Supplement

Contents

1. A simple, reproducible code example

A simulated dataset and simple code to estimate the proposed metrics is publicly available (https://osf.io/gs7fp/). The code also appears below.

```
############################## LOAD PACKAGES ##############################
library(robumeta)
library(metafor)
library(boot)
library(dplyr)
library(tidyr)

############################## HELPER FUNCTIONS ##############################
##### Get BCa-bootstrapped confidence intervals for a vector of parameters (not
    just one) #####
# boot.res: an object from boot()
# type: what kind of bootstrapping to use? (as in boot.ci())
# n.ests: how many parameters were estimated?
get_boot_CIs = function(boot.res,
                        type = "bca",
                        n.ests) {
  bootCIs = lapply( 1:n.ests, function(x) boot.ci(boot.res, type = type, index = x)
      )

  # list with one entry per estimate
  # the middle index "4" on the bootCIs accesses the stats vector
  # the final index chooses the CI lower (4) or upper (5) bound
  bootCIs = lapply( 1:n.ests, function(x) c( bootCIs[[x]][[4]][4],
                                             bootCIs[[x]][[4]][5] ) )
}

##### Return Phat(z), Phat(z0), and Phat(z) - Phat(z0) #####
# important: this function would need to be modified somewhat to use it on another
    dataset (see comments therein)
# dat: the dataset
# cluster.var: the name of the variable denoting clusters, if any
# if left to default NA, assumes studies are independent
# q: threshold for a meaningfully strong effect size (function assumes we are
    considering effects ABOVE the threshold)
# z: level of effect modifiers of interest (vector)
# z0: reference level of effect modifiers (vector)
# return.meta: should the meta-regression itself be returned, or just the
    percentages?
get_phat_reprex = function(dat,
                           cluster.var = NA,
                           q,
                           z,
```

```
                      z0,
                      return.meta = FALSE){

if ( is.na(cluster.var) ) dat$cluster = 1:nrow(dat) else dat$cluster = dat[[
    cluster.var]]

# fit the meta-regression
# linear predictor would need to be modified for other datasets
mod = robu( yi ~ Zc + Zb,
            studynum = cluster, # account for clustering
            data = dat,
            var.eff.size = vyi,
            small = TRUE )

bhat = mod$b.r # vector of coefficient estimates
t2 = mod$mod_info$tau.sq # heterogeneity estimate

# calculate the linear predictor (minus intercept) for each study
# would need to be modified for other datasets
dat$linpredZ = ( c(bhat[2]) * dat$Zc ) + ( c(bhat[3]) * dat$Zb )

##### Phat(z) #####
# point estimate, shifted to set effect modifiers to 0
dat$yi.shift = dat$yi - dat$linpredZ

# calibrated estimate, shifted to set effect modifiers to 0
calib.shift = c(bhat[1]) + sqrt( c(t2) / ( c(t2) + dat$vyi) ) * ( dat$yi.shift -
    c(bhat[1]) )

# threshold, shifted to set effect modifiers to 0
# would need to be modified for other datasets
q.shift = q - ( bhat[2]*z[1] + bhat[3]*z[2] )
# note: below assumes we are considering effects ABOVE the threshold
Phat = mean( calib.shift > c(q.shift) )

##### Phat(z0) #####
# reference level
# would need to be modified for other datasets
q.shift.ref = q - ( bhat[2]*z0[1] + bhat[3]*z0[2] )
Phat.ref = mean( calib.shift > c(q.shift.ref) )

##### Return Results #####
if ( return.meta == FALSE){
  return( c(Phat, Phat.ref, Phat - Phat.ref) )
} else {
  return( list(mod,
               c(Phat, Phat.ref, Phat - Phat.ref)) )
```

```
  }
}


############################## ANALYZE FAKE DATA ##############################
# read in the fake data
# these were generated similarly to data from the simulation study
# with tau^2 = 0.04, E[N] = 100, normal true effects, and clustering
# variable names:
# yi: point estimate
# vyi: variance
# Zc: continuous effect modifier
# Zb: binary effect modifier

# set your working directory to contain the fake data:
d = read.csv("reprex_fake_data.csv")
head(d)

# define comparisons of interest
q = .2 # threshold of interest
z = c(1, 1) # effect modifier level of interest
z0 = c(2.5, 0) # effect modifier reference level to be compared to level of
    interest

##### Point Estimates #####
# fit meta-regression and get point estimates for Phat(z), Phat(z0), and Phat(z) -
    Phat(z0)
stats = get_phat_reprex(dat = d,
                        cluster.var = "cluster",
                        q = q,
                        z = z,
                        z0 = z0,
                        return.meta = TRUE)

# look at meta-regression coefficients
stats[[1]]

# Phat(z), Phat(z0), and Phat(z) - Phat(z0) respectively
stats[[2]]

##### Bootstrapped Inference #####
# number of iterates: should be much higher in practice (e.g., 1,000),
# but kept small here for illustrative purposes
boot.reps = 300

# to easily allow for cluster-bootstrapping, nest the data
# now has one row per cluster
```

```
# works whether there is clustering or not
# because without clustering, the clusters are 1:nrow(data)
dNest = d %>% group_nest(cluster)
boot.res = boot( data = dNest,
                 parallel = "multicore",
                 R = boot.reps,
                 statistic = function(original, indices) {
                   # resample clusters with replacement
                   bNest = original[indices,]
                   b = bNest %>% unnest(data)

                   get_phat_reprex(dat = b,
                                   cluster.var = "cluster",
                                   q = q,
                                   z = z,
                                   z0 = z0,
                                   return.meta = FALSE) } )
# note: this simple bootstrapping script could, for some datasets, hit errors if
    the meta-regression
# model can't be fit for some resamples
# for a more sophisticated bootstrapping script that can proceed even if certain
    iterates fail,
# or if the BCa confidence interval construction works for some of the 3 estimators
     but not all,
# see the simulation study script "helper_MRM.R" and "doParallel.R"

# get the bootstrapped CIs for all 3 statistics respectively using BCa method
( bootCIs = get_boot_CIs(boot.res, n.ests = 3) )

##### All Results #####
# Phat(z) and its CI limits
stats[[2]][1]; bootCIs[[1]]

# Phat(z0) and its CI limits
stats[[2]][2]; bootCIs[[2]]

# Phat(z) - Phat(z0) and its CI limits
stats[[2]][3]; bootCIs[[3]]
```

2. Extended simulation results

2.1. Associations of meta-analysis characteristics with performance metrics

We regressed each of 3 performance metrics (absolute error, coverage, and confidence interval width) on main effects of 6 characteristics of the meta-analysis: k, $E[N]$, the presence of clustering, the population effect distribution, use of the BC-rare covariate contrast (used only in the regressions for $\widehat{P}_{>q}(z) - \widehat{P}_{>q}(z_0)$), and the true $P_{>q}$. Models were fit at the iterate level (2.3 million observations) with robust standard errors to account for clustering within scenarios. Because models were fit at the the iterate rather than scenario level, we did not include bias, which is only defined at the scenario level.

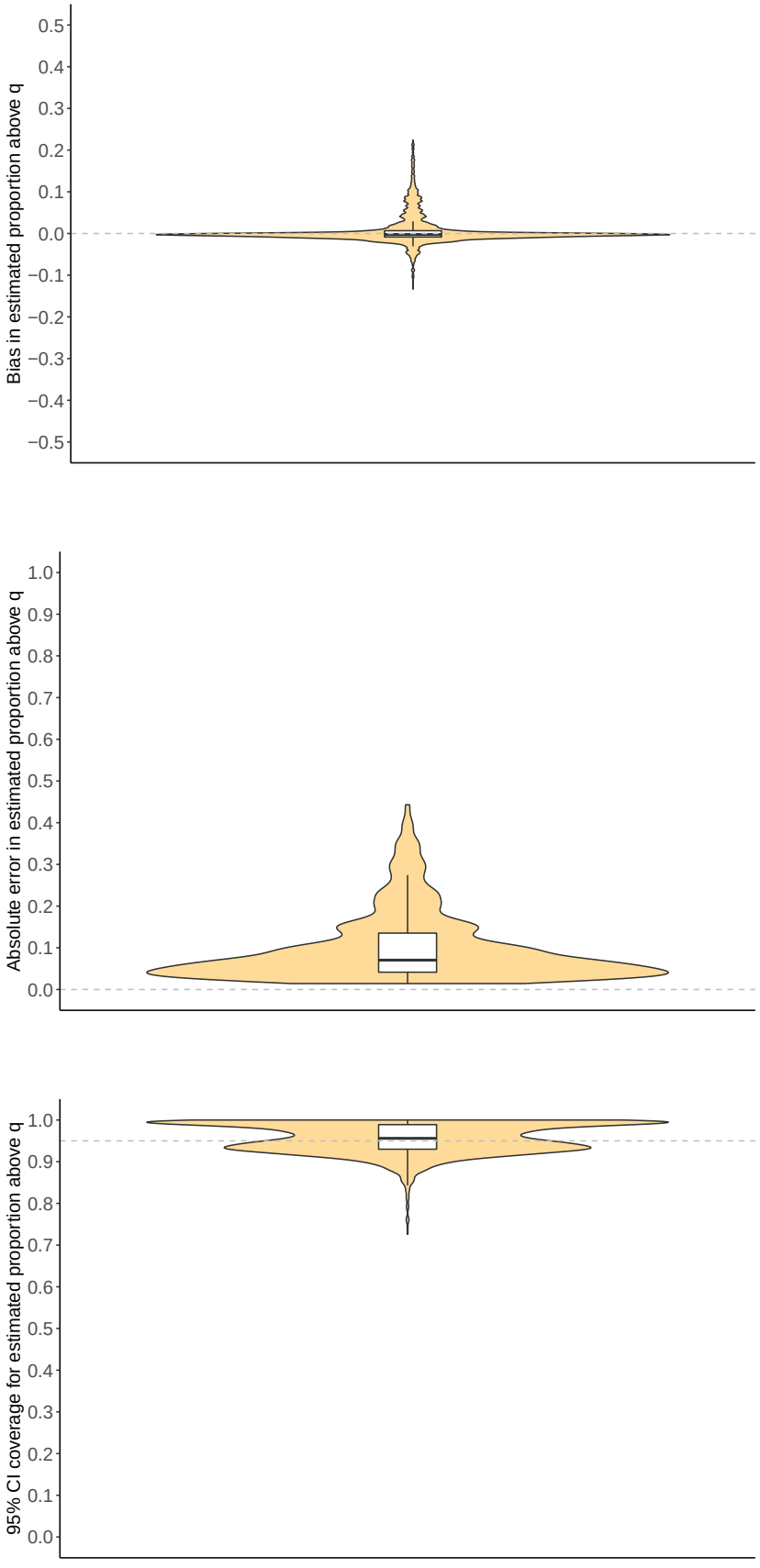
Performance metric	**Good**	**Bad**
$\widehat{P}_{>q}$ abs. error	k, E[N], normal effects	Clustered
$\widehat{P}_{>q}$ coverage	Normal effects	k, E[N], clustered
$\widehat{P}_{>q}$ CI width	k, E[N], normal effects	Clustered
$\widehat{P}_{>q}(z) - \widehat{P}_{>q}(z_0)$ abs. error	k, E[N], normal effects	Clustered, BC-rare
$\widehat{P}_{>q}(z) - \widehat{P}_{>q}(z_0)$ coverage	k, E[N], normal effects	Clustered, BC-rare
$\widehat{P}_{>q}(z) - \widehat{P}_{>q}(z_0)$ CI width	k, E[N], normal effects	Clustered, BC-rare

