Cluster Wild Bootstrapping to Handle Dependent Effect Sizes in Meta-Analysis with a Small Number of Studies

Megha Joshi

2022-04-12

Meta Analysis

- Set of statistical techniques to synthesize results from multiple studies on the same topic
- Goals of meta-analysis
    - Summarize effect size estimates across studies
    - Characterize variability in effect sizes
    - Explain the variability in effect sizes

Dependence

- Typical meta-analytic techniques (like meta-regression) involves the assumption that effect sizes are independent
- However, common for each primary study to yield more than one effect size or studies to be nested in some way creating dependence

Dependence Structure

- Correlated effects
    - Multiple outcomes measured on the same sample
    - Multiple treatment conditions compared to the same control
    - Multiple follow-up measures
    - Multiple correlations from the same sample
- Hierarchical effects
    - Studies conducted by the same lab

Example

- Tanner-Smith and Lipsey (2015) meta-analysis of the effects of brief alcohol interventions on alcohol consumption and behavior
    - 185 studies, 1446 effect sizes
    - Studies had multiple effect sizes
    - Multiple correlated outcome measures: e.g., alcohol consumption measured by frequency of consumption, quantity consumed, blood alcohol concentration
    - Repeated measures
    - Multiple comparison groups

Handling Dependence

- Ignore dependence - just run meta-regression
  - Incorrect standard errors, incorrect inference from hypothesis tests
- Ad-hoc methods
  - Selecting one effect per study
  - Analyzing subsets of data separately
  - Loss of information

Standard Multivariate Methods

- Ideal
- Requires info on covariance between effect sizes
- Primary studies often don't report
- So although this is ideal, we cannot actually use it in practice

Robust Variance Estimation

- Robust variance estimation (CR0-type CRVE) (Hedges, Tipton, and Johnson, 2010)
    - Rough approximate assumptions about dependence structure - sandwich estimators
    - Only works well when number of studies is large (> 40, Hedges Tipton, Johnson, 2010)
    - Meta-analysis in social science research typically have smaller number of studies
    - Small number of studies - CRVE - Type 1 error inflation - meta-analysts can conclude some effect is present when it is actually not

Tanner-Smith and Lipsey (2015)

Show 10 entries Search:

	study	es_num	delta	v	dv	g2age
1	2269	9587	-0.207	0.002	Frequency of use	19.7
2	2269	9591	-0.467	0.002	Quantity of use	19.7
3	2269	9586	-0.207	0.002	Frequency of use	19.7
4	2269	9590	-0.479	0.002	Quantity of use	19.7
5	2269	9589	0	0.002	Quantity of use	19.7
6	2269	9588	0	0.002	Quantity of use	19.7
7	2343	5420	-0.336	0.02	Blood alcohol concentration	20.6
8	2343	5444	-0.125	0.017	Quantity of use	20.6

Single Coefficient Tests

- Test of single coefficients
  - For example: Does the effect of brief alcohol interventions differ according to average age of the sample?

```
library(robumeta)

robu_tsl <- robu(delta ~ g2age + dv,
                 studynum = study,
                 var.eff.size = v,
                 data = tsl_dat)
```

Single Coefficient Test of Age

```
robu_tsl
```

```
## RVE: Correlated Effects Model with Small-Sample Corrections
##
## Model: delta ~ g2age + dv
##
## Number of studies = 20
## Number of outcomes = 207 (min = 1 , mean = 10.3 , median = 8 , max = 28 )
## Rho = 0.8
## I.sq = 73.89131
## Tau.sq = 0.04060939
##
##                                    Estimate StdErr t-value  dfs P(|t|>)
## 1                      X.Intercept.  0.41959 0.5758  0.7287 9.82   0.483
## 2                             g2age -0.02803 0.0291 -0.9637 9.96   0.358
## 3 dvCombined.measures..e.g...AUDIT.  0.15580 0.1261  1.2351 6.45   0.260
## 4          dvFrequency.of.heavy.use -0.03907 0.0721 -0.5421 7.53   0.603
## 5                dvFrequency.of.use  0.08189 0.0851  0.9623 8.30   0.363
## 6                dvPeak.consumption -0.00722 0.1877 -0.0385 6.42   0.970
## 7                 dvQuantity.of.use  0.02488 0.0662  0.3755 5.62   0.721
##   95% CI.L 95% CI.U Sig
## 1  -0.8665   1.7057
## 2  -0.0929   0.0368
```

