

Validation file for Equivalence Functions

```
library(knitr)
opts_chunk$set(tidy.opts=list(width.cutoff=60),tidy=TRUE)
```

Tests against Literature, existing packages, and MiniTab

Below I reproduce some examples from the literature, compare TOSTER against other R packages that have equivalence test functions, and against Minitab. MiniTab software has implemented equivalence tests and power analysis for one-sample, independent, and paired t-tests. It is excellent - by far the most user-friendly software for anyone who prefers a commercial statistics program over an open source solution (it also has equivalence tests for 2x2 crossover designs, but not for correlations or meta-analysis). It also allows users to enter summary statistics, which is used here to compare the output of the equivalence functions in TOSTER against minitab. Note that when $SD = 1$, the mean difference equals the standardized difference. The MiniTab output is followed by the R function.

TOST for two independent samples

Example 1.1

From page 196 of Stegner, B. L., Bostrom, A. G., & Greenfield, T. K. (1996). Equivalence testing for use in psychosocial and services research: An introduction with examples. *Evaluation and Program Planning*, 19(3), 193-198. [https://doi.org/10.1016/0149-7189\(96\)00011-0](https://doi.org/10.1016/0149-7189(96)00011-0)

Randomized controlled trial of two conditions of case management, with the main dependent variable being ratings on the Brief Psychiatric Rating Scale (BPRS). The means, sd's, and sample sizes are:

Control: $M = 1.5679$, $SD = 0.4285$, $n = 64$ Experimental: $M = 1.6764$, $SD = 0.4748$, $n = 70$

Calculated t-values for TOST are $t_1 = 5.383$ and $t_2 = -2.616$

These results can be reproduced with TOSTER package and the TOSTtwo.raw R code:

```
require(TOSTER)
```

```
## Loading required package: TOSTER
```

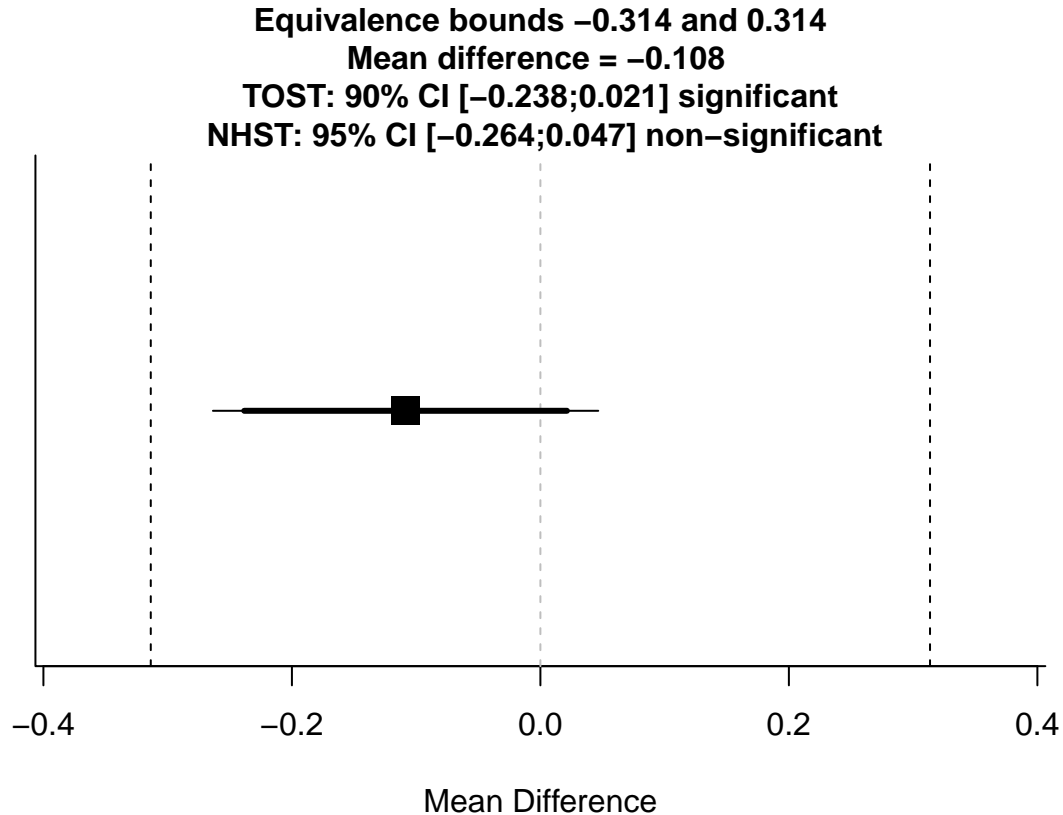
```
TOSTtwo.raw(m1 = 1.5679, m2 = 1.6764, sd1 = 0.4285, sd2 = 0.4748,
  n1 = 64, n2 = 70, low_eqbound = -0.3136, high_eqbound = 0.3136,
  alpha = 0.05, var.equal = TRUE)
```

```
## Using alpha = 0.05 Student's t-test was non-significant, t(132) = -1.384005, p = 0.1686925
```

```
##
```

```
## Using alpha = 0.05 the equivalence test based on Student's t-test was significant, t(132) = -2.61621
```

```
##
```



```
## TOST results:
##   t-value 1    p-value 1 t-value 2    p-value 2    df
## 1  5.384227 1.612713e-07 -2.616216 0.004963596 132
##
## Equivalence bounds (raw scores):
##   low bound raw high bound raw
## 1      -0.3136      0.3136
##
## TOST confidence interval:
##   Lower Limit 90% CI raw Upper Limit 90% CI raw
## 1           -0.2383608           0.02136078
```

Example 1.2

From page 225 of Limentani, G. B., Ringo, M. C., Ye, F., Bergquist, M. L., & MCSorley, E. O. (2005). Beyond the t-test: statistical equivalence testing. *Analytical Chemistry*, 77(11), 221-A.

Reproducing the 90% CI and significant equivalence test.

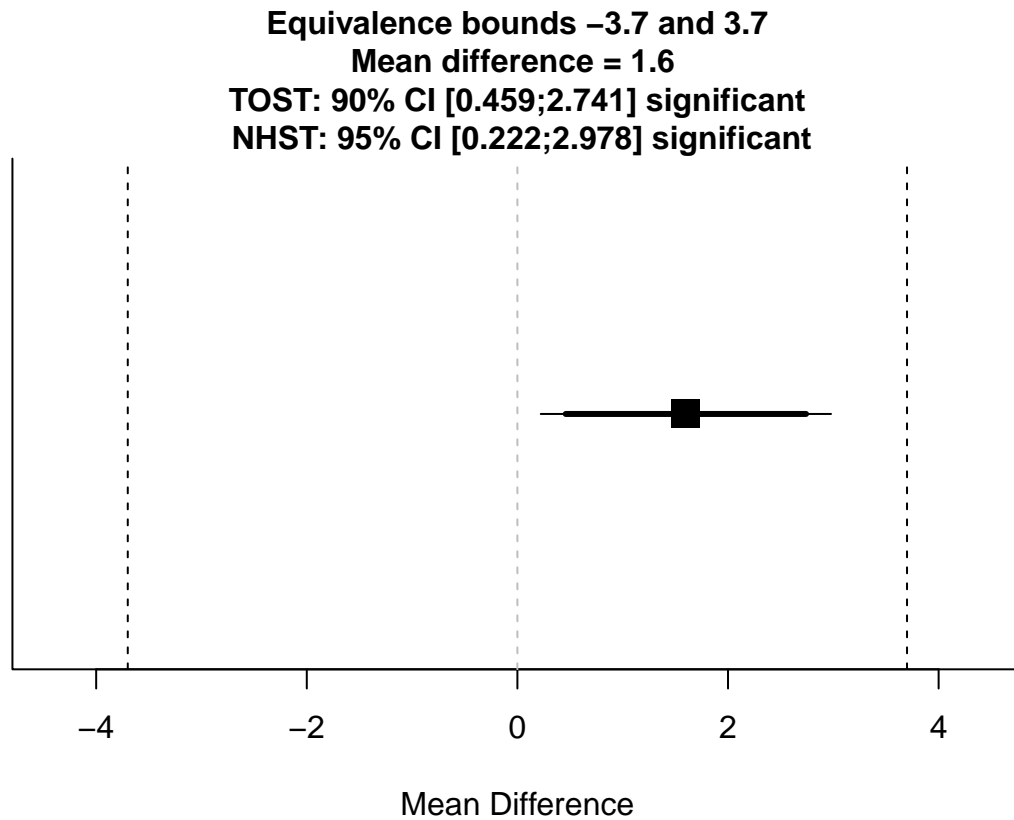
```
TOSTtwo.raw(m1 = 89.3, m2 = 87.7, sd1 = 1.9, sd2 = 1.3, n1 = 12,
            n2 = 12, low_eqbound = -3.7, high_eqbound = 3.7, alpha = 0.05,
            var.equal = TRUE)
```

```
## Using alpha = 0.05 Student's t-test was significant, t(22) = 2.407535, p = 0.02488925
```

```
##
```

```
## Using alpha = 0.05 the equivalence test based on Student's t-test was significant, t(22) = -3.15989,
```

##



```
## TOST results:
##   t-value 1    p-value 1 t-value 2    p-value 2 df
## 1  7.974961 3.109178e-08 -3.15989 0.002270472 22
##
## Equivalence bounds (raw scores):
##   low bound raw high bound raw
## 1      -3.7          3.7
##
## TOST confidence interval:
##   Lower Limit 90% CI raw Upper Limit 90% CI raw
## 1           0.4588201           2.74118
```

Example 1.3

From Meyners, M. (2012). Equivalence tests - A review. Food Quality and Preference, 26(2), 231-245. <https://doi.org/10.1016/j.foodqual.2012.05.003>, page 237:

To illustrate the approach, we consider the data presented by Bi (2005) of 100 consumers each from two cities A and B, giving overall liking scores on a 9-point scale. The equivalence margin is set to $d = 0.5$, i.e. a difference of 0.5 or less in mean values is considered negligible. With mean values of $\bar{I}_A = 7.1$ and $\bar{I}_B = 6.9$ and variances $s^2_A = 2.0$ and $s^2_B = 2.2$, the corresponding t-statistic is 0.976, and the p-values for the two one-sided test problems become $p_1 = 0.0717$ and $p_2 = 0.0003$.