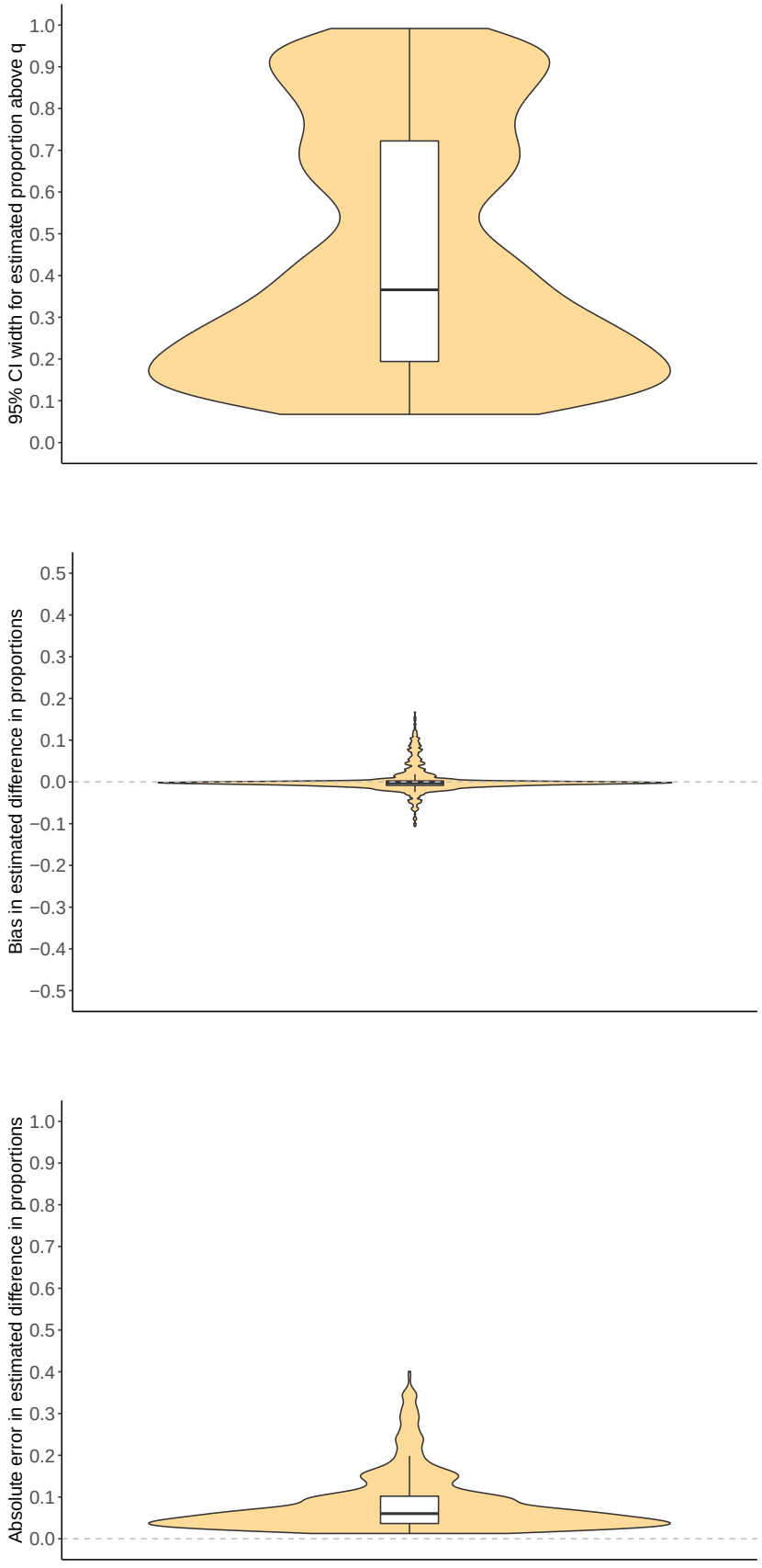
Supplementary Table 1: *"Good": characteristics associated with improved performance ($p < 0.001$). "Bad": characteristics associated with worse performance ($p < 0.001$). Abs. error: absolute error. Coverage: coverage of 95% confidence intervals. CI width: Width of 95% confidence intervals.*

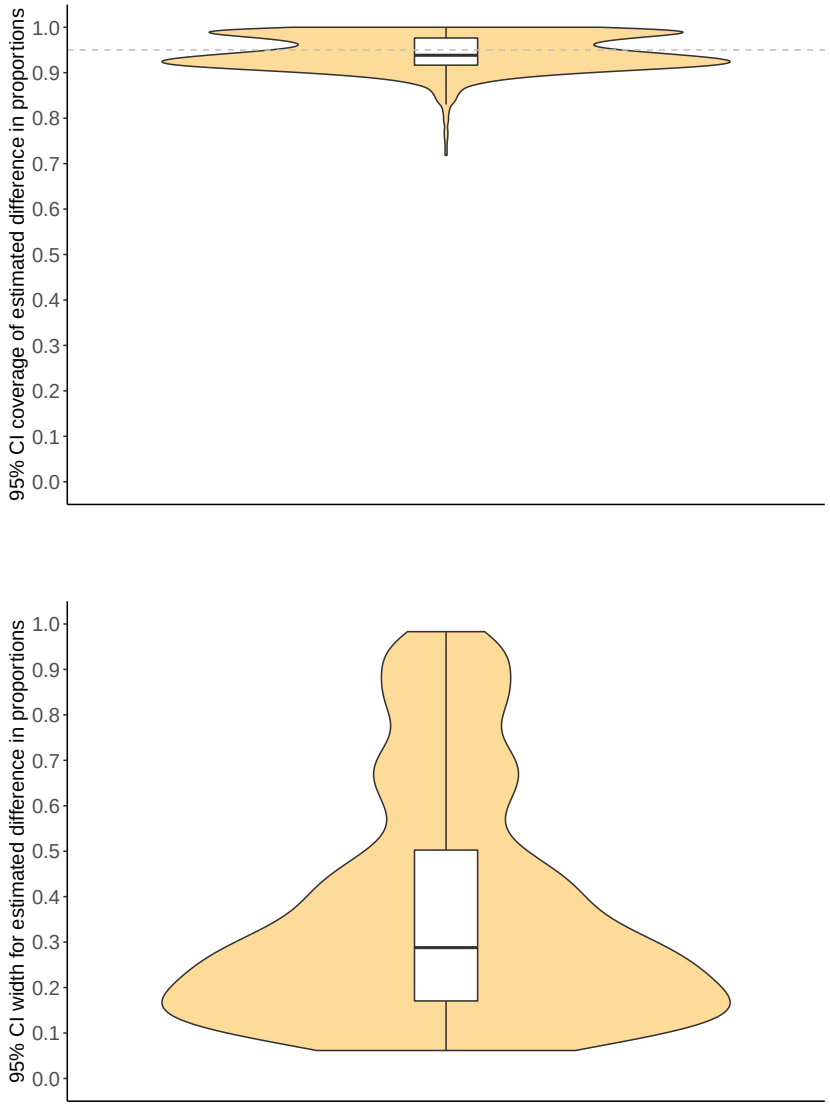
2.2. Violin plots of performance metrics (recommended scenarios)

These plots show only scenarios meeting the guidelines given in the Discussion (not clustered exponential effects and, for $\widehat{P}_{>q}(z) - \widehat{P}_{>q}(z_0)$, not the BC-rare contrast). Orange violins are mirrored density plots. White boxplots display the median, 25^{th} percentile, and 75^{th} percentile. Horizontal dashed reference lines represent perfect performance.

Plots are ordered first by the performance metric, shown on plots' Y-axes (bias in $\widehat{P}_{>q}(z)$, absolute error in $\widehat{P}_{>q}(z)$, coverage for $\widehat{P}_{>q}(z)$, 95% confidence interval width for $\widehat{P}_{>q}(z)$, bias in $\widehat{P}_{>q}(z) - \widehat{P}_{>q}(z_0)$, absolute error in $\widehat{P}_{>q}(z) - \widehat{P}_{>q}(z_0)$, coverage for $\widehat{P}_{>q}(z) - \widehat{P}_{>q}(z_0)$, 95% confidence interval width for $\widehat{P}_{>q}(z) - \widehat{P}_{>q}(z_0)$) and then by the characteristic of the meta-analysis, shown on plots' X-axes (k, $E[N]$, the presence of clustering, the population effect distribution, and the covariate contrast). The covariate contrast is omitted for $\widehat{P}_{>q}$ because it is not relevant.

