Bias in meta-analysis

Elena Kulinskaya

School of Computing Sciences,
University of East Anglia

November 2020

Meta-analysis and gold standard

Meta-analysis is a statistical technique for combining evidence from related studies. The primary aim consists in getting a better grip on some question involving an intervention effect or an environmental effect.

The gold standard of medical trials are double-blind prospective studies. A meta-analysis in conservation on the other hand often makes use of studies that were done for other purposes. It is not necessarily true that a meta-analysis leads to secure conclusions. In fact, a meta-analysis may quite possibly give misleading results.

One reason behind such failures are biases. The individual studies used in the meta-analysis are not performed with uniform standards, the data are only available summarily, there may be agenda-driven biases, relevant data may be hidden, etc.

Outline
Biases in ecology and conservation
Diagnosing bias
Quantile (QQ) plots
Funnel plots
Sensitivity analysis
Example: Valkama et al. (2008)

Glossary

- Selection bias = selection of studies not in a random manner, so that the sample obtained is not representative of the population intended to be analyzed
- Publication bias = influence of research findings on the probability of a study being published
- Dissemination bias = dependence of accessibility of research findings on the direction or strength of these findings

Elena Kulinskaya Bias in meta-analysis

Types of publication bias

- Underreporting of statistically non-significant results (the file drawer problem)
- Underreporting of results inconsistent with the current theory

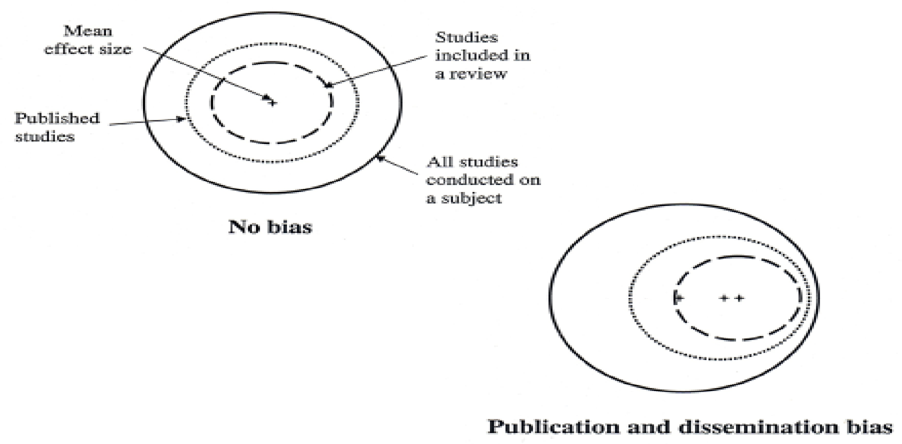
Outline
Biases in ecology and conseration
Diagnosing bias
Quantile (QQ) plots
Funnel plots
Sensitivity analysis
Example: Valkama et al. (2008)

Types of dissemination bias

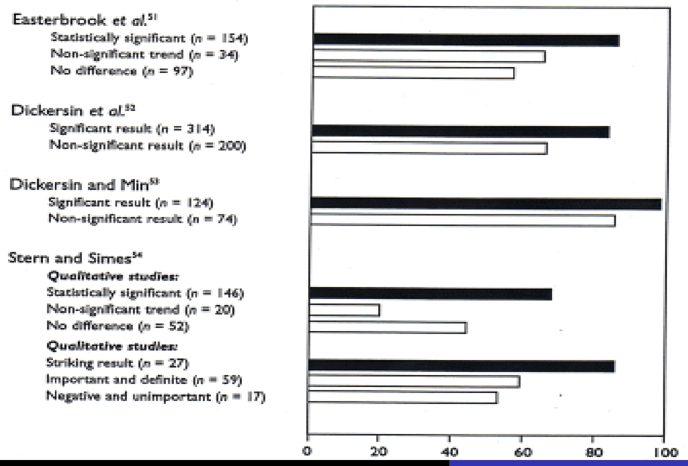
- Time-lag bias
- Place of publication bias
- Citation bias
- Database (indexing) bias
- Language bias
- Grey literature bias