We can use the TOST procedure for an independent t-test to reproduce the p-values and 90% CI:

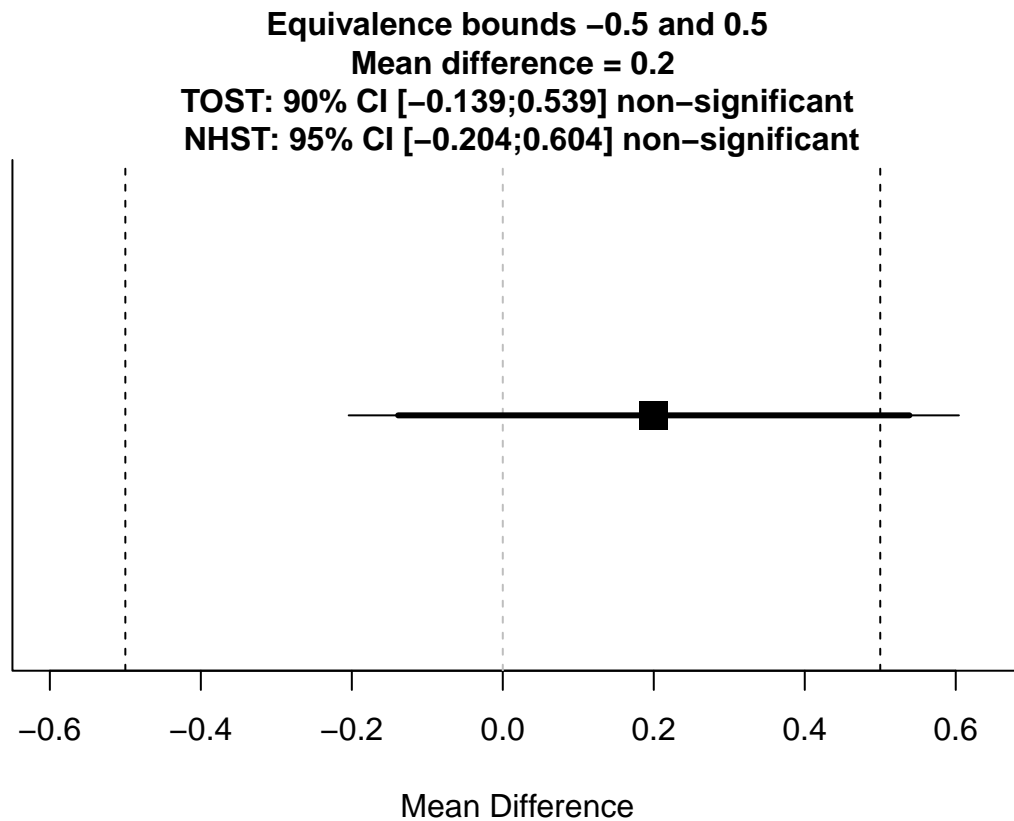
```
TOSTtwo.raw(m1 = 7.1, m2 = 6.9, sd1 = sqrt(2), sd2 = sqrt(2.2),
  n1 = 100, n2 = 100, low_eqbound = -0.5, high_eqbound = 0.5,
  alpha = 0.05, var.equal = TRUE)
```

```
## Using alpha = 0.05 Student's t-test was non-significant, t(198) = 0.9759001, p = 0.3303048
```

```
##
```

```
## Using alpha = 0.05 the equivalence test based on Student's t-test was non-significant, t(198) = -1.46385, p = 0.07241035
```

```
##
```



```
## TOST results:
```

```
##   t-value 1   p-value 1 t-value 2   p-value 2   df
## 1    3.41565 0.0003859877  -1.46385 0.07241035 198
##
```

```
## Equivalence bounds (raw scores):
```

```
##   low bound raw high bound raw
## 1      -0.5         0.5
##
```

```
## TOST confidence interval:
```

```
##   Lower Limit 90% CI raw Upper Limit 90% CI raw
## 1      -0.1386793         0.5386793
```

This outcome matches the result of (page 237): The 90% CI is [0.139; 0.539] with a p-value of approximately 0.07.

Example 1.4

From Table 1 in Quertemont, E. (2011). How to Statistically Show the Absence of an Effect. *Psychologica Belgica*, 51(2), 109. <https://doi.org/10.5334/pb-51-2-109>

Example 1: $t(8) = 1.52$; $p = 0.92$ Example 2: $t(298) = -2.23$; $p = 0.013$

Example 1

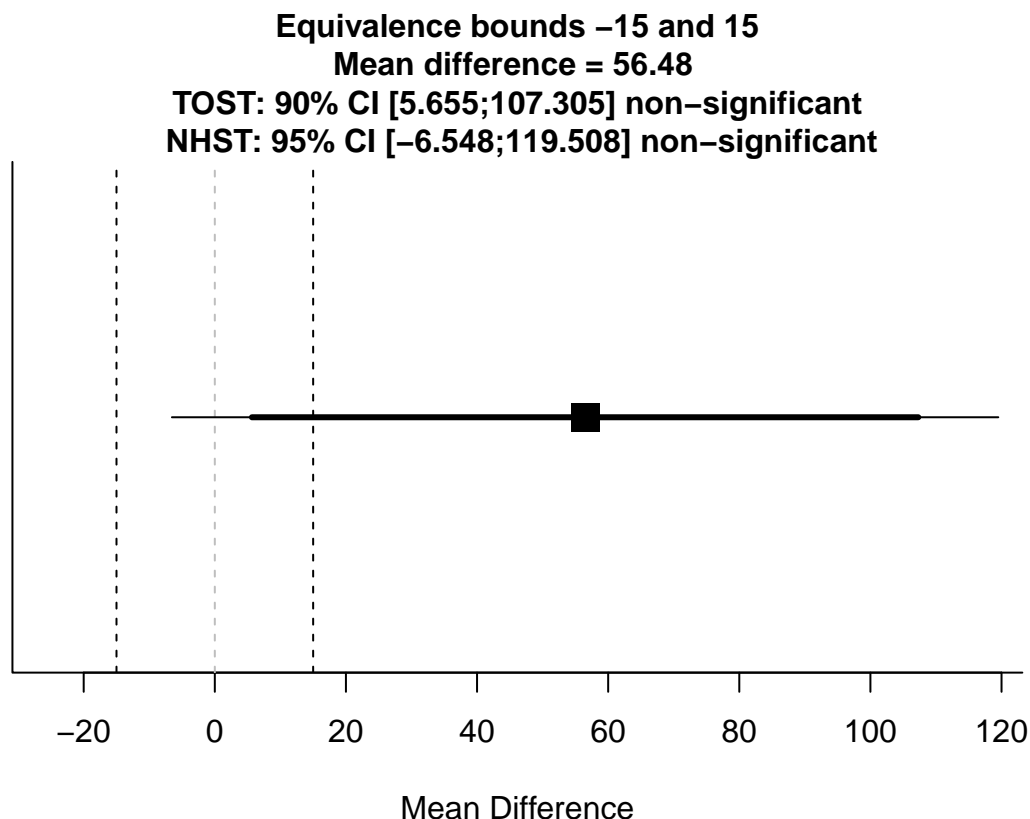
```
TOSTtwo.raw(m1 = 459.09, m2 = 402.61, sd1 = 47.53, sd2 = 38.42,  
  n1 = 5, n2 = 5, low_eqbound = -15, high_eqbound = 15, alpha = 0.05,  
  var.equal = TRUE)
```

```
## Using alpha = 0.05 Student's t-test was non-significant,  $t(8) = 2.066441$ ,  $p = 0.07262966$ 
```

```
##
```

```
## Using alpha = 0.05 the equivalence test based on Student's t-test was non-significant,  $t(8) = 1.517634$ 
```

```
##
```



```
## TOST results:
```

```
##   t-value 1  p-value 1 t-value 2 p-value 2 df
```

```
## 1  2.615247 0.01543878  1.517634  0.916208  8
```

```
##
```

```
## Equivalence bounds (raw scores):
```

```
##   low bound raw high bound raw
```

```
## 1         -15          15
```

```
##
```

```
## TOST confidence interval:
```

```
## Lower Limit 90% CI raw Upper Limit 90% CI raw
## 1 5.654791 107.3052
```