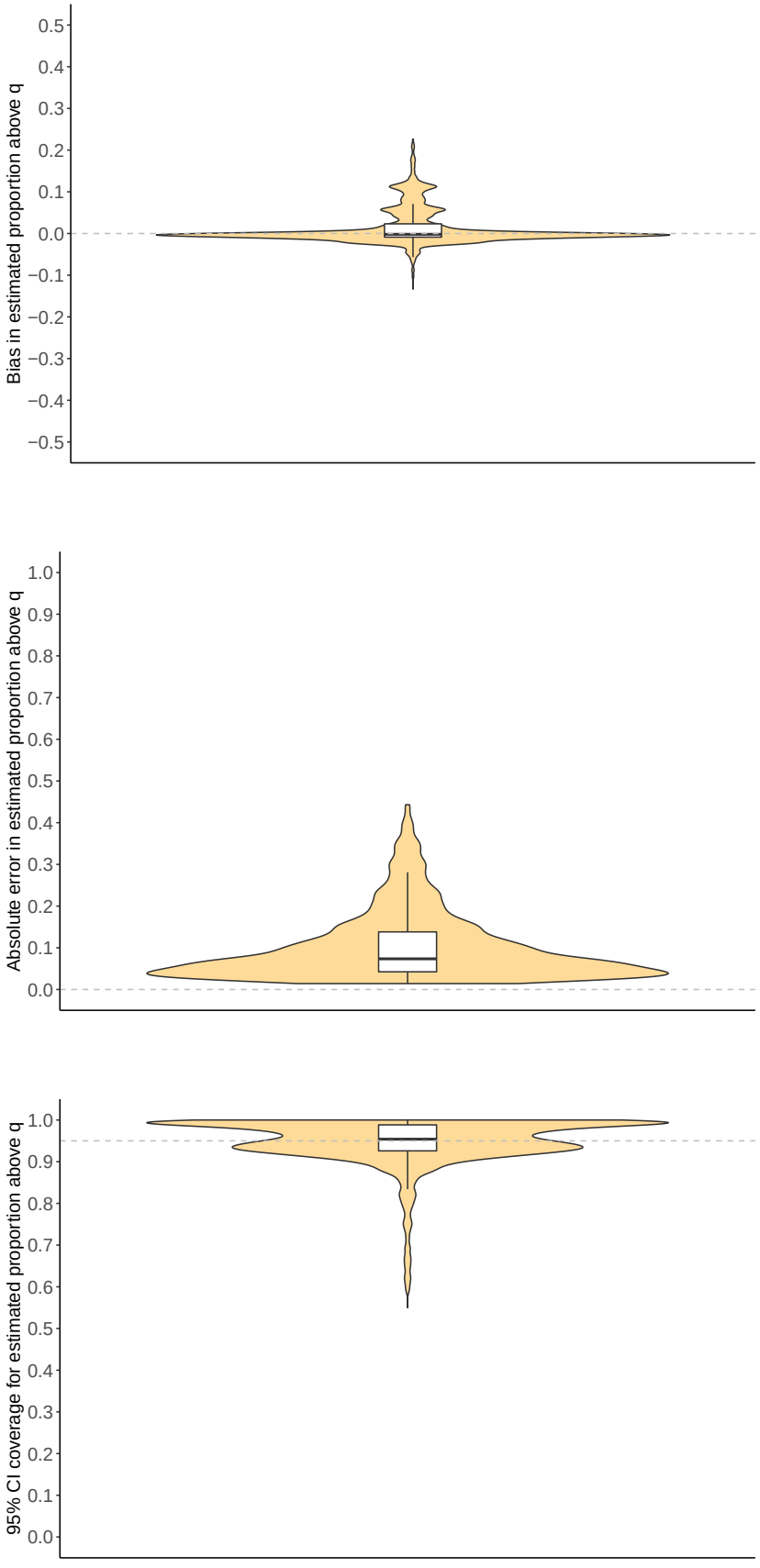

Bias in estimated proportion above q
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.5
Absolute error in estimated proportion above q
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
95% CI coverage for estimated proportion above q
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0

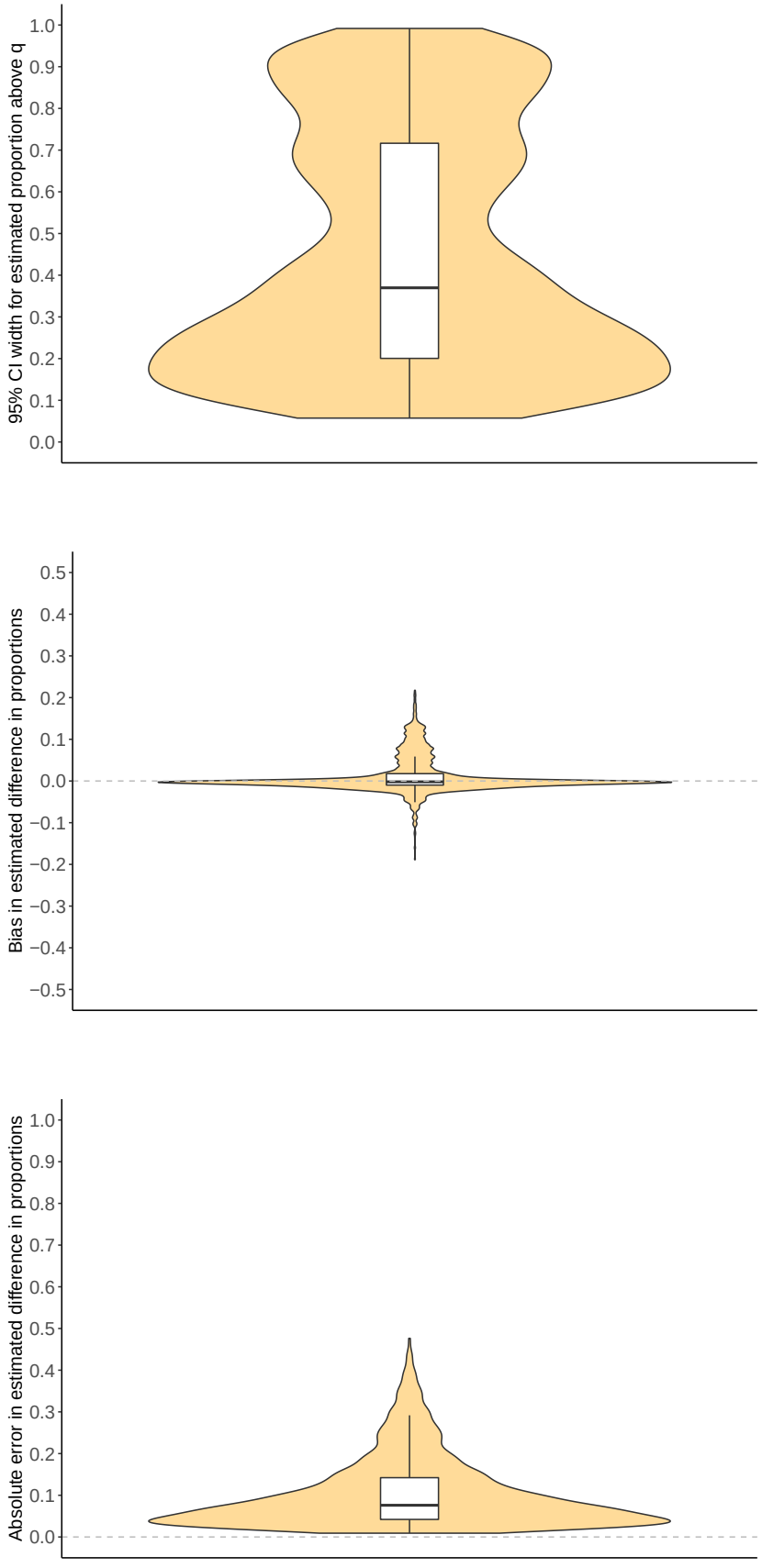

95% CI width for estimated proportion above q
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Bias in estimated difference in proportions
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.5
Absolute error in estimated difference in proportions
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0

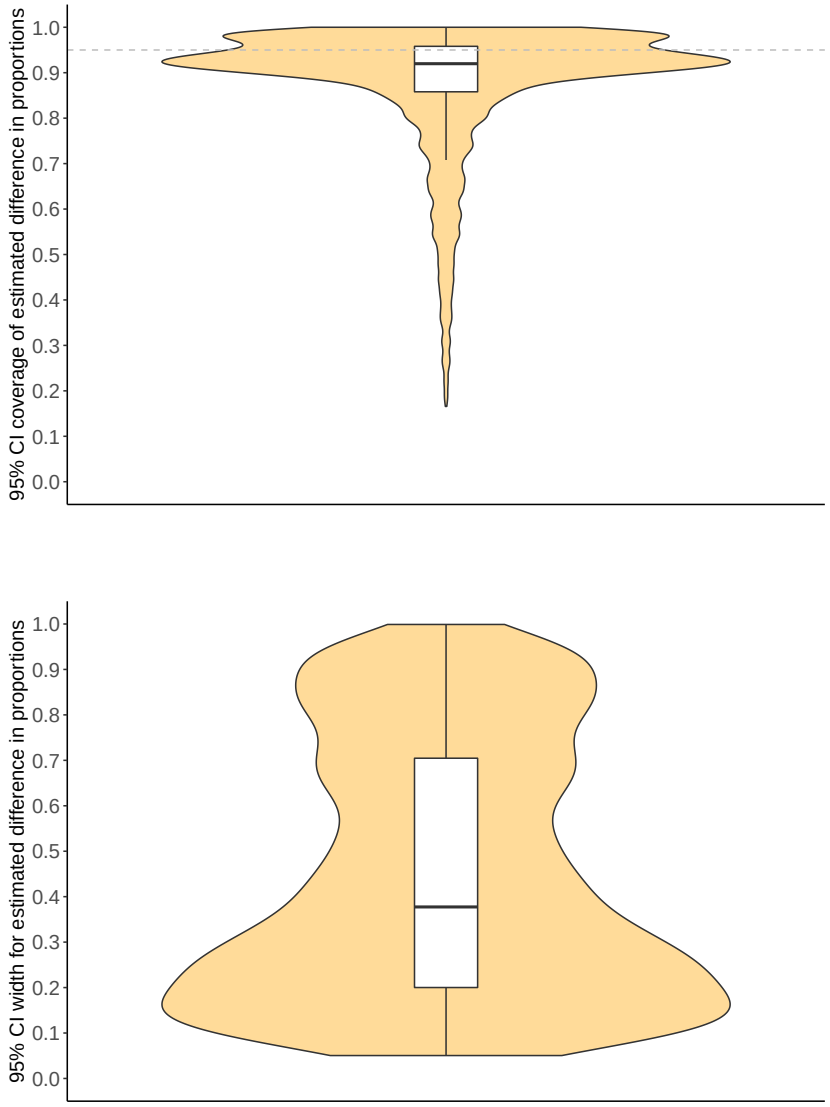


2.3. Violin plots of performance metrics (all scenarios)

These plots show all scenarios, including those that do not fulfill the guidelines given in the Discussion.