Multiple-Contrast Hypothesis Test

- Multiple-contrast hypothesis tests
  - Do effects differ across outcome measurements?

```
## RVE: Correlated Effects Model with Small-Sample Corrections
##
## Model: delta ~ g2age + dv
##
## Number of studies = 20
## Number of outcomes = 207 (min = 1 , mean = 10.3 , median = 8 , max = 28 )
## Rho = 0.8
## I.sq = 73.89131
## Tau.sq = 0.04060939
##
##                                    Estimate StdErr t-value  dfs P(|t|>)
## 1                      X.Intercept.  0.41959 0.5758  0.7287 9.82   0.483
## 2                             g2age -0.02803 0.0291 -0.9637 9.96   0.358
## 3 dvCombined.measures..e.g...AUDIT.  0.15580 0.1261  1.2351 6.45   0.260
## 4          dvFrequency.of.heavy.use -0.03907 0.0721 -0.5421 7.53   0.603
## 5                dvFrequency.of.use  0.08189 0.0851  0.9623 8.30   0.363
## 6                dvPeak.consumption -0.00722 0.1877 -0.0385 6.42   0.970
## 7                 dvQuantity.of.use  0.02488 0.0662  0.3755 5.62   0.721
```

Multiple Contrast Hypothesis: Dependent Variable

To learn more here is a vignette from `clubSandwich`.

```r
library(clubSandwich)
```

```
## Registered S3 method overwritten by 'clubSandwich':
##   method    from
##   bread.mlm sandwich
```

```r
Wald_test(robu_tsl,
          constraints = constrain_zero(3:7),
          vcov = "CR2")
```

```
##  test Fstat df_num df_denom p_val sig
##   HTZ 0.767      5     7.15 0.601
```

Implications of Multiple-Contrast Hypothesis Tests

- If effects differ for different outcome measures
    - For example, if brief alcohol interventions reduce quantity consumed but increase frequency consumed; or have no effects on quantity consumed but reduce frequency consumed
    - Useful to see if program is having intended effect for outcomes of interest

Small Sample Corrections

- Tipton (2015) for tests of single coefficients
- Tipton and Pustejovsky (2015) for multiple-contrast hypothesis tests
- Both recommended a method - HTZ test
    - CR2 correction method and using the Satterthwaite degrees of freedom for single coefficient tests
    - Extension of CR2 + Satterthwaite for multiple-contrast hypothesis tests
- HTZ controls Type 1 error rates adequately
- But, really low Type 1 error rates especially for **multiple-contrast hypothesis tests** (Tipton and Pustejovsky, 2015)
    - Indicating that the test may have low power
    - e.g., might miss the difference in the treatment effect across outcome measures when the difference actually exists

Cluster Wild Bootstrapping (CWB)

- Alternative method - examined in the econometrics literature - not in meta-analytic framework
- Bootstrapping - estimate unknown quantities by re-sampling from original data many times (Boos et al., 2013)
- CWB - re-sampling residuals by multiplying them by cluster-level random weights (Cameron, Gelbach, and Miller 2008)

CWB Algorithm

1. Fit a null model and a full model on the original data

2. Obtain residuals from the null model

3. Generate an auxiliary random variable that has mean of 0 and variance of 1 and multiply the residuals by the random variable (e.g., Rademacher weights) set to be constant within clusters (CWB)

    - Can also multiply the residuals by CR2 matrices before multiplying by weights (CWB Adjusted)

4. Obtain new outcome scores by adding the transformed residuals to the predicted values from the null model fit on the original data

