the discordant pairs using the formula above. Enter the smaller of the two discordant pairs

   psi odds ratio, but must be larger than 1, if not, calculate $OR = \frac{1}{OR}$
3. **Using Data from Option 2**
   Enter the data as the following parameters

   paid calculate the second discordant value using the formula above and enter the smaller of the two

   psi odds ratio, but must be larger than 1, if not, calculate $OR = \frac{1}{OR}$
4. **Using Data from Option 3**
   Enter the data as the following parameters

   paid calculate the second discordant value using the formula above and enter the smaller of the two

   psi calculate the second discordant value using the formula above and then use the two to calculate the odds ratio, using the formula above; OR must be larger than 1, if not, calculate $OR = \frac{1}{OR}$
5. **Using Data from Option 4**
   Enter the data as the following parameters

   paid enter the smaller of the discordant pair here.

   psi calculate odds ratio from discordant pairs, odds ratio must be larger than 1, if not, calculate $OR = \frac{1}{OR}$

4 Independent t-test

4.1 Terminology

Cohen's d effect size, can be calculated from means (M), standard deviations (SD) and group sizes (N). Cohen's d = $\frac{|M_1 - M_2|}{\text{pooled SD}}$

pooled SD = $\sqrt{\frac{(N_1-1)\times SD_1^2)+(N_2-1)\times SD_2^2}{N_1+N_2-2}}$

If sample sizes N_1 and N_2 are equal, pooled SD can be simplified to: pooled SD = $\sqrt{\frac{SD_1^2+SD_2^2}{2}}$

4.2 What data do you need and where can you find it?

What data do you need?		Where can you find it?
Option 1	**Option 2**	
Cohen's d		effect size, cohens_d in the database
	N for group 1 and group 2	number of participants in group 1 and group 2, group_1_size and group_2_size in the database
	mean for group 1 and group 2	mean of outcome variable for group 1 and group 2, often abbreviated M in reports, group_1_mean and group_2_mean in the database
	standard deviation for group 1 and group 2	standard deviation of outcome variable in group 1 and group 2, often abbreviated sd in reports, group_1_sd and group_2_sd in the database

= all needed = only one needed

4.3 How to do power analysis in G*Power

1. Choose the *t tests* test family

2. Select *Means: Difference between two independent means (two groups)*

3. Select *A priori* for the type of power analysis

4. (a) **Using Data from Option 1**
   Enter the data into the following fields:

   Effect size d Cohen's d

   Allocation ratio N2/N1 1, when the groups are planned to be of equal size. Calculate the allocation ratio depending on planned group size from $\frac{N_2}{N_1}$

   (b) **Using Data from Option 2**
   Enter the data into the following fields:

   Allocation ratio N2/N1 1, when the groups are planned to be of equal size. Calculate the allocation ratio depending on planned group size from $\frac{N_2}{N_1}$

   Effect size d Determine the effect size d

   - Hit the *Determine* => button
   - Select *n1 != n2* if the sample sizes were not equal and enter the means and the shared standard deviations
   - Select *n1 = n2* if the sample sizes were equal, and enter the means and standard deviations
   - Hit *Calculate and transfer to main window*
   - If sample sizes were not equal and standard deviations were also not equal, calculate d using the formula in the terminology section and enter the resulting value.

5. Calculate your power analysis

4.4 How to do power analysis in R

1. Use the `pwr.t.test` function from the `pwr` package

2. **Using Data from Option 1**
   Enter the data as the following parameters:

   d Cohen's d

3. **Using Data from Option 2** Enter the data as the following parameters:

   d calculate Cohen's d using the formula in the terminology section and enter the resulting value. If the full data is available, you can also use `effsize::cohen.d` or `effectsize::cohens_d`

5 Paired t-test

5.1 Terminology

Cohen's d_z effect size for paired samples (Not the same as d for independent samples!), can be calculated from means (M), standard deviations (SD) and the correlation between the two groups, or by defining a new variable from the differences between the two paired samples, and calculating d_z from the mean and standard deviation of that variable, or from the t-test-statistic and degrees of freedom:

Cohen's $d_z = \frac{M_1 - M_2}{\sqrt{SD_1^2 + SD_2^2 - 2 \times SD_1 \times SD_2 \times \text{correlation coefficient between group 1 and group 2}}}$

Cohen's $d_z = \frac{\text{mean of differences}}{\text{SD of differences}}$

Cohen's $d_z = \frac{t}{\sqrt{\text{df} + 1}}$

5.2 What data do you need and where can you find it?

What data do you need?				Where can you find it?
Option 1	**Option 2**	**Option 3**	**Option 4**	
Cohen's d_z				effect size, cohens_d in the database in the paired_t_test table, not to be confused with Cohen's d for independent samples
	mean for group 1 and group 2			mean of outcome variable for group 1 and group 2, often abbreviated M in reports, group_1_mean and group_2_mean in the database
	standard deviation for group 1 and group 2			standard deviation of outcome variable in group 1 and group 2, often abbreviated sd in reports, group_1_sd and group_2_sd in the database
	correlation between group 1 and group 2 outcomes			correlation in the database; Pearson's r correlation coefficient
		mean of differences		mean of differences in outcome variable between group 1 and group 2; differences_mean in the database
		standard deviation of differences		standard deviation of differences in outcome variable in group 1 and group 2; differences_sd in database
			t	test statistic t of the paired t test, reported in papers as t(df)=*value of the test statistic t*; t in database
			degrees of freedom	common abbreviation df, reported in papers in parentheses after the test statistic, e.g. t(*df*)=value of t. df + 1 is the number of units of measurement (commonly participants); degrees_of_freedom in database