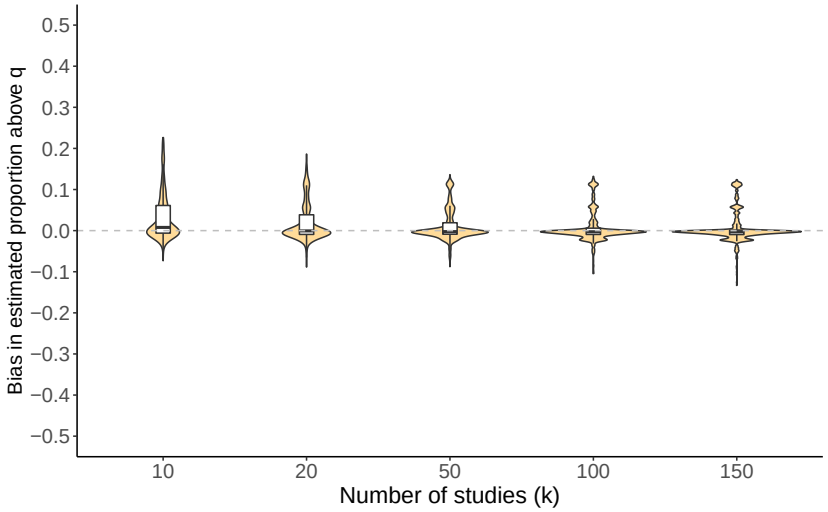

Bias in estimated proportion above q
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.5
Absolute error in estimated proportion above q
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
95% CI coverage for estimated proportion above q
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0

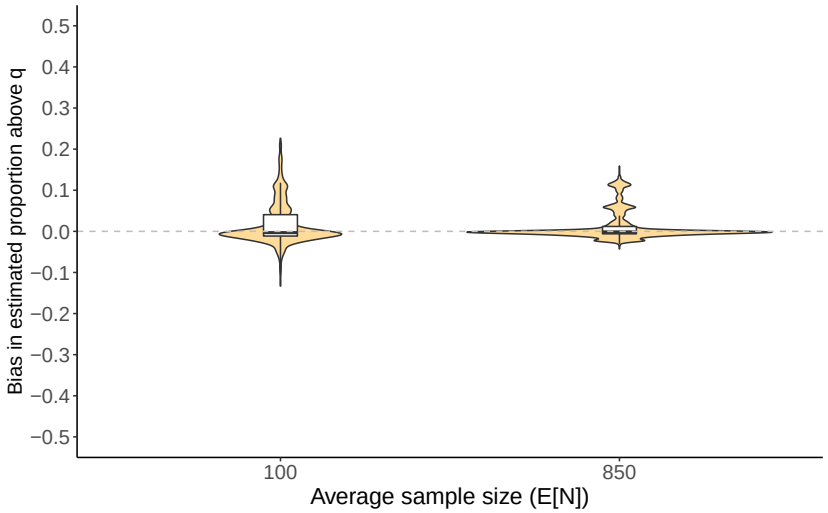

95% CI width for estimated proportion above q
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Bias in estimated difference in proportions
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.5
Absolute error in estimated difference in proportions
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0

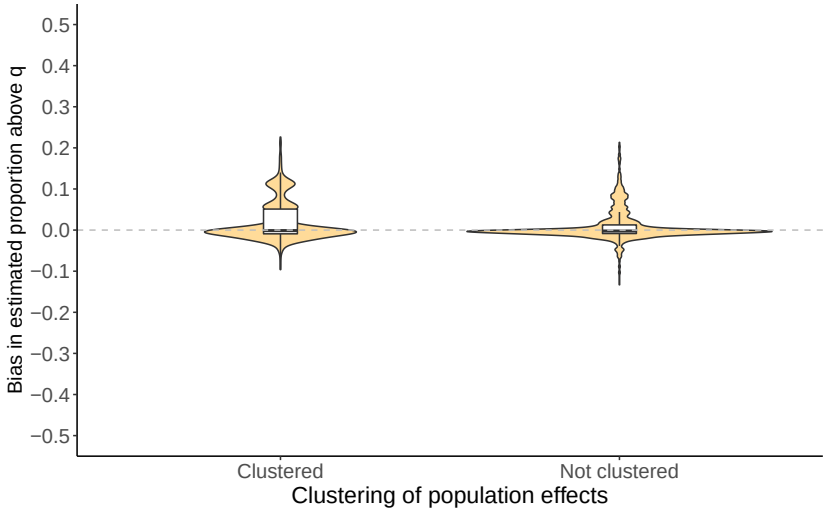


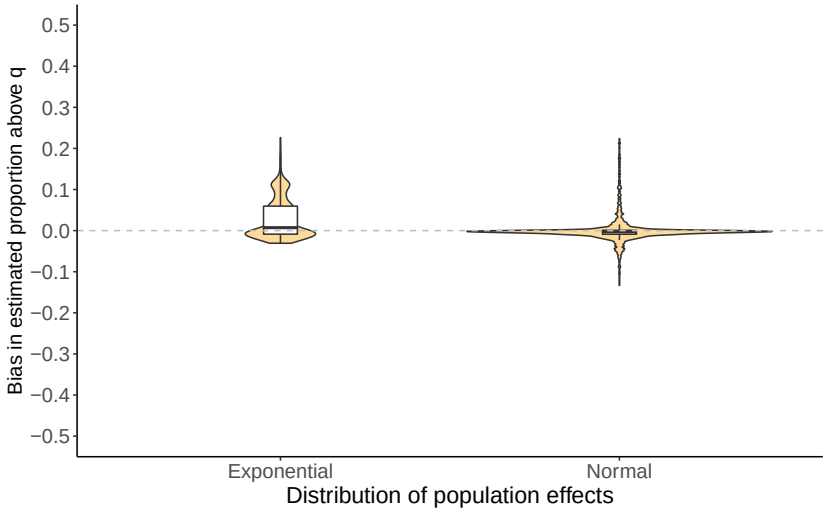
2.4. Violin plots of performance metrics stratified by meta-analysis characteristics (all scenarios)

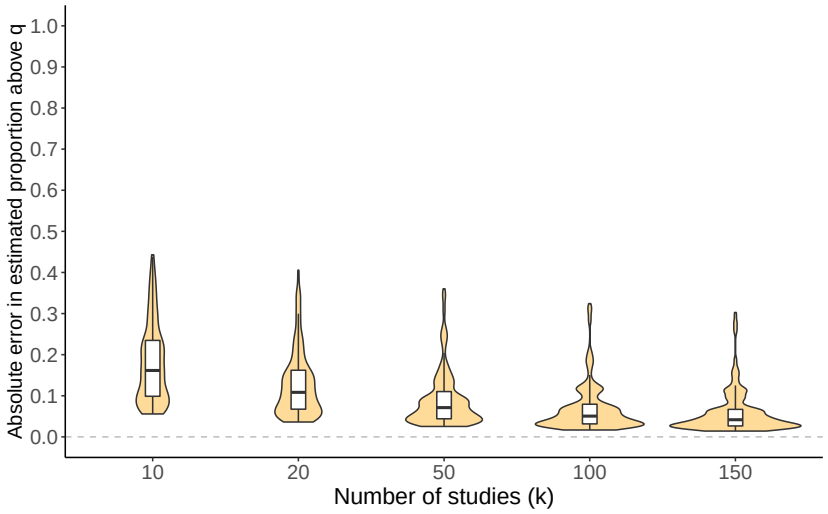
These plots show all scenarios, including those that do not fulfill the guidelines given in the Discussion.