5. Re-estimate the full model with the new calculated outcome scores and obtain the test statistic

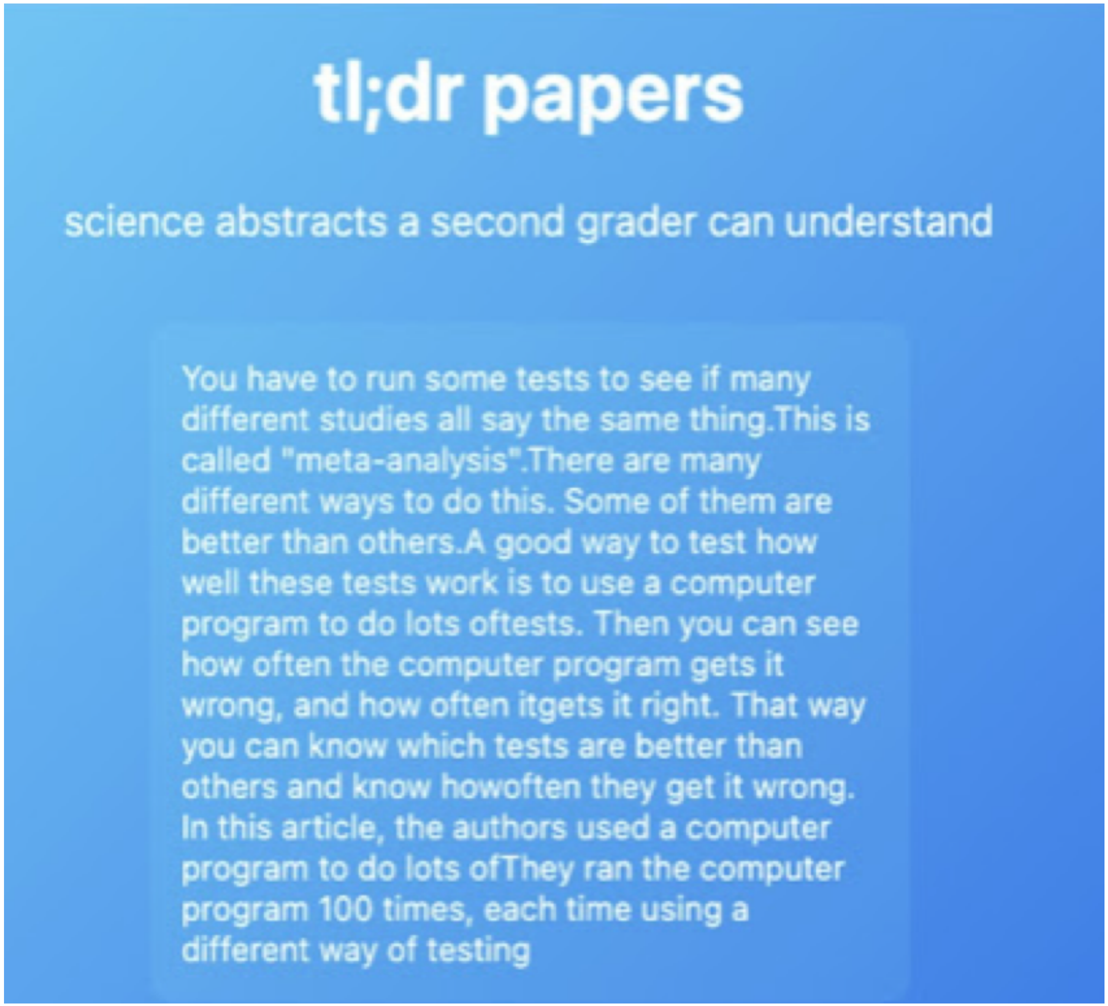
6. Repeat steps 3-5 R times. Calculate p-value:

$$p = \frac{1}{R} \sum_{r=1}^{R} I\left(F^{(r)} > F\right)$$

Research Question

To what extent does CWB improve upon the current standard test, the HTZ test, in terms of Type I error rates and power?