Example 2

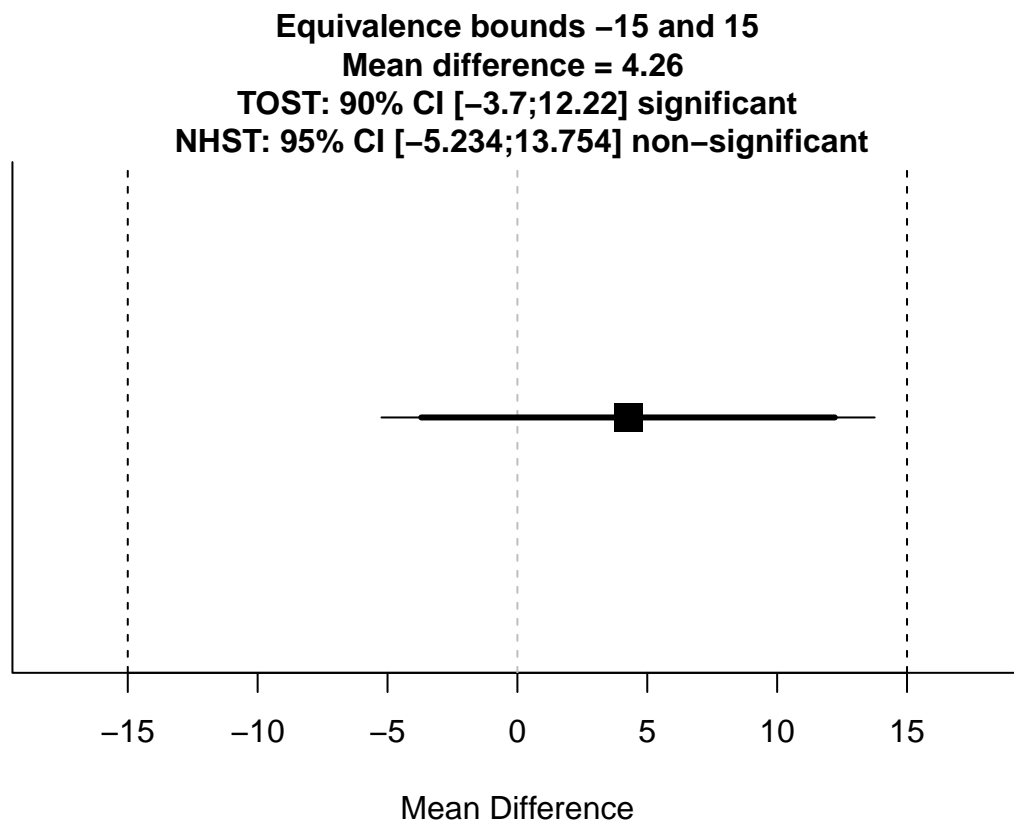
```
TOSTtwo.raw(m1 = 407.24, m2 = 402.98, sd1 = 41.77, sd2 = 41.79,
  n1 = 150, n2 = 150, low_eqbound = -15, high_eqbound = 15,
  alpha = 0.05, var.equal = TRUE)
```

```
## Using alpha = 0.05 Student's t-test was non-significant, t(298) = 0.8830225, p = 0.3779359
```

```
##
```

```
## Using alpha = 0.05 the equivalence test based on Student's t-test was significant, t(298) = -2.226212
```

```
##
```



```
## TOST results:
```

```
## t-value 1 p-value 1 t-value 2 p-value 2 df
## 1 3.992257 4.123961e-05 -2.226212 0.0133738 298
```

```
##
```

```
## Equivalence bounds (raw scores):
```

```
## low bound raw high bound raw
## 1 -15 15
```

```
##
```

```
## TOST confidence interval:
```

```
## Lower Limit 90% CI raw Upper Limit 90% CI raw
## 1 -3.700077 12.22008
```

These results perfectly match those in Table 1 of Quertemont (2011).

Example 1.5

MiniTab by default uses Welch's t-test when testing for equivalence, which does not require the assumption of equal variances, which matters with unequal sample sizes. TOSTER also presents Welch's t-test by default. Note that minitab rounds degrees of freedom down to a whole number. This is a conservative procedure. TOSTER does not. The difference is rarely noticable.

MiniTab Result Two-Sample Equivalence Test

*Method

Test mean = mean of Test Reference mean = mean of Reference Equal variances were not assumed for the analysis.

Descriptive Statistics

Variable N Mean StDev SE Mean Test 86 2,5 1,3 0,14018 Reference 102 2,3 1,1 0,10892

Difference: Mean(Test) - Mean(Reference)

Difference SE 95% CI Equivalence Interval 0,20000 0,17752 (-0,093626; 0,49363) (-0,4; 0,4)

CI is not within the equivalence interval. Cannot claim equivalence.

Test

Null hypothesis: Difference \leq -0,4 or Difference \geq 0,4 Alternative hypothesis: -0,4 < Difference < 0,4 alpha level: 0,05

Null Hypothesis DF T-Value P-Value Difference \leq -0,4 167 3,3799 0,000 Difference \geq 0,4 167 -1,1266 0,131

The greater of the two P-Values is 0,131. Cannot claim equivalence.

Reproduce results in TOSTER

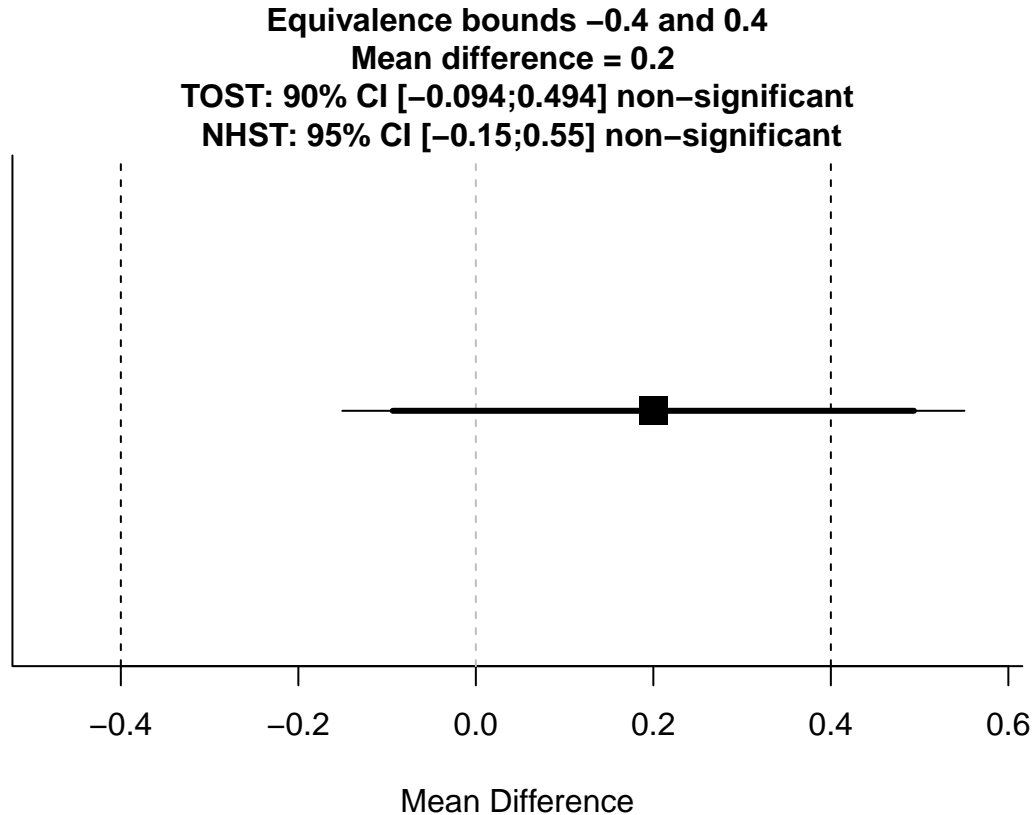
```
TOSTtwo.raw(m1 = 2.5, m2 = 2.3, sd1 = 1.3, sd2 = 1.1, n1 = 86,
            n2 = 102, low_eqbound = -0.4, high_eqbound = 0.4, alpha = 0.05)
```

```
## Using alpha = 0.05 Welch's t-test was non-significant, t(167.2924) = 1.126624, p = 0.2615151
```

```
##
```

```
## Using alpha = 0.05 the equivalence test based on Welch's t-test was non-significant, t(167.2924) =
```

```
##
```



```
## TOST results:
##   t-value 1    p-value 1 t-value 2 p-value 2      df
## 1  3.379871 0.0004512547 -1.126624 0.1307575 167.2924
##
## Equivalence bounds (raw scores):
##   low bound raw high bound raw
## 1      -0.4         0.4
##
## TOST confidence interval:
##   Lower Limit 90% CI raw Upper Limit 90% CI raw
## 1      -0.09362298      0.493623
```

Example 1.6

Now for the same data as example 1.5, but assuming equal variances:

Results MiniTab Two-Sample Equivalence Test

Method

Test mean = mean of Test Reference mean = mean of Reference Equal variances were assumed for the analysis.

Descriptive Statistics

Variable N Mean StDev SE Mean Test 86 2,5 1,3 0,14018 Reference 102 2,3 1,1 0,10892

Pooled StDev = 1,19556

Difference: Mean(Test) - Mean(Reference)

Difference SE 95% CI Equivalence Interval 0,20000 0,17502 (-0,089331; 0,48933) (-0,4; 0,4)

CI is not within the equivalence interval. Cannot claim equivalence.

Test

Null hypothesis: Difference \leq or $=$ -0,4 or Difference \geq or $=$ 0,4 Alternative hypothesis: -0,4 < Difference < 0,4 alpha level: 0,05

Null Hypothesis DF T-Value P-Value Difference \leq or $=$ -0,4 186 3,4281 0,000 Difference \geq or $=$ 0,4 186 -1,1427 0,127

The greater of the two P-Values is 0,127. Cannot claim equivalence.