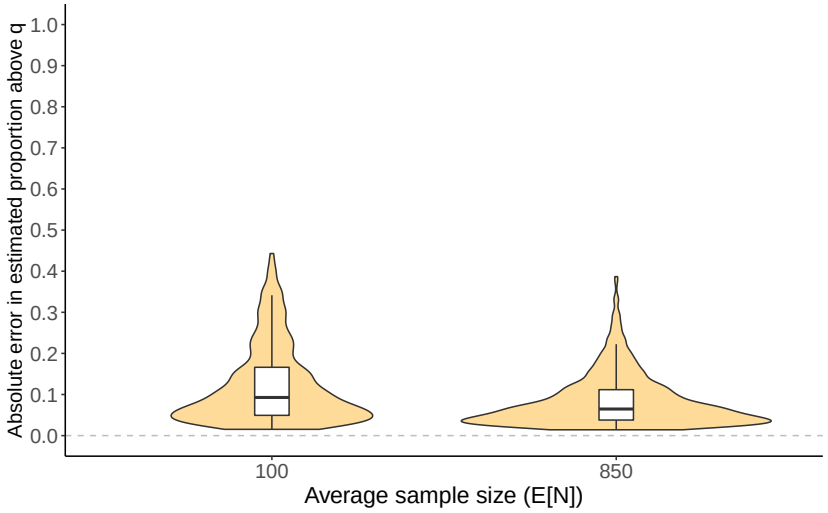

Bias in estimated proportion above q
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.5
10
20
50
100
150
Number of studies (k)


Bias in estimated proportion above q
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.5
100
850
Average sample size (E[N])


Bias in estimated proportion above q
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.5
Clustered
Not clustered
Clustering of population effects

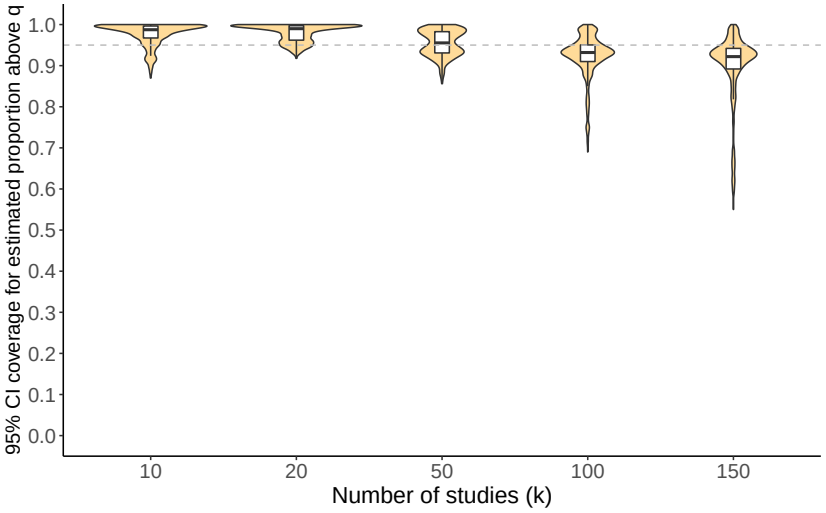

Bias in estimated proportion above q
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.5
Exponential
Normal
Distribution of population effects

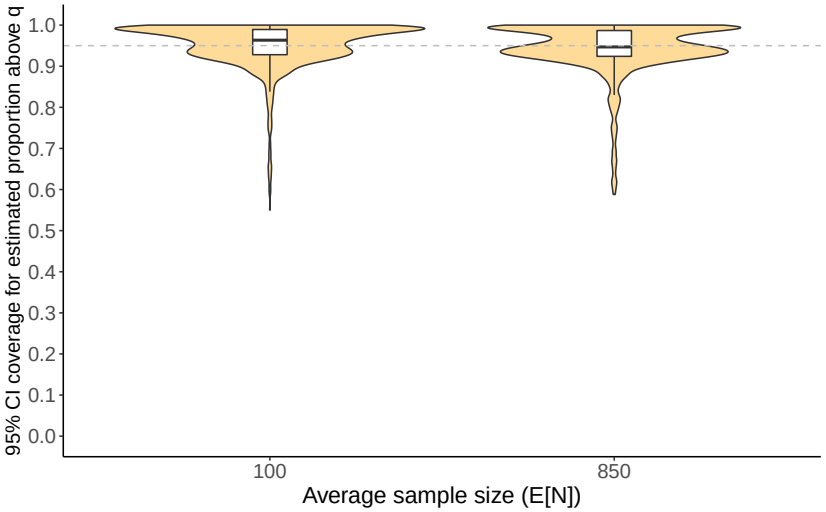

Absolute error in estimated proportion above q
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
10
20
50
100
150
Number of studies (k)

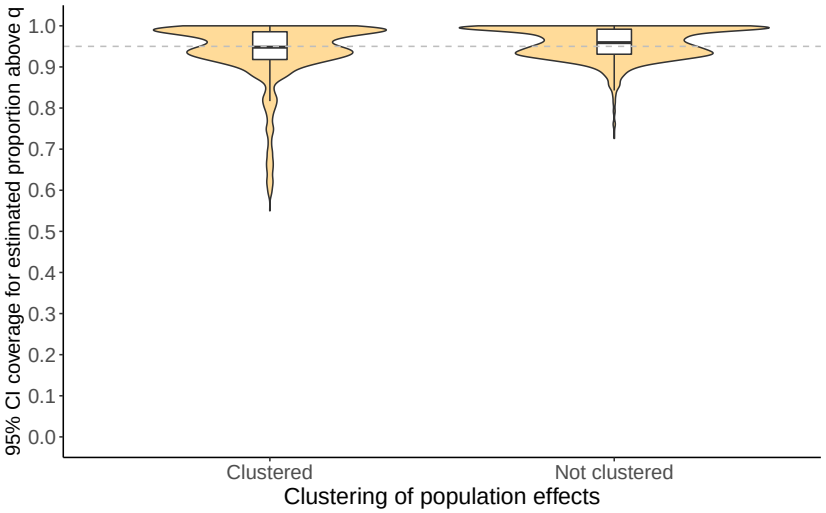

Absolute error in estimated proportion above q
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
100
850
Average sample size (E[N])

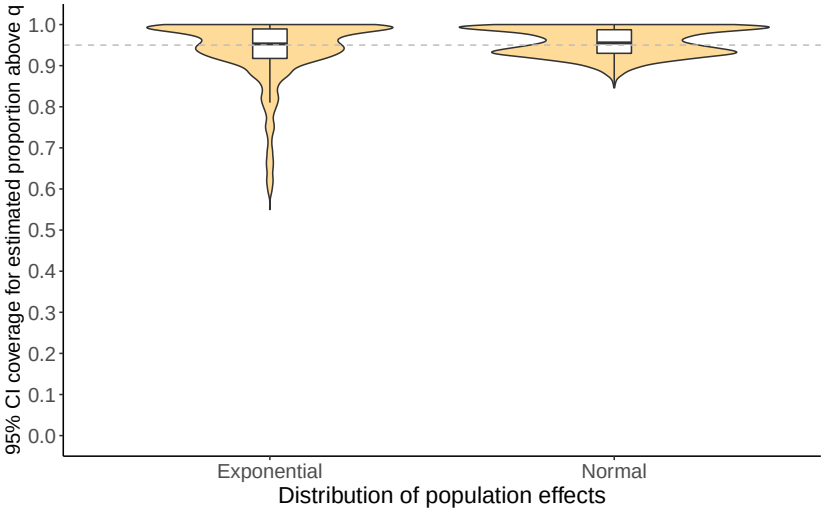

Absolute error in estimated proportion above q
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clustered
Not clustered
Clustering of population effects

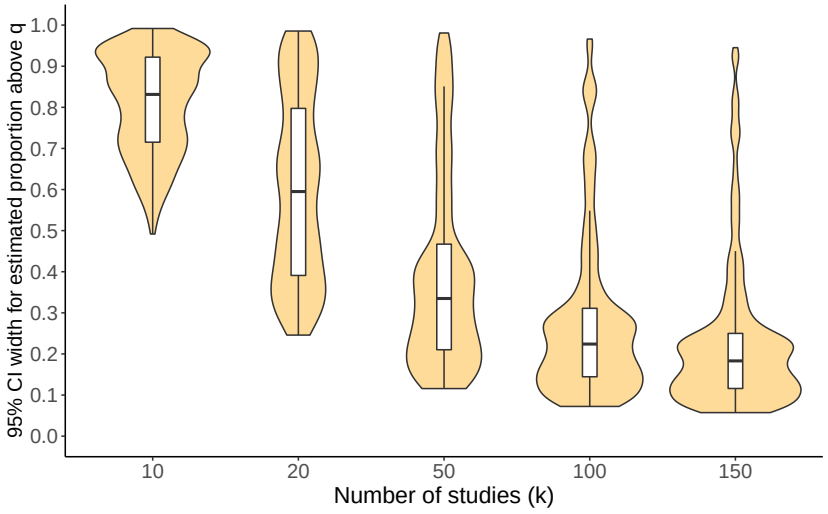

Absolute error in estimated proportion above q
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Exponential
Normal
Distribution of population effects

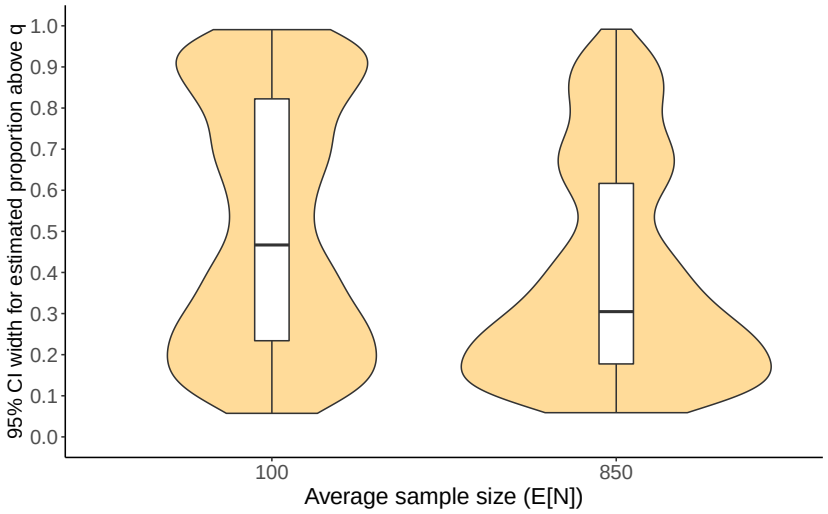

95% CI coverage for estimated proportion above q
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
10
20
50
100
150
Number of studies (k)


95% CI coverage for estimated proportion above q
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
100
850
Average sample size (E[N])