Simulation Studies



Simulation Study Methods

Data Generation

- Standardized mean differences
- Correlated effects meta-analytic data
- Study 1 covariates
  - Design matrix from Tipton and Pustejovsky (2015)
  - Replicate Tipton and Pustejovsky (2015) and study methods with design matrix that contains imbalanced, non-normal covariates
- Study 2 covariates
  - One covariate (between, within) with 3 to 5 balanced categories
  - Examine if the treatment effects are similar across multiple categories of a categorical variable (e.g., different age groups)

Estimation Methods

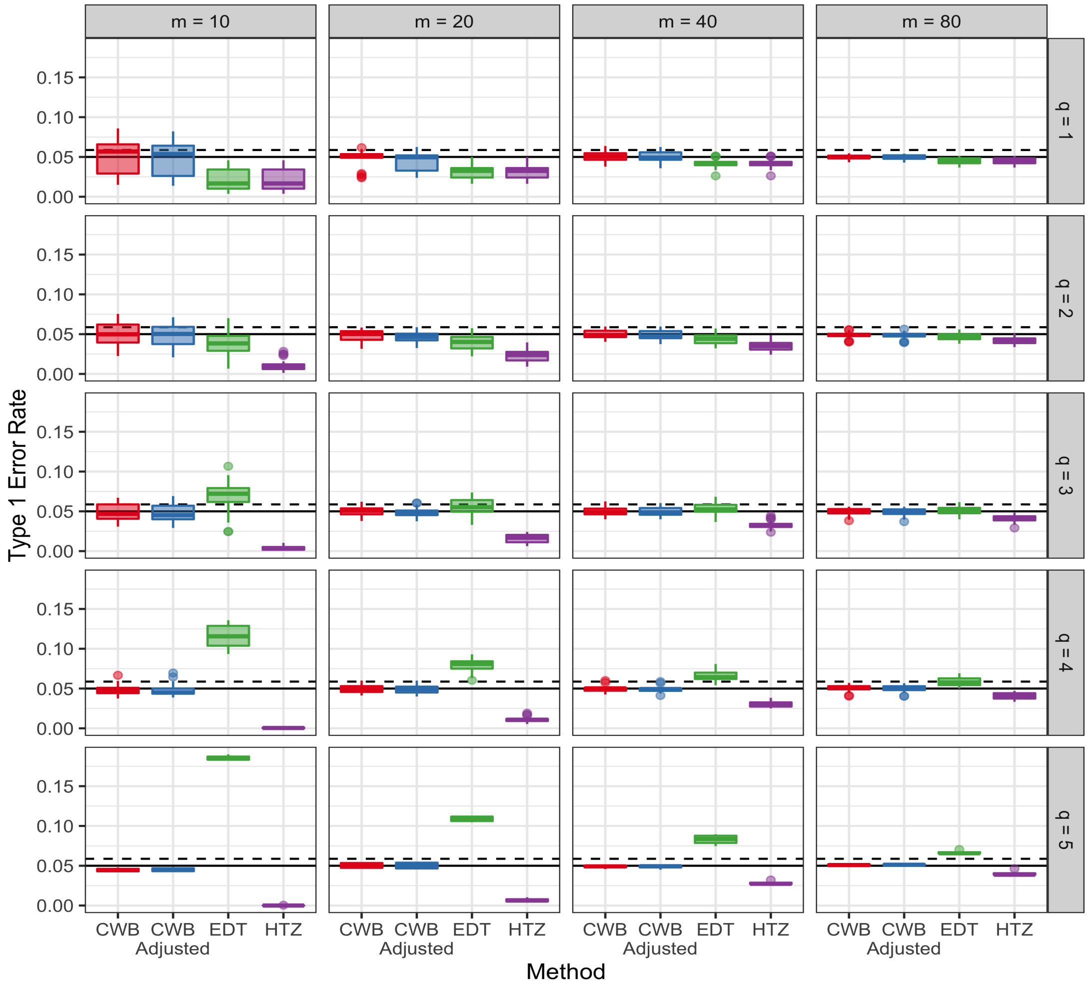
- Meta-regression - correlated effects working model
- Tests:
  - Single coefficient tests
  - Multiple-contrast hypothesis tests
- Estimation methods:
  - HTZ test
  - CWB test
  - CWB Adjusted test
- Number of bootstraps set to 399 (to save time and based on prior simulation studies)