Reproducing results with TOSTER

```
TOSTtwo.raw(m1 = 2.5, m2 = 2.3, sd1 = 1.3, sd2 = 1.1, n1 = 86,  
            n2 = 102, low_eqbound = -0.4, high_eqbound = 0.4, alpha = 0.05,  
            var.equal = TRUE)
```

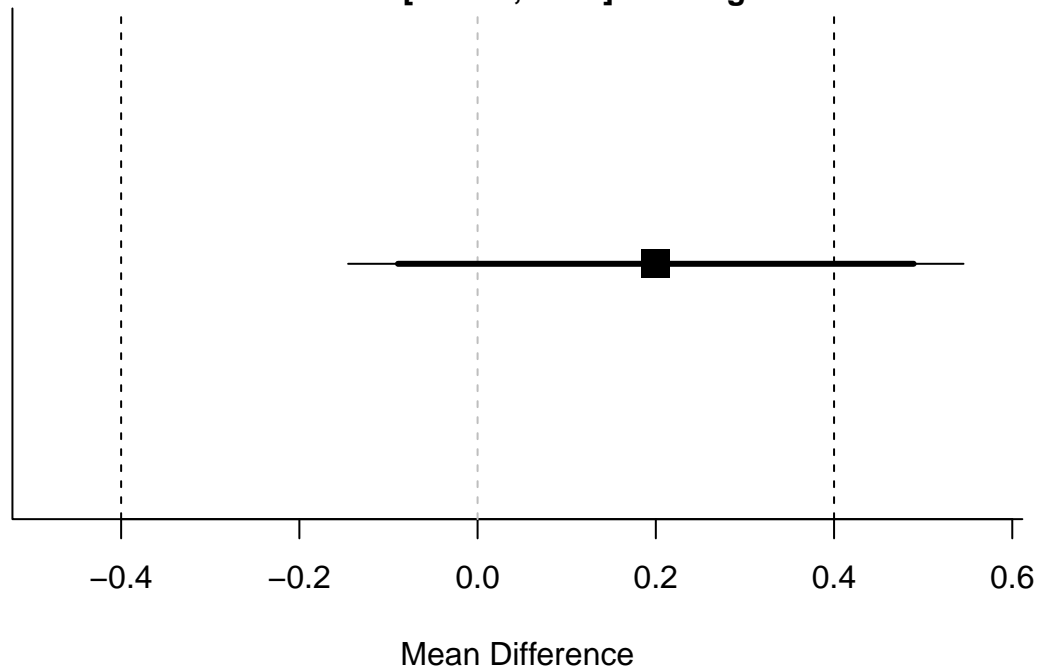
```
## Using alpha = 0.05 Student's t-test was non-significant, t(186) = 1.142695, p = 0.254634
```

```
##
```

```
## Using alpha = 0.05 the equivalence test based on Student's t-test was non-significant, t(186) = -1.1427
```

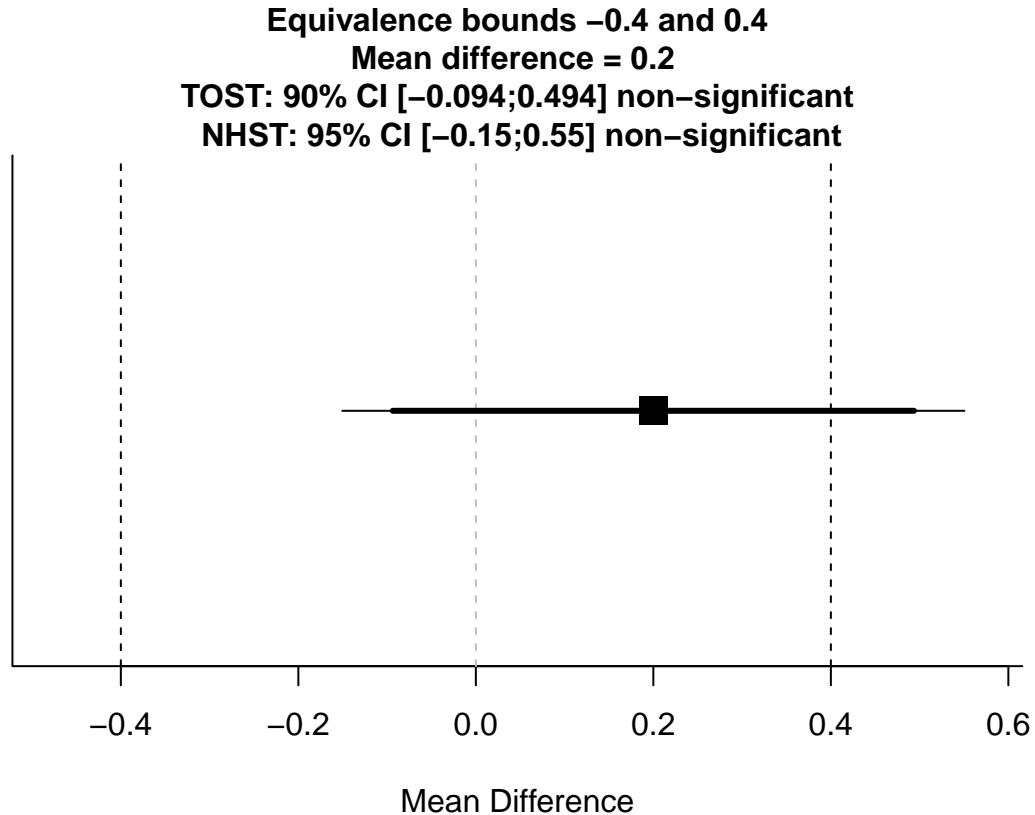
```
##
```

Equivalence bounds -0.4 and 0.4
Mean difference = 0.2
TOST: 90% CI [-0.089;0.489] non-significant
NHST: 95% CI [-0.145;0.545] non-significant



```
## TOST results:
##   t-value 1    p-value 1 t-value 2 p-value 2  df
## 1  3.428085 0.0003742019 -1.142695  0.127317 186
##
## Equivalence bounds (raw scores):
##   low bound raw high bound raw
## 1      -0.4      0.4
##
## TOST confidence interval:
##   Lower Limit 90% CI raw Upper Limit 90% CI raw
## 1      -0.08933128      0.4893313
TOSTtwo.raw(m1 = 2.5, m2 = 2.3, sd1 = 1.3, sd2 = 1.1, n1 = 86,
            n2 = 102, low_eqbound = -0.4, high_eqbound = 0.4, alpha = 0.05)

## Using alpha = 0.05 Welch's t-test was non-significant, t(167.2924) = 1.126624, p = 0.2615151
##
## Using alpha = 0.05 the equivalence test based on Welch's t-test was non-significant, t(167.2924) = 1.126624, p = 0.2615151
##
```



```
## TOST results:
##   t-value 1    p-value 1 t-value 2 p-value 2      df
## 1  3.379871 0.0004512547 -1.126624 0.1307575 167.2924
##
## Equivalence bounds (raw scores):
##   low bound raw high bound raw
## 1      -0.4         0.4
##
## TOST confidence interval:
##   Lower Limit 90% CI raw Upper Limit 90% CI raw
## 1      -0.09362298      0.493623
```

Paired t-test

Example 2.1

We can compare the `TOSTpaired.raw` function against the equivalence package.

Paired t-test in equivalence package (from example in `tost` function) and `TOSTpaired.raw`.

```
require(equivalence)
```

```
## Loading required package: equivalence
## Loading required package: lattice
## Loading required package: boot
```

```

##
## Attaching package: 'boot'

## The following object is masked from 'package:lattice':
##
##      melanoma

## Loading required package: PairedData
## Loading required package: MASS
## Loading required package: gld
## Loading required package: mvtnorm
## Loading required package: ggplot2

##
## Attaching package: 'PairedData'

## The following object is masked from 'package:base':
##
##      summary

data(ufc)
# Remove missing data
ufc <- ufc[complete.cases(ufc[, 9:10]), ]

tost(ufc$Height.m.p, ufc$Height.m, epsilon = 1, paired = TRUE)

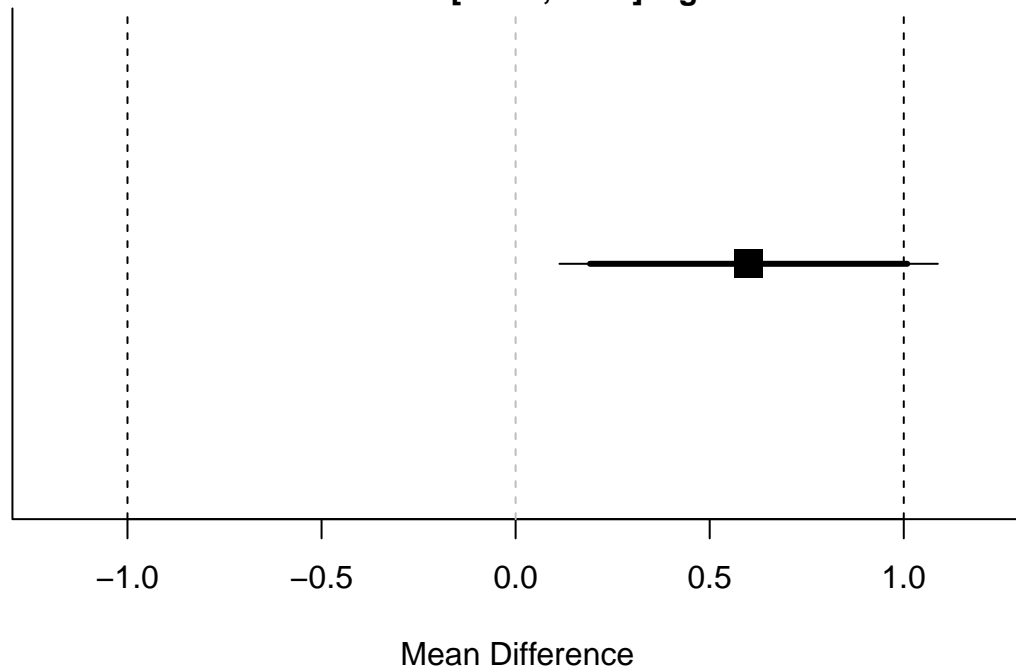
##
## Paired TOST
##
## data:  ufc$Height.m.p and ufc$Height.m
## df = 376
## sample estimates:
## mean of the differences
##          0.6003913
##
## Epsilon: 1
## 95 percent two one-sided confidence interval (TOST interval):
##  0.1916419 1.0091407
## Null hypothesis of statistical difference is: not rejected
## TOST p-value: 0.05389588

TOSTpaired.raw(n = length(ufc$Height.m.p), m1 = mean(ufc$Height.m.p,
  na.rm = TRUE), m2 = mean(ufc$Height.m, na.rm = TRUE), sd1 = sd(ufc$Height.m.p,
  na.rm = TRUE), sd2 = sd(ufc$Height.m, na.rm = TRUE), r12 = cor(ufc$Height.m,
  ufc$Height.m.p, use = "pairwise.complete.obs"), low_eqbound = -1,
  high_eqbound = 1, alpha = 0.05)

## Using alpha = 0.05 the NHST t-test was significant, t(376) = 2.42201, p = 0.01590709
##
## Using alpha = 0.05 the equivalence test was non-significant, t(376) = -1.612042, p = 0.05389588
##