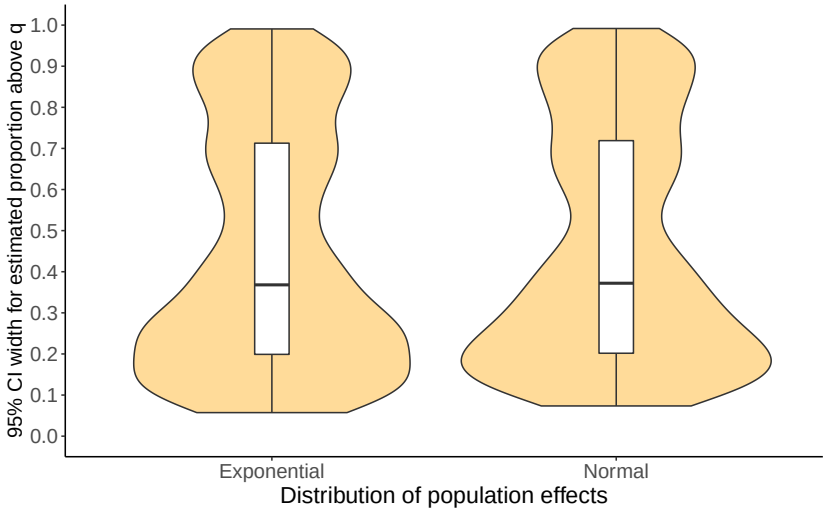

95% CI coverage for estimated proportion above q
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clustered
Not clustered
Clustering of population effects

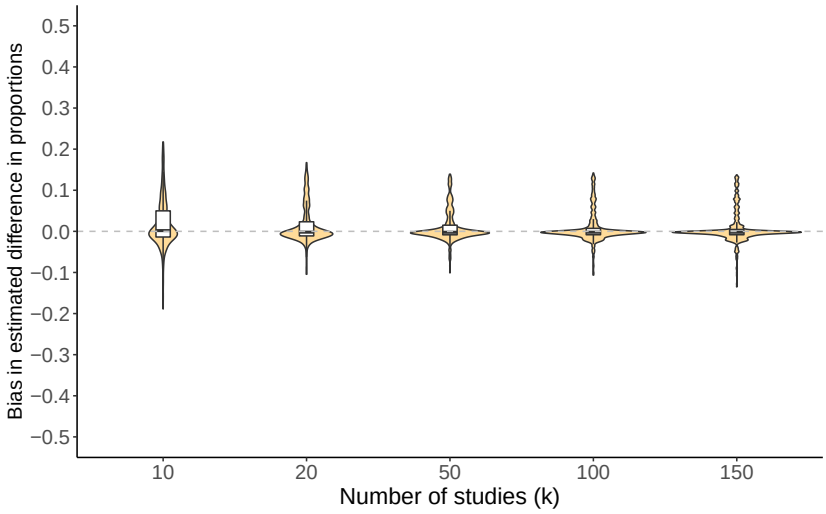

95% CI coverage for estimated proportion above q
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Exponential
Normal
Distribution of population effects

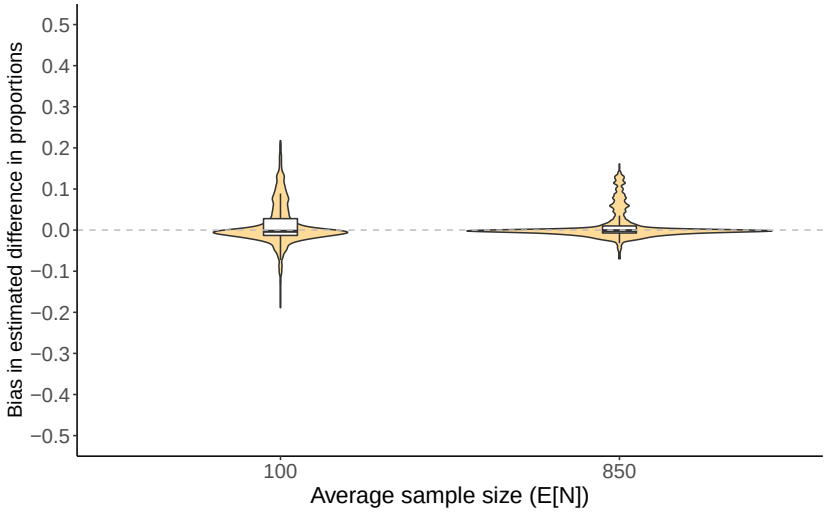

95% CI width for estimated proportion above q
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
10
20
50
100
150
Number of studies (k)

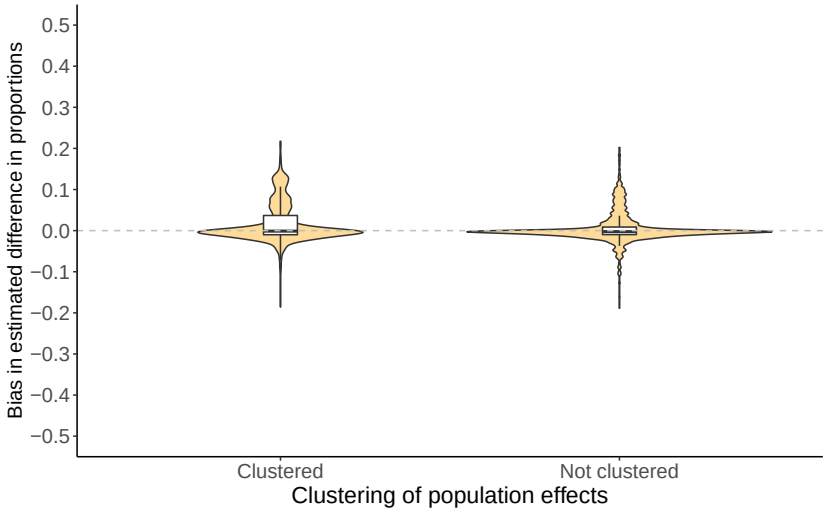

95% CI width for estimated proportion above q
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
100
850
Average sample size (E[N])

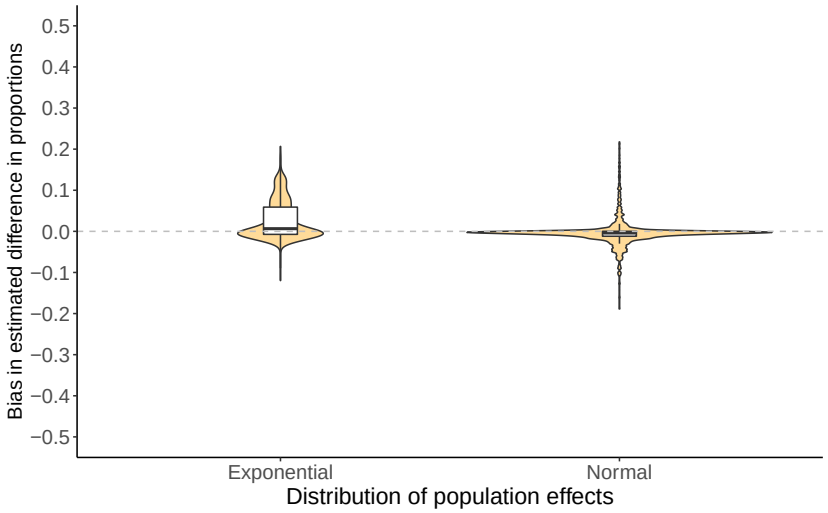

95% CI width for estimated proportion above q
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clustered
Not clustered
Clustering of population effects

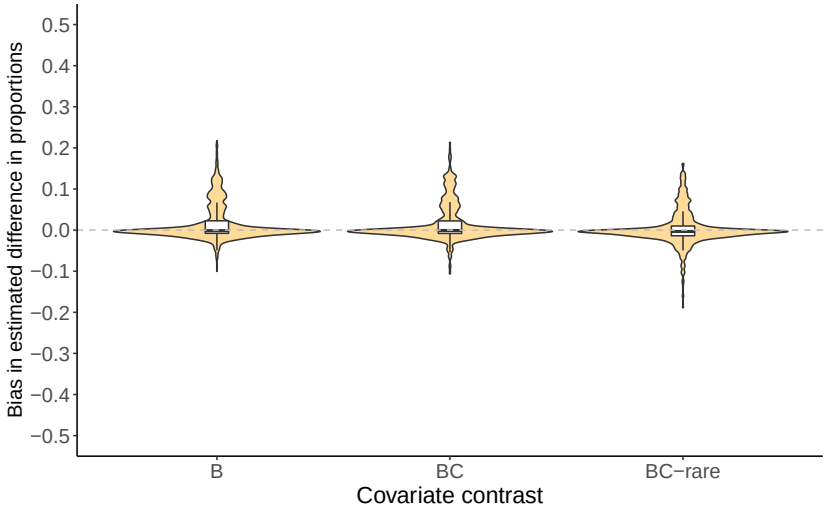

95% CI width for estimated proportion above q
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Exponential
Normal
Distribution of population effects

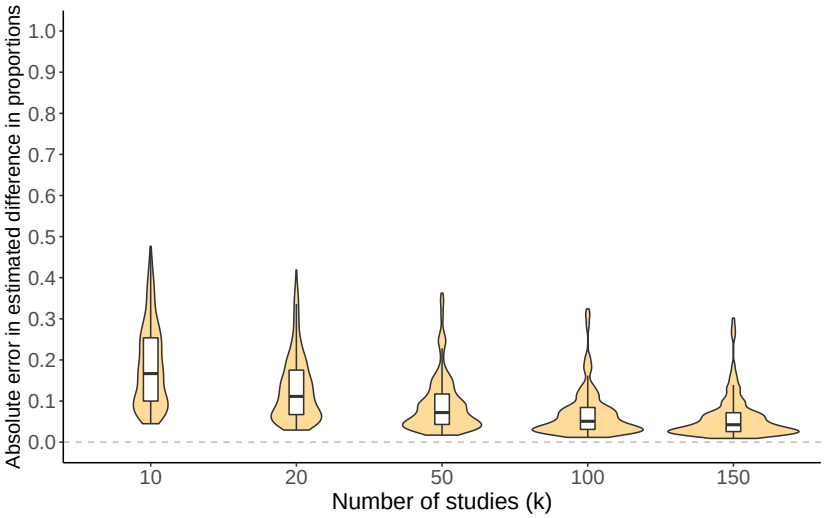

Bias in estimated difference in proportions
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.5
10
20
50
100
150
Number of studies (k)