Performance Criteria

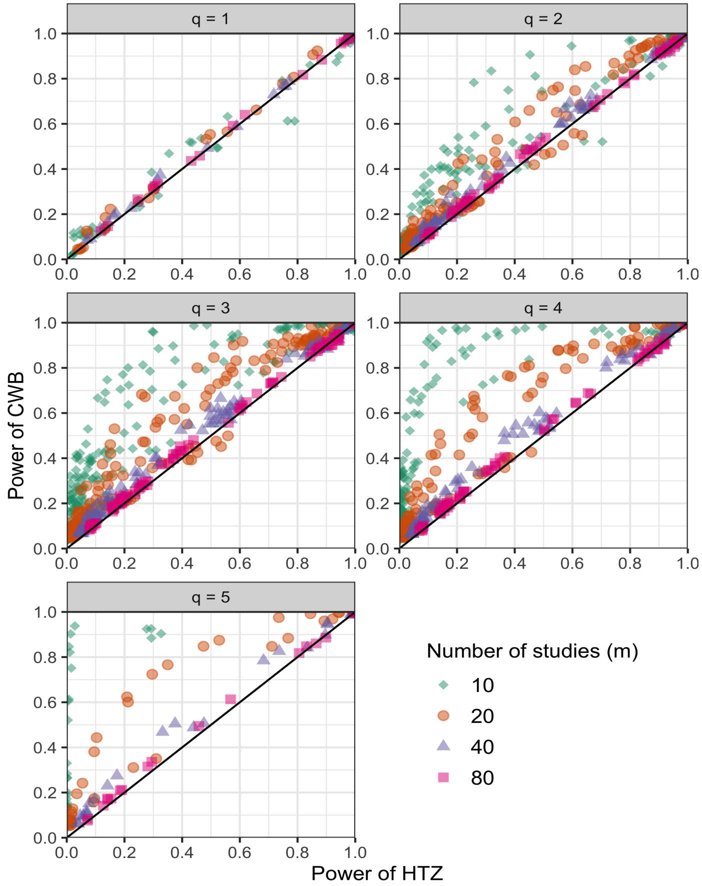
- Focus on hypothesis testing
  - Type I error rate - probability of rejecting a null hypothesis when false true
  - Power - probability of rejecting null hypothesis when null is false