```

Equivalence bounds -1 and 1
Mean difference = 0.6
TOST: 90% CI [0.192;1.009] non-significant
NHST: 95% CI [0.113;1.088] significant



```
## TOST results:
##   t-value 1  p-value 1 t-value 2   p-value 2  df
## 1 -1.612042 0.05389588  6.456061 1.660469e-10 376
##
## Equivalence bounds (raw scores):
##   low bound raw high bound raw
## 1          -1          1
##
## TOST confidence interval:
##   Lower Limit 90% CI raw Upper Limit 90% CI raw
## 1           0.1916419           1.009141
```

Example 2.2

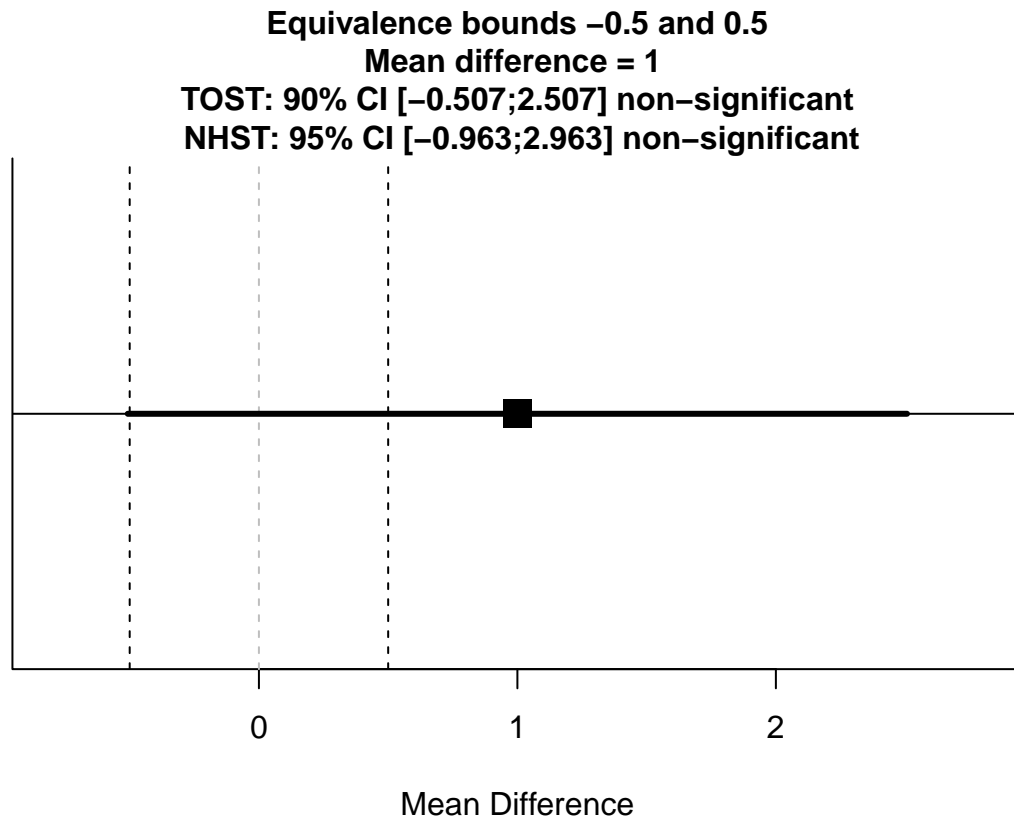
Quertemont also provides an example for paired t-tests in the appendix (page 124): $M_{dif} = 1$, $SD_{dif} = 1.58$, $d = .5$, $t = 0.71$, $df = 4$, $p = .74$ with a 95% CI from -0.96 to + 2.96.

```
morning <- c(3, 4, 4, 5, 4)
evening <- c(1, 4, 1, 4, 5)
```

```
TOSTpaired.raw(n = 5, m1 = mean(morning), m2 = mean(evening),
  sd1 = sd(morning), sd2 = sd(evening), r12 = cor(morning,
    evening), low_eqbound = -0.5, high_eqbound = 0.5, alpha = 0.05)
```

```
## Using alpha = 0.05 the NHST t-test was non-significant, t(4) = 1.414214, p = 0.2301996
##
```

```
## Using alpha = 0.05 the equivalence test was non-significant, t(4) = 0.7071068, p = 0.7407407
##
```



```
## TOST results:
##   t-value 1 p-value 1 t-value 2  p-value 2 df
## 1 0.7071068 0.7407407   2.12132 0.05059575  4
##
## Equivalence bounds (raw scores):
##   low bound raw high bound raw
## 1         -0.5          0.5
##
## TOST confidence interval:
##   Lower Limit 90% CI raw Upper Limit 90% CI raw
## 1          -0.5074433          2.507443
```

This output matches the calculations in Quertemont (2011).

Example 2.3

MiniTab does not allow users to perform equivalence tests for paired samples t-test based on summary data, but this equivalence test can be calculated based on the raw data.

Data: DV1<-c(4,2,4,3,5,4,3,4,5,4,2,3,4,5,5) DV2<-c(3,4,2,3,5,4,3,4,5,3,4,3,2,3,4)

Results MiniTab Equivalence Test with Paired Data: DV1, DV2

Method

Test mean = mean of DV1 Reference mean = mean of DV2

Descriptive Statistics

Variable N Mean StDev SE Mean DV1 15 3,8000 1,0142 0,26186 DV2 15 3,4667 0,91548 0,23637

Difference: Mean(DV1) - Mean(DV2)

Difference StDev SE 95% CI Equivalence Interval 0,33333 1,23443 0,31873 (-0,22804; 0,89471) (-0,8; 0,8)

CI is not within the equivalence interval. Cannot claim equivalence.

Test

Null hypothesis: Difference \leq or $=$ -0,8 or Difference \geq or $=$ 0,8 Alternative hypothesis: $-0,8 < \text{Difference} < 0,8$ alpha level: 0,05

Null Hypothesis DF T-Value P-Value Difference \leq or $=$ -0,8 14 3,5558 0,002 Difference \geq or $=$ 0,8 14 -1,4642 0,083

The greater of the two P-Values is 0,083. Cannot claim equivalence.

Reproduce Results with TOSTER

```
DV1 <- c(4, 2, 4, 3, 5, 4, 3, 4, 5, 4, 2, 3, 4, 5, 5)
DV2 <- c(3, 4, 2, 3, 5, 4, 3, 4, 5, 3, 4, 3, 2, 3, 4)
```