= all needed   = only one needed

5.3 How to do power analysis in G*Power

1. Choose the *t tests* test family
2. Select *Means: Difference between two dependent means (matched pairs)*
3. Select *A priori* for the type of power analysis
4. (a) **Using Data from Option 1**
   Enter the data into the following fields:

   Effect size dz Cohen's d_z

   (b) **Using Data from Option 2**
   Enter the data into the following fields:

   Effect size dz Determine the effect size d_z

   - Hit the *Determine* => button

- Select *From group parameters* and enter the means and standard deviations of the two groups and the correlation coefficient for the correlation between groups
- Hit *Calculate and transfer to main window*

(c) **Using Data from Option 3**
Enter the data into the following fields:

Effect size dz Determine the effect size d_z

- Hit the *Determine* => button
- Select *From differences* and enter the mean of difference and SD of difference
- Hit *Calculate and transfer to main window*

(d) **Using Data from Option 3**
Enter the data into the following fields:

Effect size dz Determine the effect size d_z using the third formula in the terminology section

5. Calculate your power analysis

5.4 How to do power analysis in R

1. Use the `pwr.t.test` function from the `pwr` package, with `paired=TRUE`
2. **Using Data from Option 1**
Enter the data as the following parameters:

   d Cohen's d_z

3. **Using Data from Option 2** Enter the data as the following parameters:

   d calculate Cohen's d_z using the formula in the terminology section and enter the resulting value; if the full data are available, you can also use the `effectsize::cohens_d` function with `paired=TRUE` parameter

4. **Using Data from Option 3** Enter the data as the following parameters:

   d calculate Cohen's d_z using the formula in the terminology section and enter the resulting value; if the full data are available, you can also use the `effectsize::cohens_d` function with `paired=TRUE` parameter

6 Wilcoxon Rank-Sum Test

6.1 Terminology

Wilcoxon rank-sum test also known as Mann-Whitney U test or Wilcoxon Mann-Whitney test

A.R.E. asymptotic relative efficiency power relative to a two-sample t-test for independent samples

response distribution distribution of response (dependent) variable

minimal A.R.E. Can be calculated for response distributions with finite variance; since they are least efficient (relative to the t-test), they yield the largest necessary sample sizes, i.e. the most conservative estimate of required sample sizes

Cohen's d effect size, can be calculated from means (M), standard deviations (SD) and group sizes (N). Cohen's d $= \frac{|M_1-M_2|}{\text{pooled SD}}$

pooled SD $= \sqrt{\frac{(N_1-1)\times SD_1^2)+(N_2-1)\times SD_2^2}{N_1+N_2-2}}$

If sample sizes N_1 and N_2 are equal, pooled SD can be simplified to: pooled SD $= \sqrt{\frac{SD_1^2+SD_2^2}{2}}$

6.2 What data do you need and where can you find it?

What data do you need?		Where can you find it?
Option 1	**Option 2**	
response distribution		If unsure, select the conservative possibility of minimal A.R.E.
Cohen's d		effect size, cohens_d in the database
	N for group 1 and group 2	number of participants in group 1 and group 2, group_1_size and group_2_size in the database
	mean for group 1 and group 2	mean of outcome variable for group 1 and group 2, often abbreviated M in reports, group_1_mean and group_2_mean in the database
	standard deviation for group 1 and group 2	standard deviation of outcome variable in group 1 and group 2, often abbreviated sd in reports, group_1_sd and group_2_sd in the database

= all needed   = only one needed

6.3 How to do power analysis in G*Power

1. Choose the *t tests* test family

2. Select *Means: Wilcoxon-Mann-Whitney test (two groups)*

3. Select *A priori* for the type of power analysis

4. (a) **Using Data from Option 1**
   Enter the data into the following fields:

   Parent distribution Select a response distribution. If unsure, select *min ARE*, as this provides the most conservative estimate

   Effect size d Cohen's d

   (b) **Using Data from Option 2**
   Enter the data into the following fields:

   Parent distribution Select a response distribution. If unsure, select *min ARE*, as this provides the most conservative estimate

   Effect size d Determine the effect size d

   - Hit the *Determine* => button
   - Select *n1 != n2* if the sample sizes were not equal and enter the means and the shared standard deviations
   - Select *n1 = n2* if the sample sizes were equal, and enter the means and standard deviations
   - Hit *Calculate and transfer to main window*
   - If sample sizes were not equal and standard deviations were also not equal, calculate d using the formula in the terminology section and enter the resulting value.