Study 1 Results

Type I Error

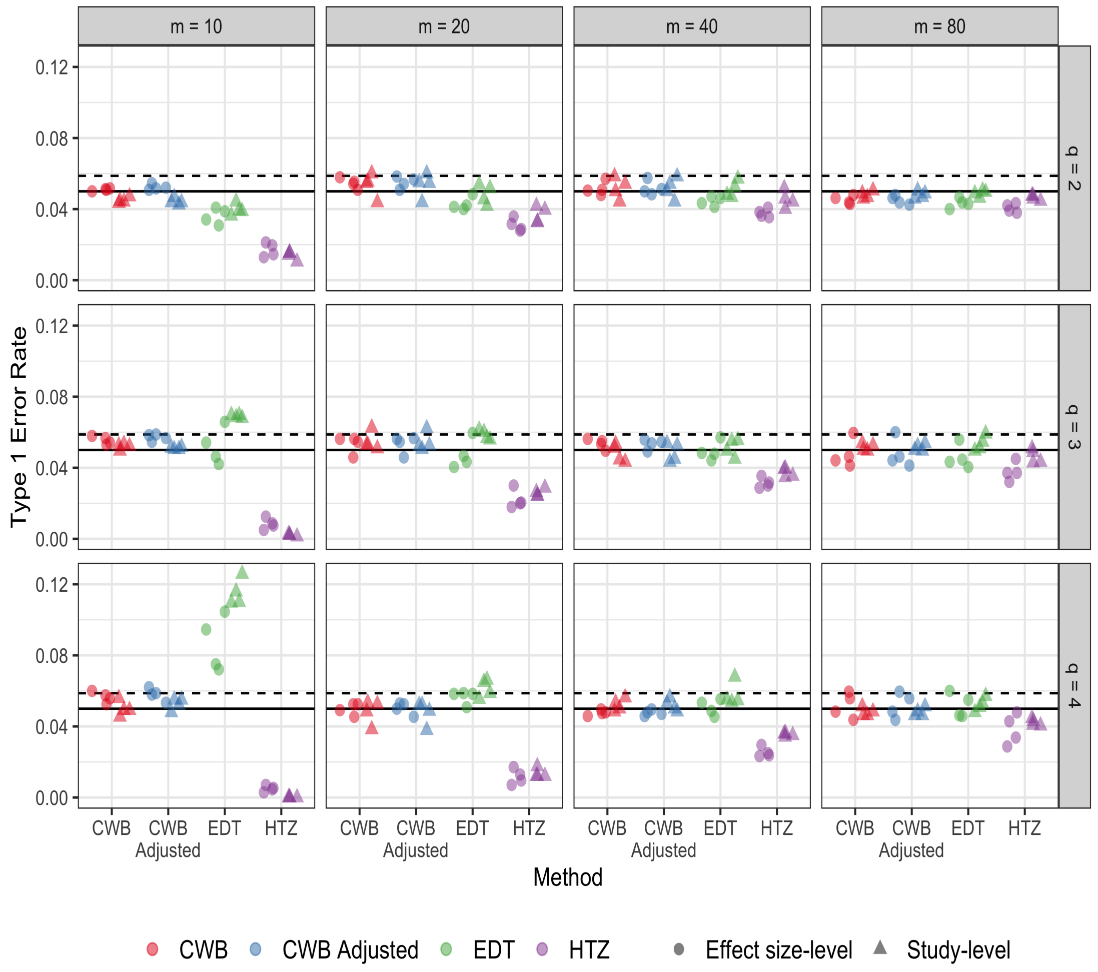

m = 10
m = 20
m = 40
m = 80
q = 1
q = 2
q = 3
q = 4
q = 5
Type 1 Error Rate
0.00
0.05
0.10
0.15
CWB
CWB Adjusted
EDT
HTZ
Method

Relative Power



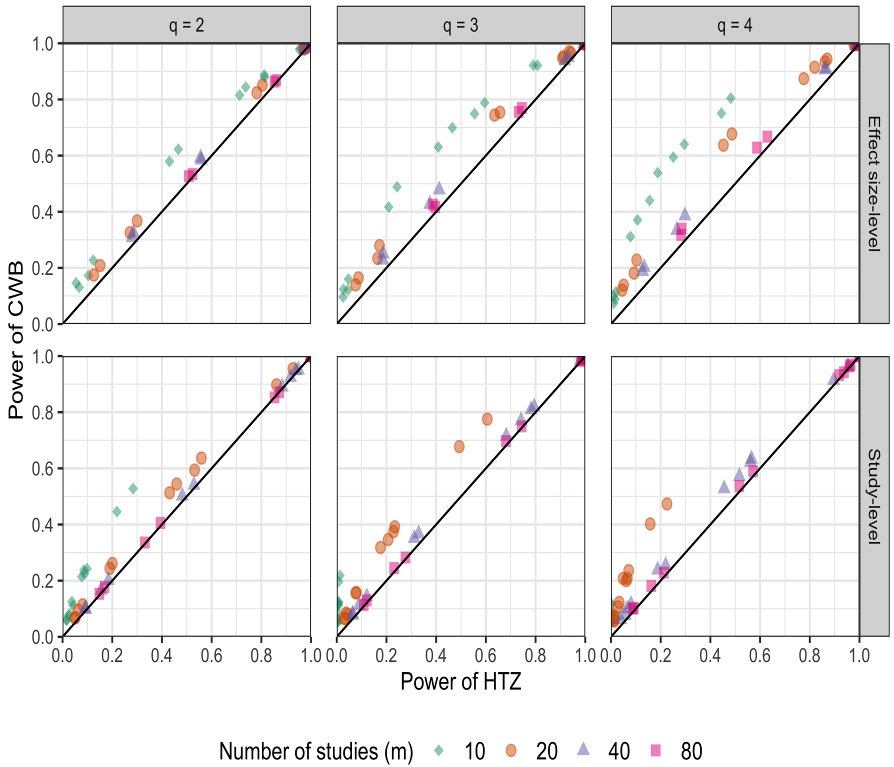
Study 2 Results

Type I Error



m = 10
m = 20
m = 40
m = 80
q = 2
q = 3
q = 4
Type 1 Error Rate
0.00
0.04
0.08
0.12
CWB
CWB Adjusted
EDT
HTZ
Method
CWB
CWB Adjusted
EDT
HTZ
Effect size-level
Study-level