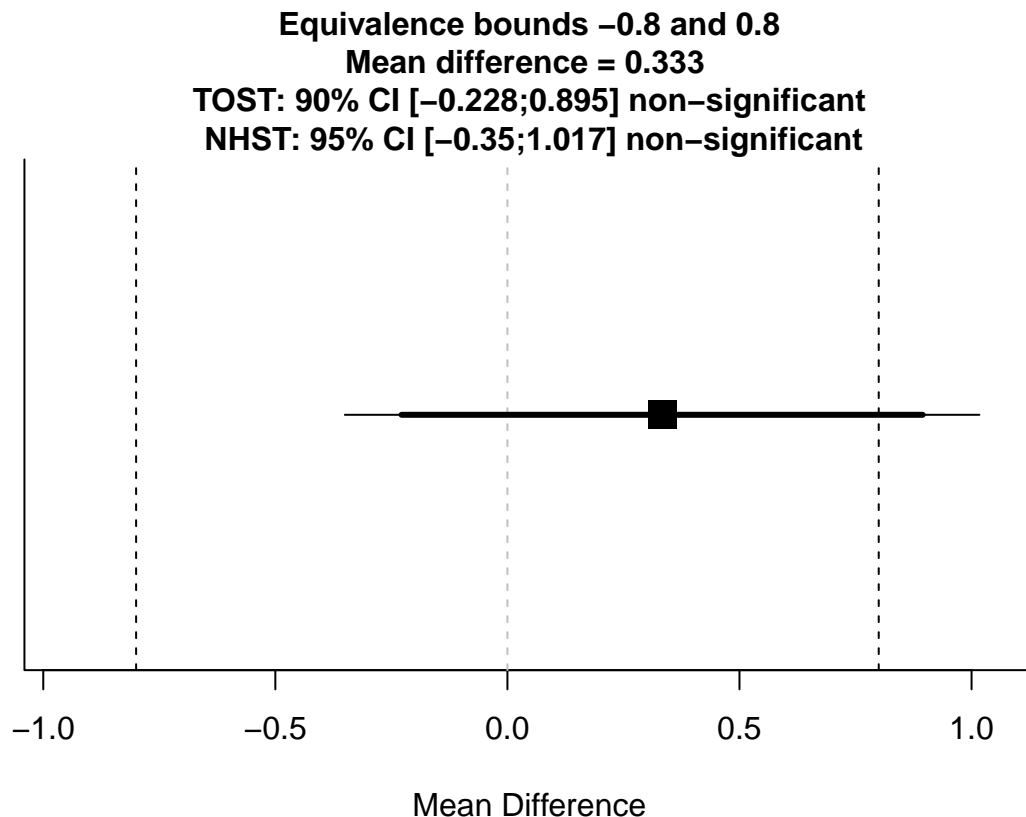
```
TOSTpaired.raw(n = length(DV1), m1 = mean(DV1), m2 = mean(DV2),
  sd1 = sd(DV1), sd2 = sd(DV2), r12 = cor(DV1, DV2), low_eqbound = -0.8,
  high_eqbound = 0.8, alpha = 0.05)
```

```
## Using alpha = 0.05 the NHST t-test was non-significant, t(14) = 1.045825, p = 0.3133636
```

```
##
```

```
## Using alpha = 0.05 the equivalence test was non-significant, t(14) = -1.464155, p = 0.08262144
```

```
##
```



```
## TOST results:
##   t-value 1  p-value 1 t-value 2  p-value 2 df
## 1 -1.464155 0.08262144  3.555805 0.00158212 14
##
## Equivalence bounds (raw scores):
##   low bound raw high bound raw
## 1      -0.8      0.8
##
## TOST confidence interval:
##   Lower Limit 90% CI raw Upper Limit 90% CI raw
## 1      -0.2280449      0.8947115
```

One-sample t-test

Example 3.1

Result MiniTab One-Sample Equivalence Test

Method

Target = 0

Descriptive Statistics

N Mean StDev SE Mean 100 -0,3 1 0,10000

Difference: Mean - Target

Difference SE 95% CI Equivalence Interval -0,30000 0,10000 (-0,46604; 0) (-0,5; 0,5)

CI is within the equivalence interval. Can claim equivalence.

Test

Null hypothesis: Difference \leq -0,5 or Difference \geq 0,5 Alternative hypothesis: $-0,5 < \text{Difference} < 0,5$ alpha level: 0,05

Null Hypothesis DF T-Value P-Value Difference \leq -0,5 99 2,0000 0,024 Difference \geq 0,5 99 -8,0000 0,000

The greater of the two P-Values is 0,024. Can claim equivalence.

Reproduce with TOSTER:

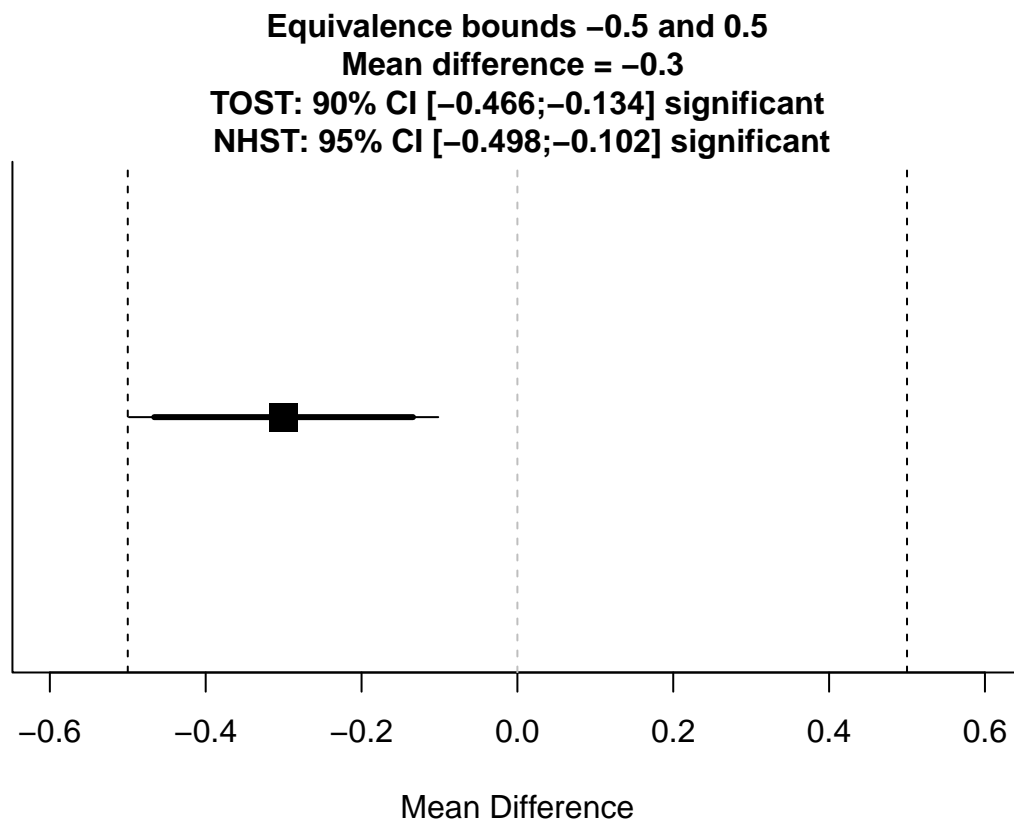
```
TOSTone.raw(m = -0.3, mu = 0, sd = 1, n = 100, low_eqbound = -0.5,  
            high_eqbound = 0.5, alpha = 0.05)
```

```
## Using alpha = 0.05 the NHST one-sample t-test was significant, t(99) = -3, p = 0.003415508
```

```
##
```

```
## Using alpha = 0.05 the equivalence test was significant, t(99) = 2, p = 0.02411985
```

```
##
```



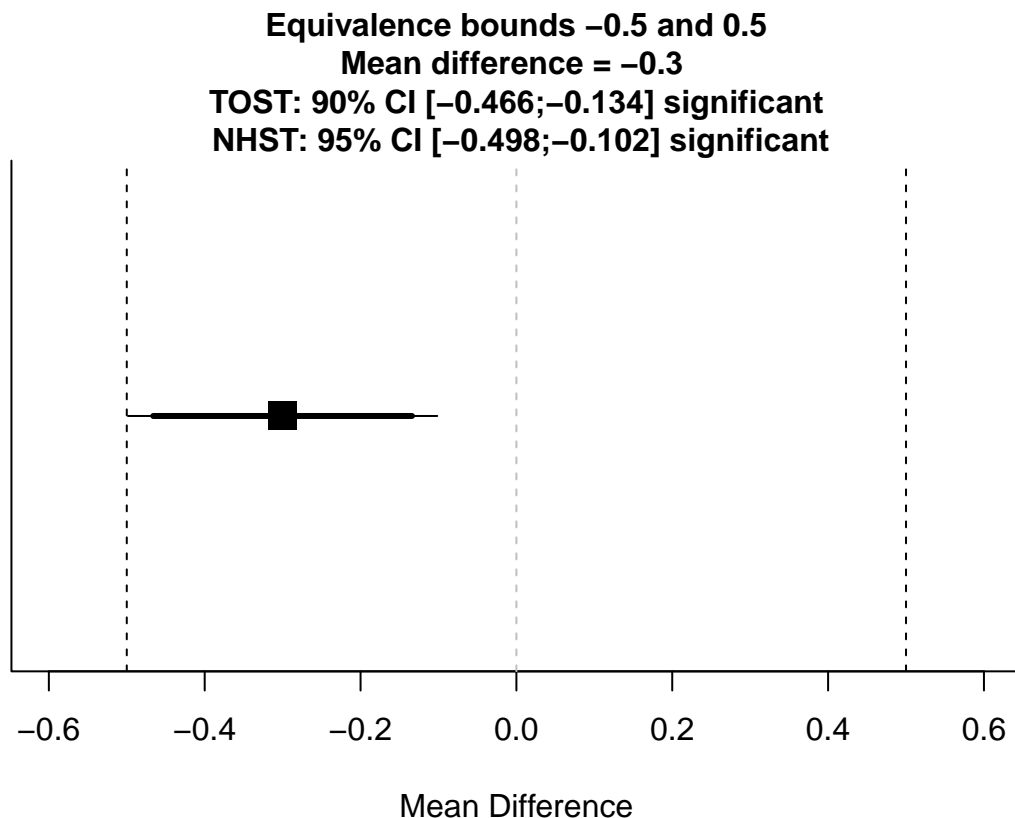
```
## TOST results:
```

```
##   t-value 1  p-value 1 t-value 2    p-value 2 df
## 1          2 0.02411985      -8 1.200152e-12 99
##
## Equivalence bounds (raw scores):
##   low bound raw high bound raw
## 1          -0.5          0.5
##
## TOST confidence interval:
##   Lower Limit 90% CI raw Upper Limit 90% CI raw
## 1          -0.4660391      -0.1339609
```