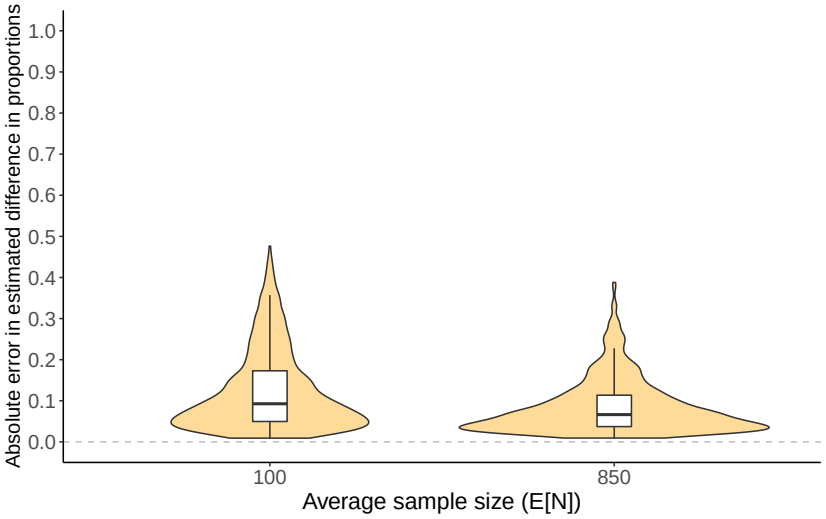

Bias in estimated difference in proportions
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.5
100
850
Average sample size (E[N])

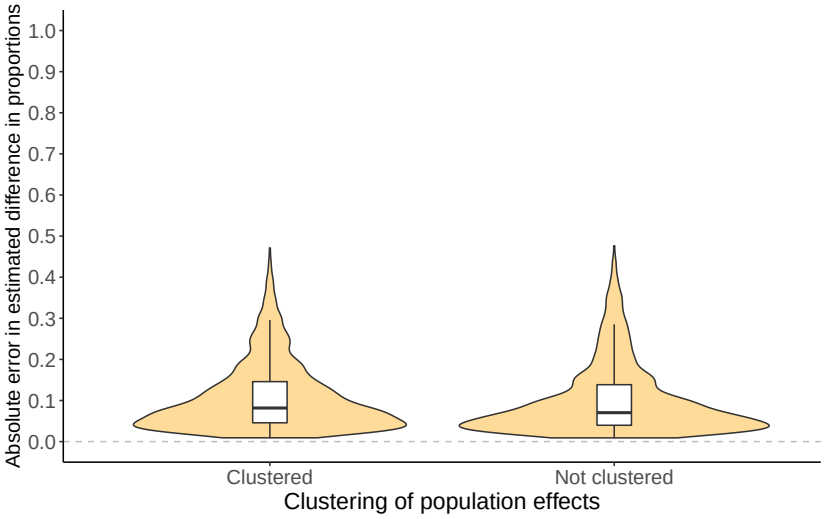

Bias in estimated difference in proportions
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.5
Clustered
Not clustered
Clustering of population effects

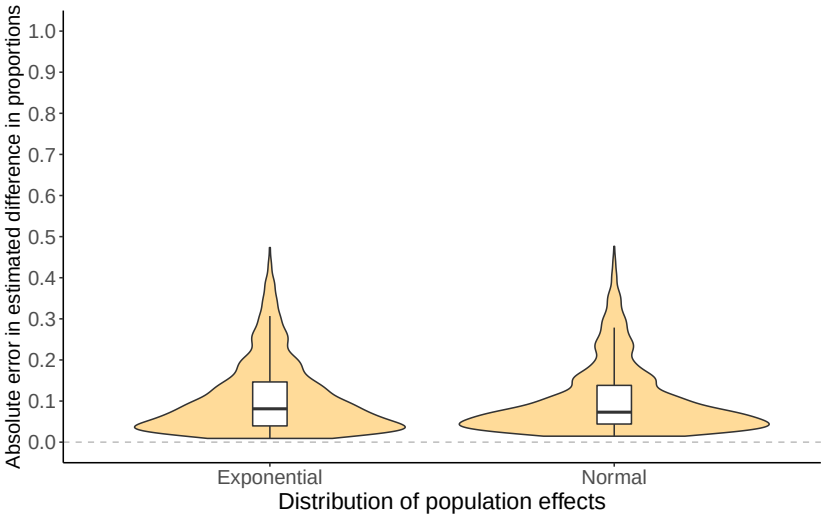

Bias in estimated difference in proportions
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.5
Exponential
Normal
Distribution of population effects

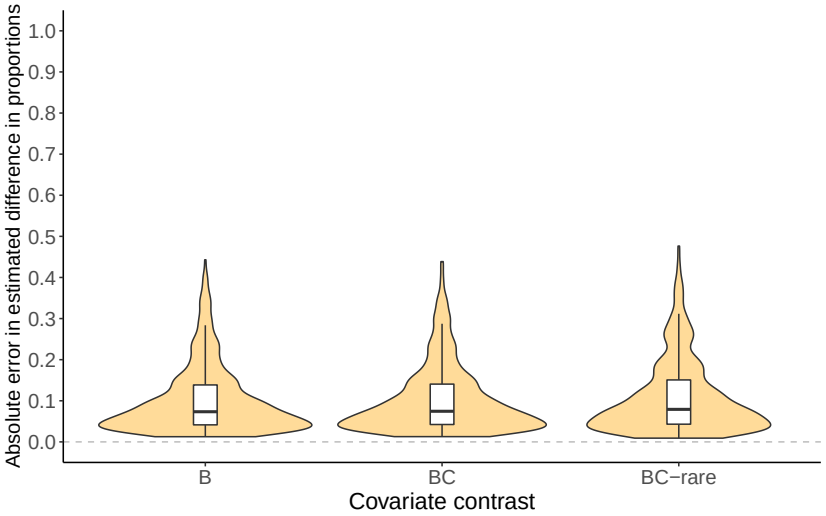

Bias in estimated difference in proportions
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.5
B
BC
BC−rare
Covariate contrast

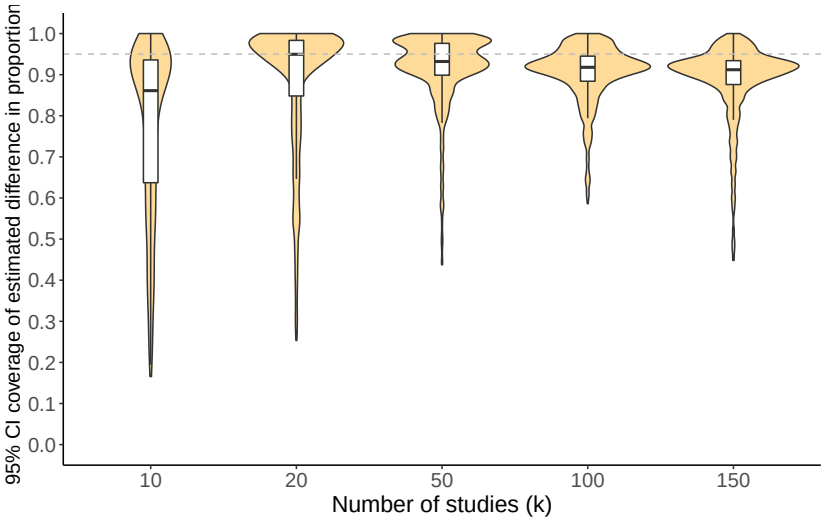

Absolute error in estimated difference in proportions
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
10
20
50
100
150
Number of studies (k)

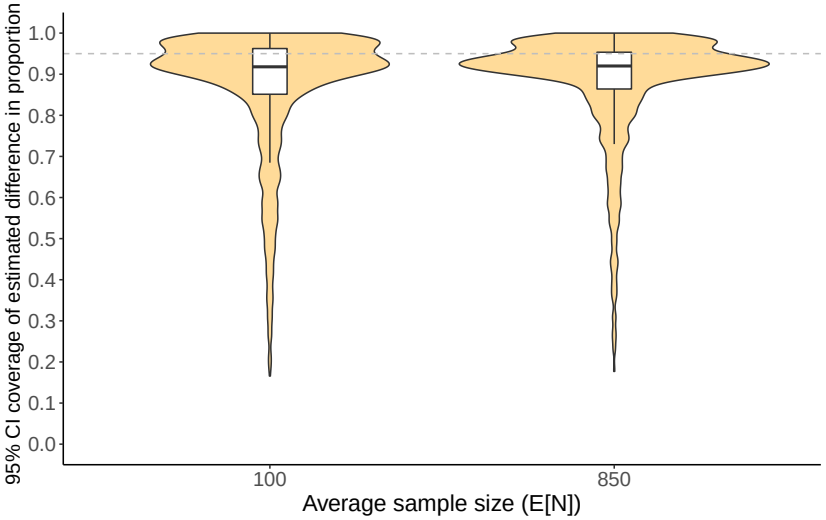

Absolute error in estimated difference in proportions
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
100
850
Average sample size (E[N])


Absolute error in estimated difference in proportions
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clustered
Not clustered
Clustering of population effects

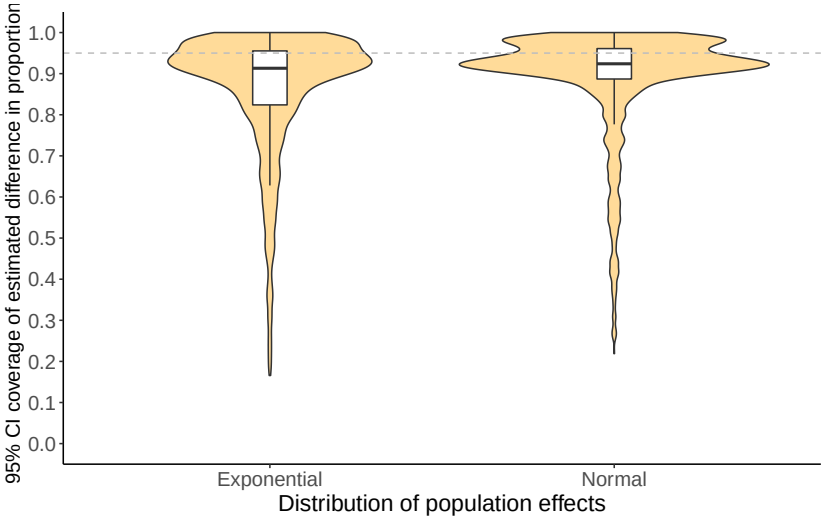

Absolute error in estimated difference in proportions
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Exponential
Normal
Distribution of population effects

