BUCSS Web Applications: Bias-Uncertainty Corrected Sample Size

AUTHORS

Samantha F. Anderson, Arizona State University samantha.f.anderson@asu.edu
Ken Kelley, University of Notre Dame, kkelley@nd.edu

PURPOSE

Each web application returns the necessary per-group sample size to achieve a desired level of statistical power for a planned study of a particular design based on information obtained from a prior study. The effect from the prior study can be corrected for publication bias and/or uncertainty to provide a sample size that will achieve more accurate statistical power for a planned study, when compared to approaches that use a sample effect size at face value or rely on sample size only. The app also outputs the corrected noncentrality parameter.

GENERAL OVERVIEW

Researchers often use the sample effect size from a prior study as an estimate of the population value of an expected future effect in sample size planning. However, sample effect size estimates should not usually be used at face value to plan sample size, due to both publication bias and uncertainty (for definitions, consult Anderson & Maxwell, 2017). The approach implemented in our web applications uses the observed *t*-value (or *F*-value) and sample size from a previous study to correct the noncentrality parameter associated with the effect of interest for publication bias and/or uncertainty. This new estimated noncentrality parameter is output and then used to calculate the necessary per-group sample size to achieve the desired level of power in the planned study. The approach uses a likelihood function of a truncated non-central *F* distribution, where the truncation occurs due to small effect sizes being unobserved due to publication bias. The numerator of the likelihood function is simply the density of a noncentral *F* distribution. The denominator is the power of the test, which serves to truncate the distribution. Thus, the ratio of the numerator and denominator is a truncated noncentral *F* distribution. (See Taylor & Muller, 1996, Equation 2.1).

Assurance is defined as the proportion of times that power will be at or above the desired level, if the sample size planning procedure were to be repeated many times, and each time new data were collected and analyzed (consult Anderson & Maxwell, 2017 for more on assurance). For example, assurance = .5 means that power will be above the desired level half of the time, but below the desired level the other half of the time. Selecting assurance = .5 (selecting the noncentrality parameter at the 50th percentile of the likelihood distribution) results in a median-unbiased estimate of the population noncentrality parameter and does not correct for uncertainty. In order to correct for uncertainty, assurance > .5 can be selected, which corresponds to selecting the noncentrality parameter associated with the (1 - assurance) quantile of the likelihood distribution. Thus, one could have, for example, 99% assurance that the study will yield 80% power. Setting assurance < .5 is not recommended.

If the prior study of interest has been subjected to publication bias, the α-level for the previous study can typically be set to .05. If the prior study has not been subjected to publication bias (e.g., a pilot study), the α-level of the previous study can be set to 1 to indicate no publication bias (meaning that the knowledge from the study was not subjected to an α-level in order to have seen the results). Alternative α-levels can also be accommodated to represent differing amounts of publication bias. For example, setting the previous study α to .20 would reflect less severe publication bias than the default of .05, because those results would be known provided that the *p*-value was less than .20. In essence, setting the prior study α at .20 assumes that studies with *p*-values less than .20 are published, whereas those with larger *p*-values are not.

In some cases, the corrected noncentrality parameter for a given level of assurance will be estimated to be zero. This is an indication that, at the desired level of assurance, the previous study's effect cannot be accurately estimated due to high levels of uncertainty and/or bias. When this happens, subsequent sample size planning is not possible with the chosen specifications. Two alternatives are recommended. First, users can select a lower value of assurance (e.g., .80 instead of .95). Second, users can reduce the influence of publication bias by setting the α-level of the prior study at a value greater than .05 (making less than a full adjustment for publication bias). We encourage users to make the adjustments as minimal as possible.

These web applications assume that the planned study will have equal *n* (cell sample sizes). Unequal *n* in the prior study is handled in the following way.
1. ANOVA designs with between-subjects factors: If the user enters an *N* (total sample size) that is not divisible by the number of between-subjects cells, the function calculates *n* by dividing *N* by the number of cells and both rounding up and rounding down the result. The suggested sample size for the planned study is calculated using both of these values of *n*, and the function returns the larger of these two suggestions, to be conservative.
2. Independent-*t*: If the user enters an odd value for *N*, no information is available on the exact group sizes. The function calculates *n* by dividing *N* by 2 and both rounding up and rounding down the result. The suggested sample size for the planned study is calculated using both of these values of *n*, and the function returns the larger of these two suggestions, to be conservative. If the user enters a vector for *n* with two different values, specific information is available on the exact group sizes. *n* is calculated as the harmonic mean of these two values (a measure of effective sample size). Again, this is rounded both up and down. The suggested sample size for the planned study is calculated using both of these values of *n*, and the function returns the larger of these two suggestions, to be conservative.

DESIGNS

Independent-*t*
Dependent-*t*
Between-subjects ANOVA – one or two way omnibus effects
Between-subjects ANOVA – general*
Within-subjects ANOVA – one or two way omnibus effects^
Within-subjects ANOVA – general*^
Split-Plot – two factors; omnibus effects^
Split-Plot – general*^
Multiple Linear Regression – test of single predictor
Multiple Linear Regression – test of joint predictors
Multiple Linear Regression – test of all predictors

* General functions can accommodate >2 factors, contrasts, etc.
^ Assumes sphericity for the within-subjects effects

RESOURCES

Anderson, S. F., & Kelley, K. (2017). BUCSS: Bias Uncertainty Corrected Sample Size (Version 0.0.2 or greater) [R Package]. Manual available at: https://cran.r-project.org/web/packages/BUCSS/index.html

Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample size planning for more accurate statistical power: A method adjusting sample effect sizes for uncertainty and publication bias. *Psychological Science, 28*, 1547-1562.

Anderson, S. F., & Maxwell, S. E. (2017). Addressing the 'replication crisis': Using original studies to design replication studies with appropriate statistical power. *Multivariate Behavioral Research, 52*, 305-324. doi:10.1080/00273171.2017.1289361

Taylor, D. J., & Muller, K. E. (1996). Bias in linear model power and sample size calculation due to estimating noncentrality. *Communications in Statistics: Theory and Methods, 25*, 1595-1610.