Which should be identical to the TOSTone test (because when $SD = 1$, unstandardized scores equal standardized scores)

```
TOSTone(m = -0.3, mu = 0, sd = 1, n = 100, low_eqbound_d = -0.5,
        high_eqbound_d = 0.5, alpha = 0.05)
```

```
## Using alpha = 0.05 the NHST one-sample t-test was significant, t(99) = -3, p = 0.003415508
##
## Using alpha = 0.05 the equivalence test was significant, t(99) = 2, p = 0.02411985
##
```



```
## TOST results:
##   t-value 1  p-value 1 t-value 2    p-value 2 df
## 1          2 0.02411985      -8 1.200152e-12 99
##
```

```
## Equivalence bounds (Cohen's d):
##   low bound d high bound d
## 1      -0.5      0.5
##
## Equivalence bounds (raw scores):
##   low bound raw high bound raw
## 1      -0.5      0.5
##
## TOST confidence interval:
##   Lower Limit 90% CI raw Upper Limit 90% CI raw
## 1      -0.4660391      -0.1339609
```

Example 3.2

MiniTab Result One-Sample Equivalence Test

Method

Target = 0

Descriptive Statistics

N Mean StDev SE Mean 100 1 1,2 0,12000

Difference: Mean - Target

Difference SE 95% CI Equivalence Interval 1,0000 0,12000 (0; 1,1992) (-0,8; 0,8)

CI is not within the equivalence interval. Cannot claim equivalence.

Test

Null hypothesis: Difference < or = -0,8 or Difference > or = 0,8 Alternative hypothesis: -0,8 < Difference < 0,8 alpha level: 0,05

Null Hypothesis DF T-Value P-Value Difference < or = -0,8 99 15,000 0,000 Difference > or = 0,8 99 1,6667 0,951

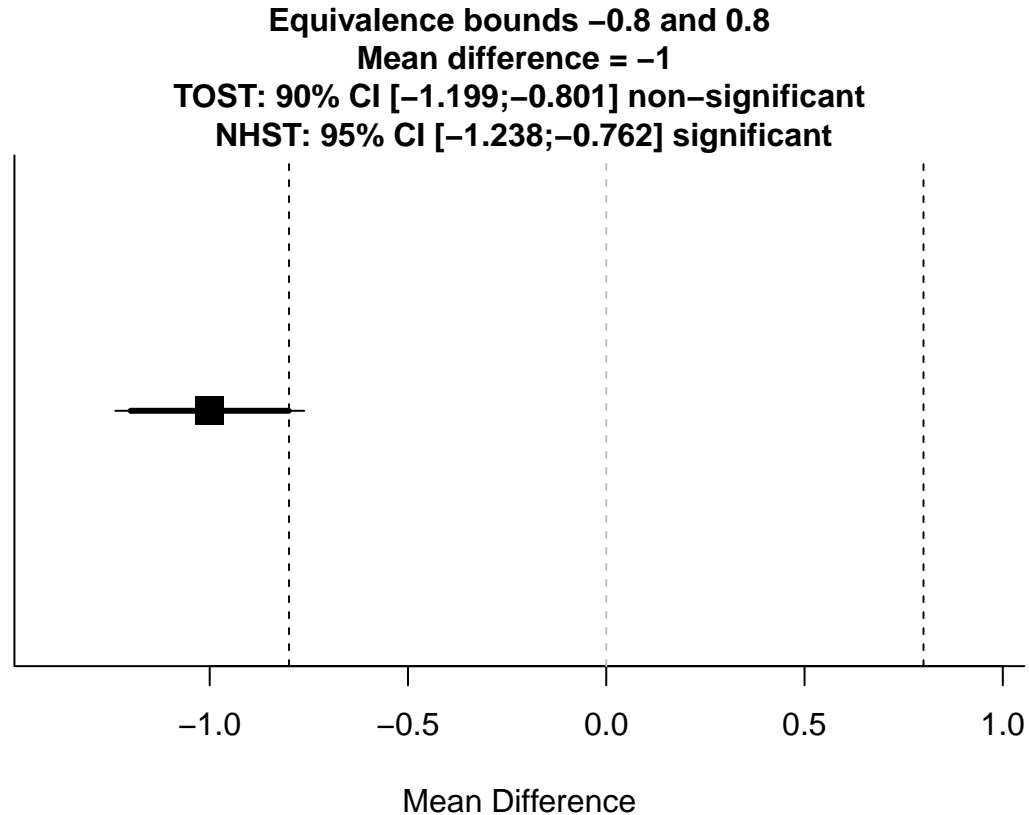
The greater of the two P-Values is 0,951. Cannot claim equivalence.

Reproducing the results with TOSTER

```
TOSTone.raw(m = -1, mu = 0, sd = 1.2, n = 100, low_eqbound = -0.8,
  high_eqbound = 0.8, alpha = 0.05)
```

```
## Using alpha = 0.05 the NHST one-sample t-test was significant, t(99) = -8.333333, p = 4.611473e-13
##
```

```
## Using alpha = 0.05 the equivalence test was non-significant, t(99) = -1.666667, p = 0.9506292
##
```



```
## TOST results:
##   t-value 1 p-value 1 t-value 2    p-value 2 df
## 1 -1.666667 0.9506292      -15 1.553349e-27 99
##
## Equivalence bounds (raw scores):
##   low bound raw high bound raw
## 1      -0.8      0.8
##
## TOST confidence interval:
##   Lower Limit 90% CI raw Upper Limit 90% CI raw
## 1      -1.199247      -0.8007531
```

Equivalence test for correlations

Example 4.1

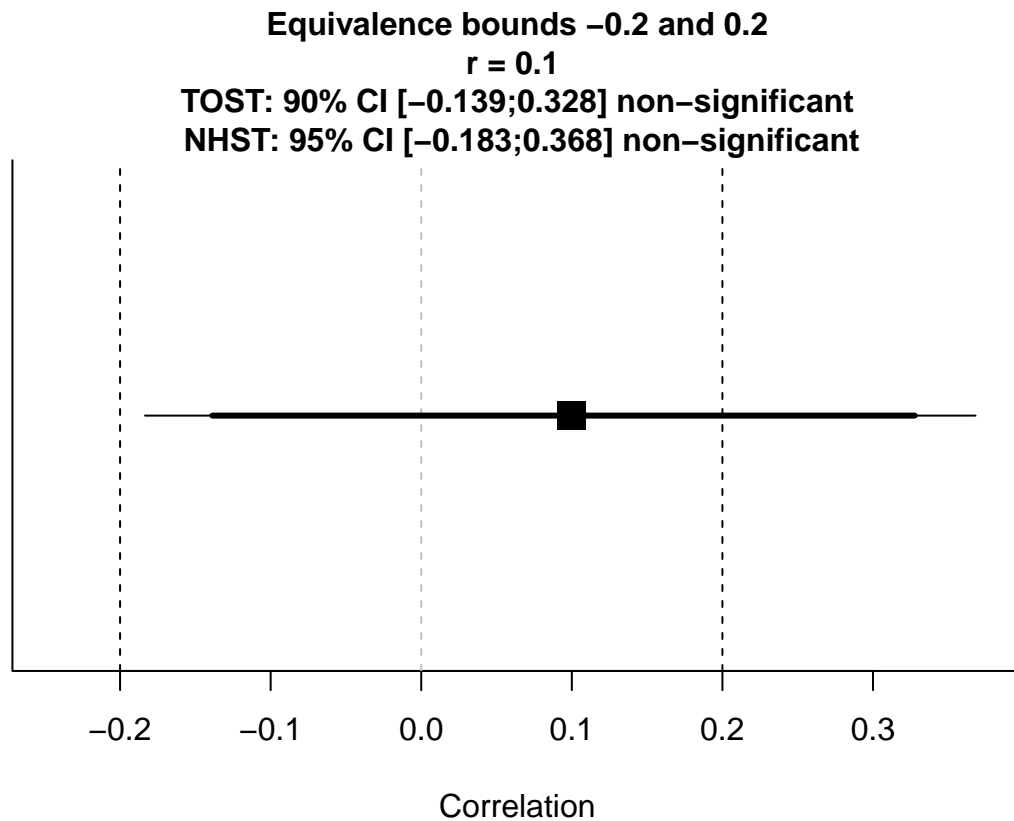
Quertemont (2011) provides an example in the appendix (page 126): Suppose a sample with $N = 50$, $r = .1$, and $d = .2$, and calculated that $p(Z < -0.70) = .24$. The confidence interval for r is -0.18 to 0.37

We can reproduce this result using the TOSTr function:

```
TOSTr(n = 50, r = 0.1, low_eqbound_r = -0.2, high_eqbound_r = 0.2,
      alpha = 0.05)
```

```
## Using alpha = 0.05 the NHST t-test was non-significant, p = 0.4895926
```

```
##
## Using alpha = 0.05 the equivalence test was non-significant, p = 0.2413396
##
```



```
## TOST results:
##   p-value 1   p-value 2
## 1 0.2413396 0.01886717
##
## Equivalence bounds (r):
##   low bound r   high bound r
## 1      –0.2      0.2
##
## TOST confidence interval:
##   Lower Limit 90% CI raw Upper Limit 90% CI raw
## 1      –0.1386915      0.3277112
```

The resulting p-value and 95% CI match the calculations by Quertemont (2011).