Relative Power



Conclusion

- Dependent effect sizes - common
- Ignore them - incorrect standard errors and inferences
- Use RVE - Type 1 error inflation - false discovery rate high
- Use small sample correction HTZ test - may miss effects that are present - particularly for multiple-contrast hypothesis tests
- Use CWB - balances Type 1 error rates and also provides more power than existing corrections

wildmeta

wildmeta

- The main function in the package is `Wald_test_cwb()`
- Works with meta-regressions models fit using `robumeta::robu()` and `metafor::rma.mv()`

```
Wald_test_cwb(
  full_model,
  constraints,
  R,
  cluster = NULL,
  auxiliary_dist = "Rademacher",
  adjust = "CR0",
  type = "CR0",
  test = "Naive-F",
  seed = NULL
)
```

TSL Data

Show 10 entries Search:

	study	es_num	delta	v	dv	g2age
1	2269	9587	-0.207	0.002	Frequency of use	19.7
2	2269	9591	-0.467	0.002	Quantity of use	19.7
3	2269	9586	-0.207	0.002	Frequency of use	19.7
4	2269	9590	-0.479	0.002	Quantity of use	19.7
5	2269	9589	0	0.002	Quantity of use	19.7
6	2269	9588	0	0.002	Quantity of use	19.7
7	2343	5420	-0.336	0.02	Blood alcohol concentration	20.6
8	2343	5444	-0.125	0.017	Quantity of use	20.6
9	2343	5449	-0.155	0.019	Blood alcohol concentration	20.6
10	2343	5443	-0.053	0.017	Peak consumption	20.6

robumeta Model

```r
library(wildmeta)
library(clubSandwich)
library(robumeta)

robu_tsl <- robu(delta ~ g2age + dv,
                 studynum = study,
                 var.eff.size = v,
                 small = TRUE,
                 data = tsl_dat)
```

```r
robu_tsl
```

```
## RVE: Correlated Effects Model with Small-Sample Corrections
##
## Model: delta ~ g2age + dv
##
## Number of studies = 20
## Number of outcomes = 207 (min = 1 , mean = 10.3 , median = 8 , max = 28 )
## Rho = 0.8
## I.sq = 73.89131
## Tau.sq = 0.04060939
##
##                                   Estimate StdErr t-value  dfs P(|t|>)
## 1                      X.Intercept.  0.41959 0.5758  0.7287 9.82   0.483
## 2                             g2age -0.02803 0.0291 -0.9637 9.96   0.358
## 3 dvCombined.measures..e.g...AUDIT.  0.15580 0.1261  1.2351 6.45   0.260
## 4          dvFrequency.of.heavy.use -0.03907 0.0721 -0.5421 7.53   0.603
## 5                dvFrequency.of.use  0.08189 0.0851  0.9623 8.30   0.363
## 6                dvPeak.consumption -0.00722 0.1877 -0.0385 6.42   0.970
## 7                 dvQuantity.of.use  0.02488 0.0662  0.3755 5.62   0.721
##   95% CI.L 95% CI.U Sig
## 1  -0.8665   1.7057
## 2  -0.0929   0.0368
## 3  -0.1477   0.4593
## 4  -0.2071   0.1290
## 5  -0.1131   0.2769
## 6  -0.4593   0.4449
## 7  -0.1399   0.1897
```

Wald_test_cwb()