5. Calculate your power analysis

6.4 How to do power analysis in R

Since power analysis for the Wilcoxon rank-sum test makes distributional assumptions, even if the test itself does not, simulation is necessary to estimate necessary sample size. For a more simple solution, use G*Power [5].

7 Wilcoxon Signed-Rank Test

7.1 Terminology

A.R.E. asymptotic relative efficiency: power relative to a two-sample t-test for independent samples

response distribution distribution of response (dependent) variable

minimal A.R.E. Can be calculated for response distributions with finite variance; since they are least efficient (relative to the t-test), they yield the largest necessary sample sizes, i.e. the most conservative estimate of required sample sizes

Cohen's d_z effect size for paired samples (Not the same as d for independent samples!), can be calculated from means (M), standard deviations (SD) and the correlation between the two groups, or by defining a new variable from the differences between the two paired samples, and calculating d_z from the mean and standard deviation of that variable, or from the z score and the number of participants (i.e. the number of observations divided by 2)

Cohen's $d_z = \frac{M_1 - M_2}{\sqrt{SD_1^2 + SD_2^2 - 2 \times SD_1 \times SD_2 \times \text{correlation coefficient between group 1 and group 2}}}$

Cohen's $d_z = \frac{\text{mean of differences}}{\text{SD of differences}}$

Cohen's $d_z = \frac{\text{z score}}{\sqrt{\text{number of participants}}}$

7.2 What data do you need and where can you find it?

What data do you need?				Where can you find it?
Option 1	**Option 2**	**Option 3**	**Option 4**	
response distribution				If unsure, select the conservative possibility of minimal A.R.E.
Cohen's d_z				effect size, cohens_d in the database in the paired_t_test table, not to be confused with Cohen's d for independent samples
	mean for group 1 and group 2			mean of outcome variable for group 1 and group 2, often abbreviated M in reports, group_1_mean and group_2_mean in the database
	standard deviation for group 1 and group 2			standard deviation of outcome variable in group 1 and group 2, often abbreviated sd in reports, group_1_sd and group_2_sd in the database
	correlation between group 1 and group 2 outcomes			correlation in the database; Pearson's r correlation coefficient
		mean of differences		mean of differences in outcome variable between group 1 and group 2; differences_mean in the database
		standard deviation of differences		standard deviation of differences in outcome variable in group 1 and group 2; differences_sd in database
			z score	z score test statistic, z_score in database
			number of participants	number of subjects, i.e. number of subjects, can choose either group_1_size or group_2_size in the database - they should be the same

= all needed   = only one needed

7.3 How to do power analysis in G*Power

1. Choose the *t tests* test family

2. Select *Means: Wilcoxon signed-rank test (matched pairs)*
3. Select *A priori* for the type of power analysis
4. (a) **Using Data from Option 1**
   Enter the data into the following fields:

   Parent distribution Select a response distribution. If unsure, select *min ARE*, as this provides the most conservative estimate

   Effect size dz Cohen's d_z

   (b) **Using Data from Option 2**
   Enter the data into the following fields:

   Parent distribution Select a response distribution. If unsure, select *min ARE*, as this provides the most conservative estimate

   Effect size dz Determine the effect size d_z

   - Hit the *Determine* => button
   - Select *From group parameters* and enter the means and standard deviations of the two groups and the correlation coefficient for the correlation between groups
   - Hit *Calculate and transfer to main window*

   (c) **Using Data from Option 3**
   Enter the data into the following fields:

   Parent distribution Select a response distribution. If unsure, select *min ARE*, as this provides the most conservative estimate

   Effect size dz Determine the effect size d_z

   - Hit the *Determine* => button
   - Select *From differences* and enter the mean of difference and SD of difference
   - Hit *Calculate and transfer to main window*

   (d) **Using Data from Option 4**
   Enter the data into the following fields:

   Parent distribution Select a response distribution. If unsure, select *min ARE*, as this provides the most conservative estimate

   Effect size dz Determine the effect size d_z using the third formula in the terminology section
5. Calculate your power analysis

7.4 How to do power analysis in R

Since power analysis for the Wilcoxon signed-rank test makes distributional assumptions, even if the test itself does not, simulation is necessary to estimate necessary sample size. For a more simple solution, use G*Power [5].

References

[1] Thom Baguley. Standardized or simple effect size: What should be reported? *British Journal of Psychology*, 100(3):603–617, 2009.

[2] Jacob Cohen. *Statistical Power Analysis for the Behavioral Sciences*. L. Erlbaum Associates, Hillsdale, N.J, 2nd ed edition, 1988.

[3] Julian di Stephano. How much power is enough? Against the development of an arbitrary convention for statistical power calculations. *Functional Ecology*, 17(5):707–709, 2003.

[4] Paul D. Ellis. *The Essential Guide to Effect Sizes. Statistical Power, Meta-Analysis, and the Interpretation of Research Results*. Cambridge University Press, 2010.

[5] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2):175–191, May 2007.

[6] Robert Rosenthal. Parametric Measures of Effect Size. In Harris Cooper and Larry Hedges, editors, *The Handbook of Research Synthesis*, pages 231–244. Russel Sage Foundation, New York, 1994.