Supplement A

We provide R code below to illustrate how to compute power using an estimated effect size according to the methods of safeguard power (Perugini, Gallucci, & Costantini, 2014), the power calibrated effect size approach (PCES; McShane & Bökenholt, 2014), and the Bias Uncertainty Corrected Sample Size (BUCSS) method (Anderson & Maxwell, 2017; Taylor & Muller, 1996).

Consider a paired sample test for a mean difference with $\delta = .41$ (see Gignac & Szodorai, 2016). Suppose we collect a sample of $N = 150$ and conduct a two-sided t-test. Note that changing N to the values of 10, 20, 25, and 50 with `set.seed(1214)` will reproduce values in Table 1 for $R = 1$ study.

```
set.seed(1214)
N <- 150                    #sample size for each pair
delta <- .41                #population effect size
alpha <- .05                #type I error rate
y <- rnorm(N,delta,1)       #generated data
t.test(y)
```

```
##
##  One Sample t-test
##
## data:  y
## t = 3.2462, df = 149, p-value = 0.001445
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.1129977 0.4645802
## sample estimates:
## mean of x
## 0.2887889
```

Classical power for the procedure is calculated with the following code:

```
library(pwr)
pwr.t.test(n=N, d=delta, sig.level=alpha,type="one.sample",
           alternative="two.sided")
```

```
##
##      One-sample t test power calculation
##
##               n = 150
##               d = 0.41
##       sig.level = 0.05
##           power = 0.9987727
##     alternative = two.sided
```

Observed power is calculated with the following code, where estimated Cohen's d takes the place of the population δ.

```
pwr.t.test(n=N, d=mean(y)/sd(y), sig.level=alpha, type="one.sample",
           alternative="two.sided")
```

```
##
##      One-sample t test power calculation
##
##               n = 150
##               d = 0.2650498
##       sig.level = 0.05
```

```
##           power = 0.897105
##     alternative = two.sided
```

Safeguard power is calculated from the lower bound of a confidence interval (CI), called a percentile. The lower bound of a 60% CI is associated with the 20th percentile and the lower bound of a 80% CI is associated with a 10th percentile. By using the 20th percentile, calculated power is expected to be a lower bound for 80% of estimated effect sizes.

Safeguard power using the 20th and 10th percentiles are calculated with the following code:

```
percentile20 <- t.test(y, conf.level=.6)$conf.int[1]
percentile10 <- t.test(y, conf.level=.8)$conf.int[1]

pwr.t.test(n=N, d=percentile20, sig.level=alpha,type="one.sample",
           alternative="two.sided") #from lower bound of 60% CI
```

```
##
##      One-sample t test power calculation
##
##               n = 150
##               d = 0.2137009
##       sig.level = 0.05
##           power = 0.739054
##     alternative = two.sided
```

```
pwr.t.test(n=N, d=percentile10, sig.level=alpha,type="one.sample",
           alternative="two.sided") #from lower bound of 80% CI
```

```
##
##      One-sample t test power calculation
##
##               n = 150
##               d = 0.174271
##       sig.level = 0.05
##           power = 0.5638345
##     alternative = two.sided
```

PCESs are estimated effect sizes that have been calibrated for sampling variability as quantified by the standard error of estimate, $S.E.$. For calibration, a target assurance (e.g., 80% average power) is specified. Calculated N will then be associated with the target assurance from the distribution of power values.

The code below computes the PCES from closed form equations and then computes N for 80% assurance. These equations are presented in Table 2 of McShane and Böckenholt (2016). An web application to calculate N via the PCES method is available at https://blakemcshane.shinyapps.io/pces/.

```
Delta <- mean(y)            #estimated mean difference effect size
sigma2 <- var(y)            #variance of mean difference
nu2 <- t.test(y)$stderr^2   #squared standard error of estimate
zcrit <- qnorm(1-alpha/2)   #critical value for two-sided test
zbeta <- qnorm(.2)          #critical value of type II error

pces <- (zcrit*Delta+zbeta*sqrt(Delta^2+nu2*(zcrit^2-zbeta^2)))/(zcrit+zbeta)
pces                #power calibrated effect size
```

```
## [1] 0.2585794
```

```
7.85*sigma2/(pces^2) #N for 80% assurance
```

```
## [1] 139.376
```

BUCSS can be computed using the R package `BUCSS`. The code below is an example of how to calculate N for an effect size that has been adjusted for bias and uncertainty due to sampling variability. It is assumed that p-values in published research are censored at $p = .05$. The code takes in the t-test statistic as the input for `t.observed`, sample size of the original study for `N`, the Type I error for the test in `alpha.prior`, the Type I error for the planned test in `alpha.planned`, assurance level in `assurance`, and power for assurance in `power`.

When `assurance = .5` and `power = .8`, $N = 128$ for 50% of calculated power having values of 80% power or higher will be returned followed by the adjusted noncentrality parameter $\breve{\lambda} = 3.062$.

```
library(BUCSS)

ss.power.dt(t.observed=t.test(y)$statistic, N=N, alpha.prior=.05,
            alpha.planned=.05, assurance=.5, power=.8, step=.001)
```

```
## [[1]]
## [1] 128
##
## [[2]]
## [1] 3.062
```

When `assurance = .8` and `power = .8`, $N = 327$ for 80% of calculated power having values of 80% power or higher will be returned followed by the adjusted noncentrality parameter $\breve{\lambda} = 1.905$.

```
ss.power.dt(t.observed=t.test(y)$statistic, N=N, alpha.prior=.05,
            alpha.planned=.05, assurance=.8, power=.8, step=.001)
```

```
## [[1]]
## [1] 327
##
## [[2]]
## [1] 1.905
```

For more examples using the BUCSS package, see Anderson (2020). Users unfamiliar with R can conduct power analysis with the BUCSS approach using web applications at DesigningExperiments.com.

References

Anderson, S. F. (2020). Using prior information to plan appropriately powered regression studies: A tutorial using BUCSS. *Psychological Methods*. doi: 10.1037/met0000366

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*(3), 305–324. doi: 10.1080/00273171.2017.1289361

McShane, B. B., & Böckenholt, U. (2016). Planning sample sizes when effect sizes are uncertain: The power-calibrated effect size approach. *Psychological Methods*, *21*(1), 47–60. doi: 10.1037/met0000036

Perugini, M., Gallucci, M., & Costantini, G. (2014). Safeguard power as a protection against imprecise power estimates. *Perspectives on Psychological Science*, *9*(3), 319–332. doi: 10.1177/1745691614528519

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(7), 1595–1610. doi: 10.1080/03610929608831787