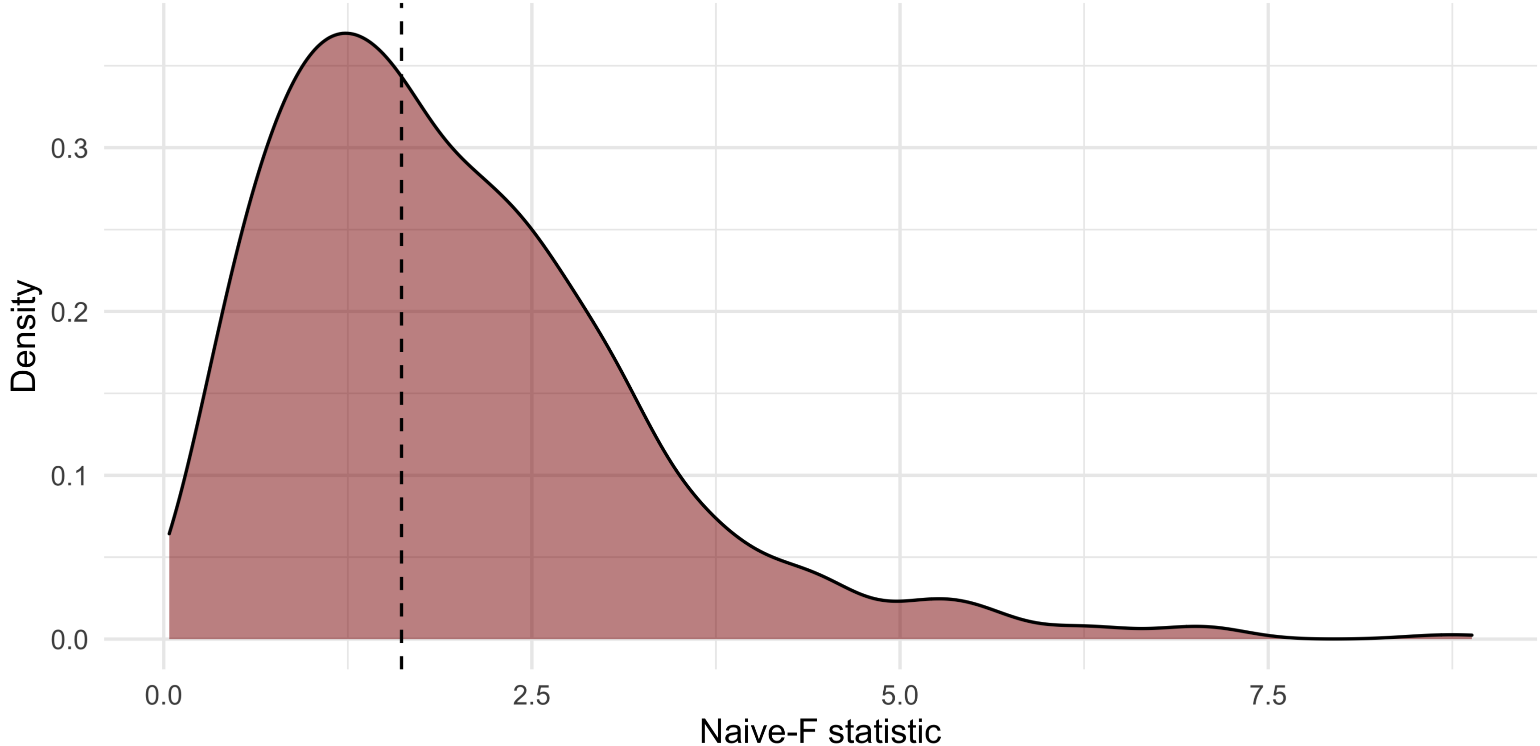
```
robu_res <- Wald_test_cwb(robu_comp_tsl,
                          constraints = constrain_zero(3:7),
                          R = 999,
                          seed = 20200218)

robu_res
```

```
##   Test Adjustment CR_type Statistic   R     p_val
## 1  CWB        CR0     CR0   Naive-F 999 0.5375375
```

Bootstrap Distribution Plot

```
plot(robu_res,
     fill = "darkred",
     alpha = 0.5)
```



Links

- `wildmeta`
- Pre-print
- Research Synthesis Methods

THANK YOU!

Questions?