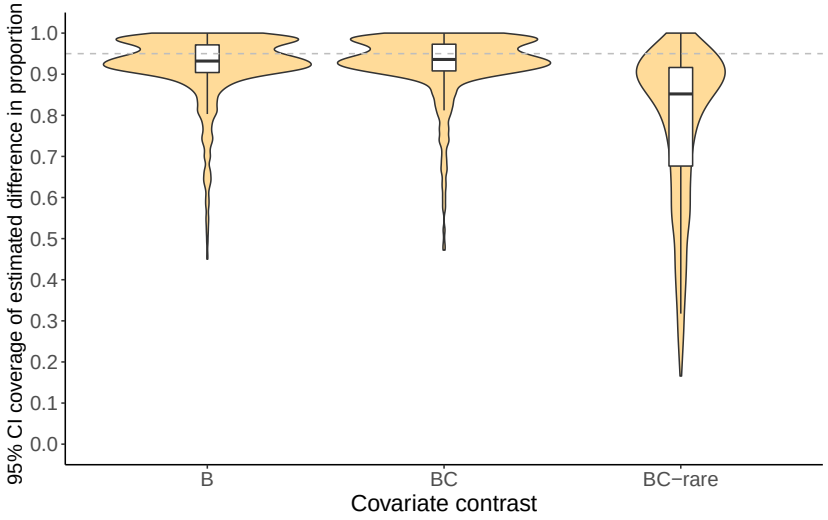

Absolute error in estimated difference in proportions
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
B
BC
BC−rare
Covariate contrast

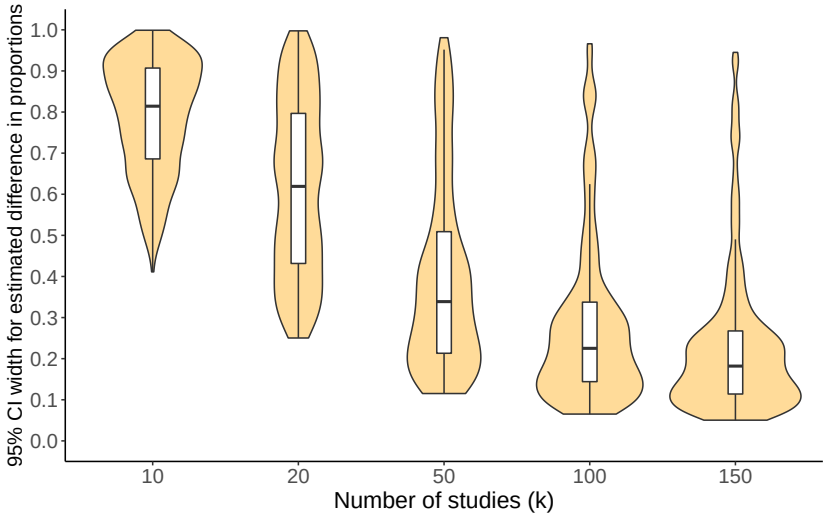

95% CI coverage of estimated difference in proportion
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
10
20
50
100
150
Number of studies (k)

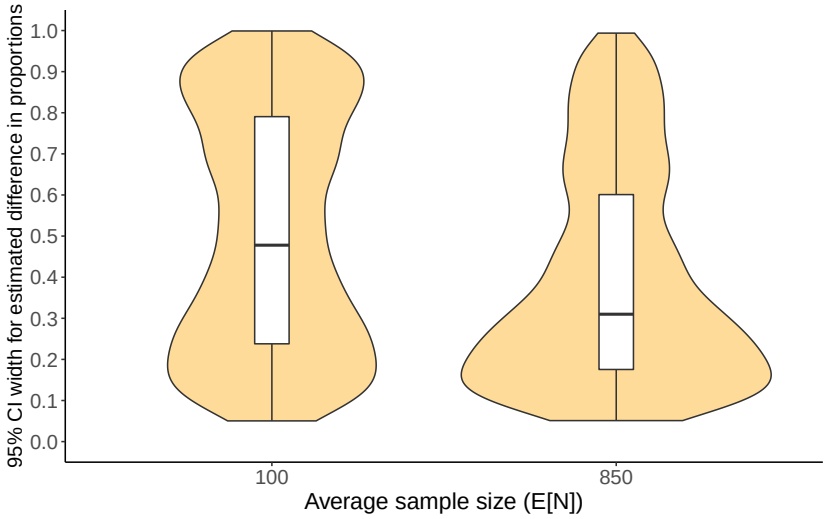

95% CI coverage of estimated difference in proportion
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
100
850
Average sample size (E[N])

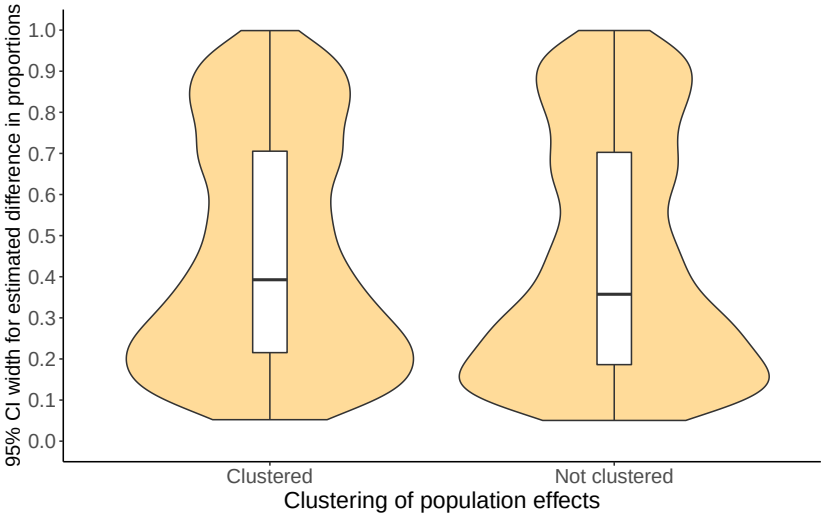

95% CI coverage of estimated difference in proportion
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clustered
Not clustered
Clustering of population effects

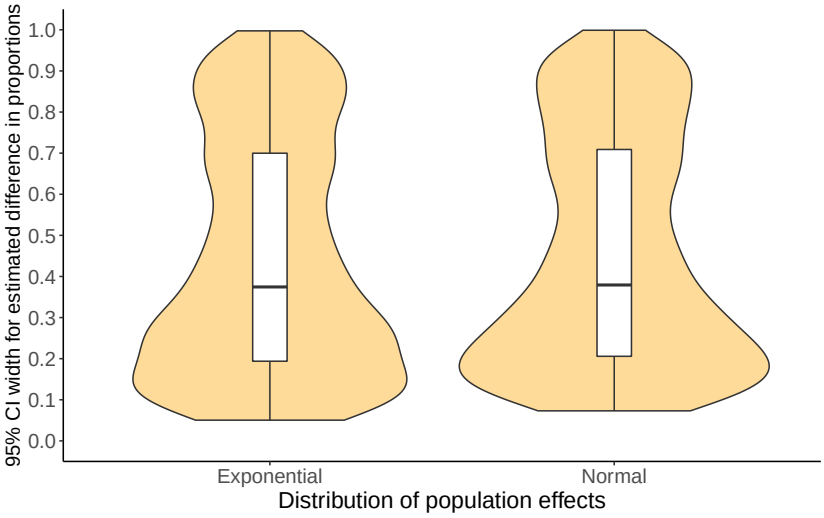

95% CI coverage of estimated difference in proportion
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Exponential
Normal
Distribution of population effects

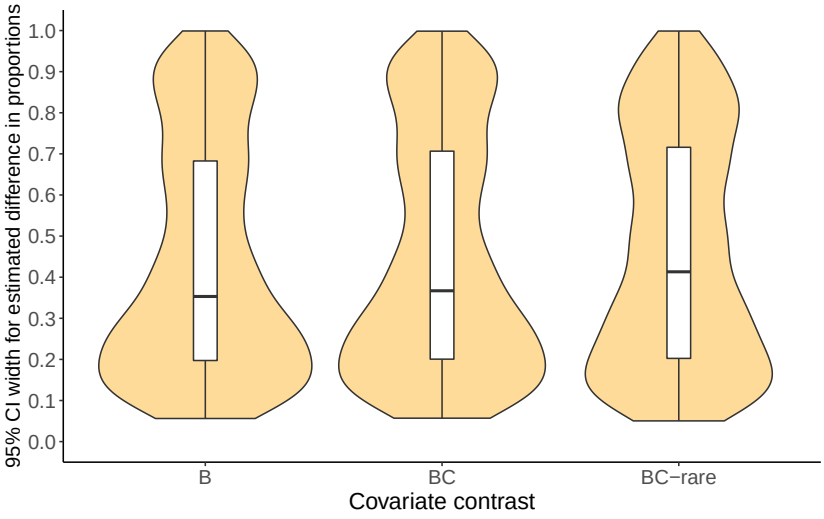

95% CI coverage of estimated difference in proportion
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
B
BC
BC−rare
Covariate contrast


95% CI width for estimated difference in proportions
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
10
20
50
100
150
Number of studies (k)


95% CI width for estimated difference in proportions
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
100
850
Average sample size (E[N])


95% CI width for estimated difference in proportions
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clustered
Not clustered
Clustering of population effects


95% CI width for estimated difference in proportions
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Exponential
Normal
Distribution of population effects


95% CI width for estimated difference in proportions
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
B
BC
BC−rare
Covariate contrast