Example 4.2

The R code from Goertzen & Cribbie, page 536, was used:

```
# Running an original two t test procedure for equivalence
equivint <- 0.3
corxy <- 0.02
n <- 100
```

```

alpha <- 0.05

zei <- log((1 + equivint)/(1 - equivint))/2
zcorxy <- log((1 + corxy)/(1 - corxy))/2
equivt1_fz <- (zcorxy - zei)/(1/sqrt(n - 3))
pvalue1_fz <- pnorm(equivt1_fz)
equivt2_fz <- (zcorxy + zei)/(1/sqrt(n - 3))
pvalue2_fz <- 1 - pnorm(equivt2_fz)
ifelse(pvalue1_fz <= alpha & pvalue2_fz <= alpha, decis_fz <- "The null hypothesis that the correlation
  decis_fz <- "The null hypothesis that the correlation between var1 and var2 falls outside of the eq

## [1] "The null hypothesis that the correlation between var1 and var2 falls outside of the equivalence
pvalue1_fz

## [1] 0.002176282
pvalue2_fz

## [1] 0.0005863917

```

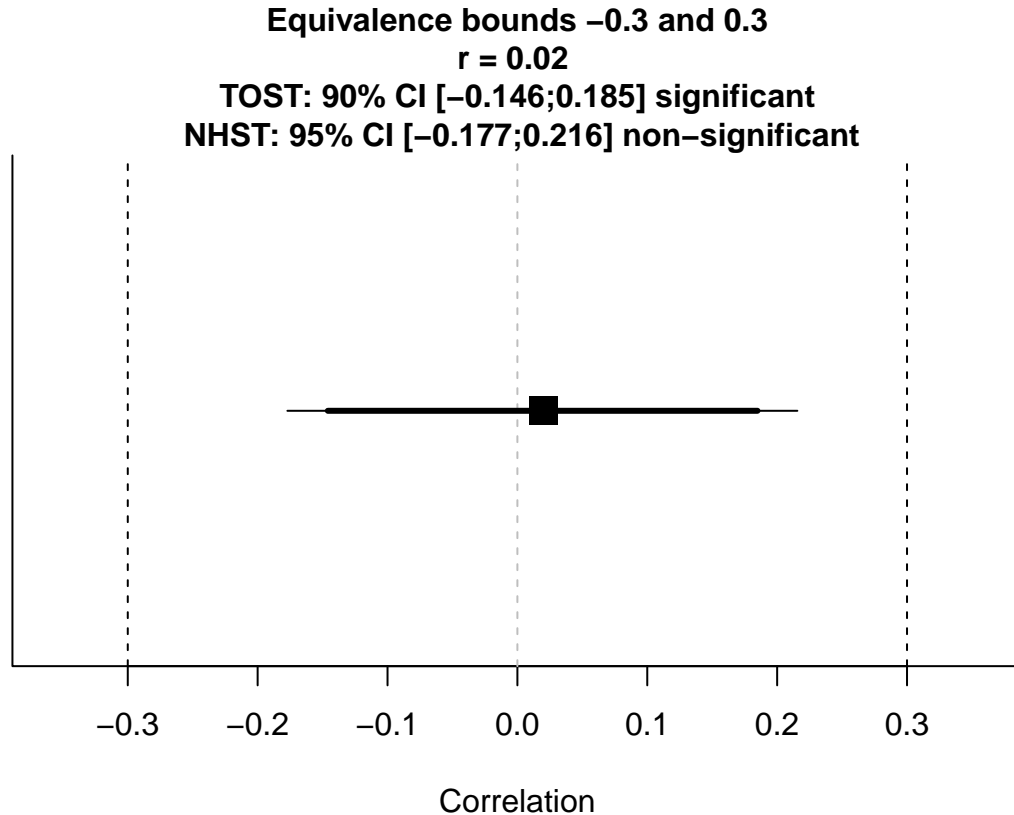
Which is identical to the p-values observed in the spreadsheet, and returned by the the TOSTr function:

```

TOSTr(n = 100, r = 0.02, low_eqbound_r = -0.3, high_eqbound_r = 0.3,
  alpha = 0.05)

## Using alpha = 0.05 the NHST t-test was non-significant, p = 0.8434322
##
## Using alpha = 0.05 the equivalence test was significant, p = 0.002176282
##

```



```
## TOST results:
##      p-value 1      p-value 2
## 1 0.002176282 0.0005863917
##
## Equivalence bounds (r):
##   low bound r high bound r
## 1      -0.3      0.3
##
## TOST confidence interval:
##   Lower Limit 90% CI raw Upper Limit 90% CI raw
## 1          -0.145957      0.1848622
```

Meta-analysis based on Cohen's d

Example 5.1

We can reproduce an example for a meta-analysis from From Rogers, J. L., Howard, K. I., & Vessey, J. T. (1993). Using significance tests to evaluate equivalence between two experimental groups. *Psychological Bulletin*, 113(3), page 557.

For example, we obtain the following calculations for the cognitive versus behavioral contrast. Traditional z : $z = ES/SE = 0.12/0.09 = 1.333$, $p = .091$. Traditional confidence interval: $ES \pm (z_{\alpha/2})(SE) = 0.12 \pm (1.96)(0.09)$, or -0.056 to 0.296. Equivalence z : $z_1 = (ES + 0.20)/SE = (0.12 + 0.20)/0.09 = 3.556$, $p = .000$. $z_2 = (ES - 0.20)/SE = (0.12 - 0.20)/0.09 = -0.889$, $p = .187$. So we table the larger p value of 0.187. Equivalence confidence interval: $ES \pm (z_{\alpha})(SE) = 0.12 \pm (1.645)(0.09)$, or -0.028 to 0.268.

We can recreate this with the TOSTmeta function:

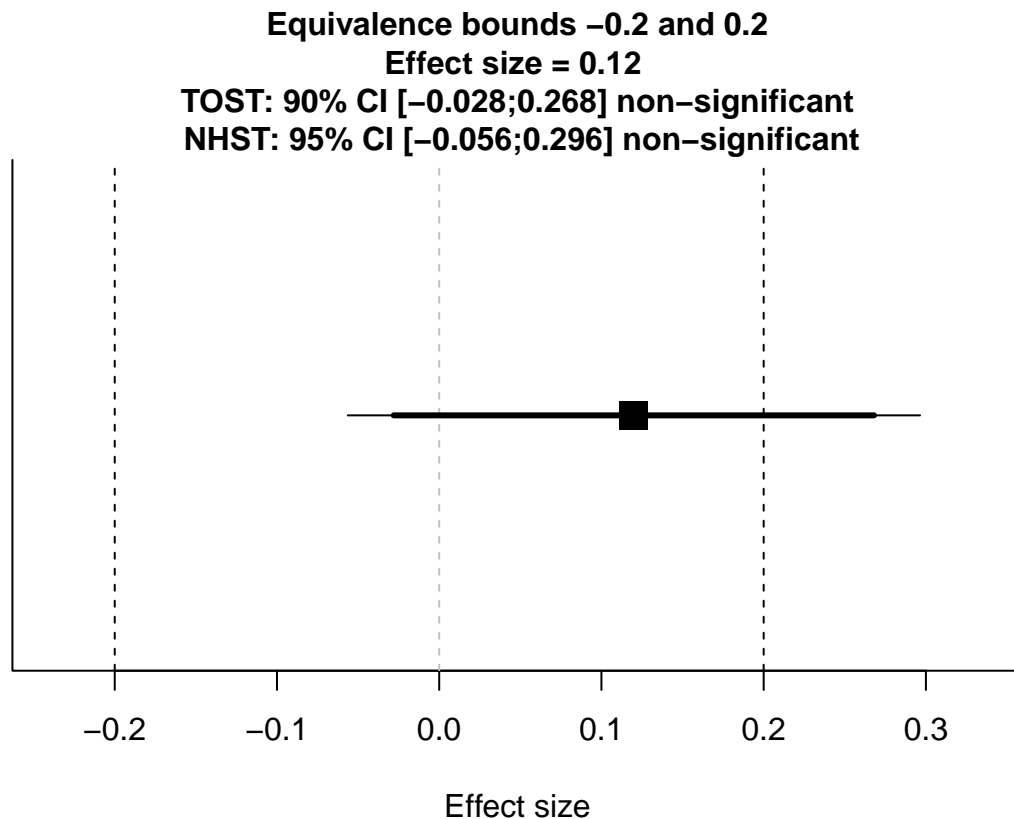
```
TOSTmeta(ES = 0.12, var = 0.0081, se = 0.09, low_eqbound_d = -0.2,  
         high_eqbound_d = 0.2, alpha = 0.05)
```

```
## Using alpha = 0.05 the meta-analysis was non-significant, Z = 1.333333, p = 0.09121122
```

```
##
```

```
## Using alpha = 0.05 the equivalence test was non-significant, Z = -0.8888889, p = 0.1870314
```

```
##
```



```
## TOST results:
```

```
##   Z-value 1   p-value 1   Z-value 2   p-value 2
```

```
## 1   3.555556 0.0001885906 -0.8888889 0.1870314
```

```
##
```

```
## Equivalence bounds (Cohen's d):
```

```
##   low bound d   high bound d
```

```
## 1       -0.2         0.2
```

```
##
```

```
## TOST confidence interval:
```

```
##   Lower Limit 90% CI   Upper Limit 90% CI
```

```
## 1       -0.02803683       0.2680368
```