Elena Kulinskaya Bias in meta-analysis

Publication and dissemination bias

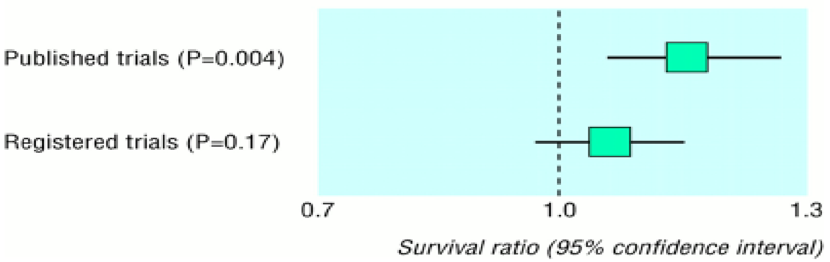


Publish or perish



Publication bias in medicine

Clinical trials assessing survival of patients with advanced ovarian cancer treated with combination chemotherapy compared with monotherapy with alkylating agent



Egger, M. et al. BMJ 1998;316:61-66

Csada et al (1996) The "file-drawer problem" of non-significant results: does it apply to biological research? *Oikos* **76**: 591-593

Journal	Total No of articles	Statistical tests	Non-significant results
American Naturalist	30	26	0
Ecology	22	18	3
Evolution	56	34	4
Nature	20	14	0
Oecologia	69	63	6
Science	20	14	0

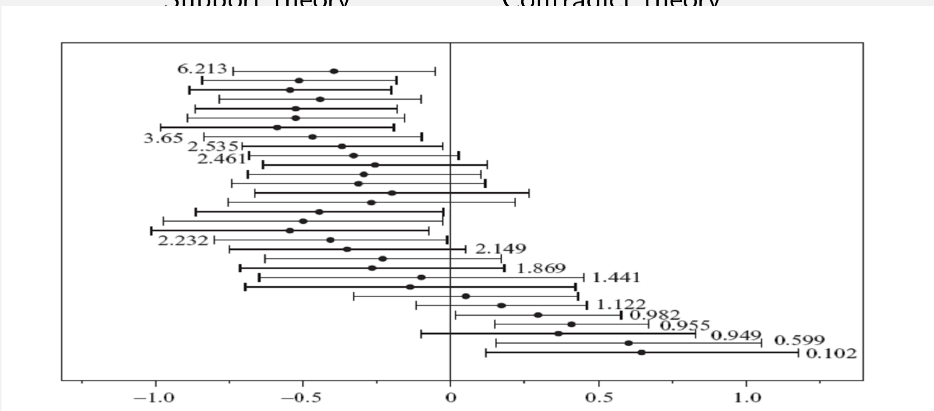
Only 8.6% of the reviewed papers presented non-significant results

Publication bias in ecology, Koricheva 2003 Oikos 102: 397-401

- 93 PhD dissertations on ecological topics defended during 1982-1998 in 12 different universities (6 Finnish and 6 Swedish).
- 187 manuscripts which were neither published nor accepted for publication at the time of thesis defense.
- Percentage of non-significant results: 49% published studies vs 41% unpublished, p=0.021.
- Manuscripts with significant and non-significant results may be submitted to different journals that differ in rejection rates.
- The proportion of non-significant results in a study tended to be negatively associated with the Impact Factor of the journal in which the study was published (r=-0.167, N=126, P=0.063).

The current theory effect, Leimu and Koricheva 2004 Proc R Soc L 271: 1961-1966

Support theory Contradict theory



Dealing with publication bias

- Avoiding it:

  Including unpublished studies and grey literature;
  Using studies in which testing the hypothesis of interest is not the main aim.
- Diagnosing it:

  Direct methods (comparison of results of published vs unpublished studies);
  Statistical methods.
- Sensitivity analyses:

  'Fail-safe number';
  'Trim and fill method (Duval & Tweedie 2000)

Diagnosing Bias

- To some extent, one can diagnose whether or not bias is present by taking a look at the estimated effects $\hat{\theta}_k$ and their standard errors s_k from the K studies.
- If the true effect were constant and equal to θ, then the studentized differences $\frac{\hat{\theta}_k - \hat{\theta}}{s_k}$ should behave approximately like a sample from $N(0, 1). If all the studies were of sufficiently large size n_k > 15$, then one could compare them to the normal distribution.
- Now, suppose the effect being estimated is biased and equal to $\hat{\theta}_k = \theta + b_k$. In this case, s_k does not take account of the bias, s_k estimates the variability around the biased effect θ_k , NOT around the common effect θ.

Diagnosing Bias (continued)

- The estimated combined effect derived from the $\hat{\theta}_k$ will produce an estimate of θ + average bias.
- Unless we assume that the average bias is zero, we are stuck, because the average bias is confounded with the common effect.
- If the biases b_k cancel each other out, we can model them as being random. This is the origin of the random effects model, which says that the individual biases b_k are of no concern, only the size of their variation matters.

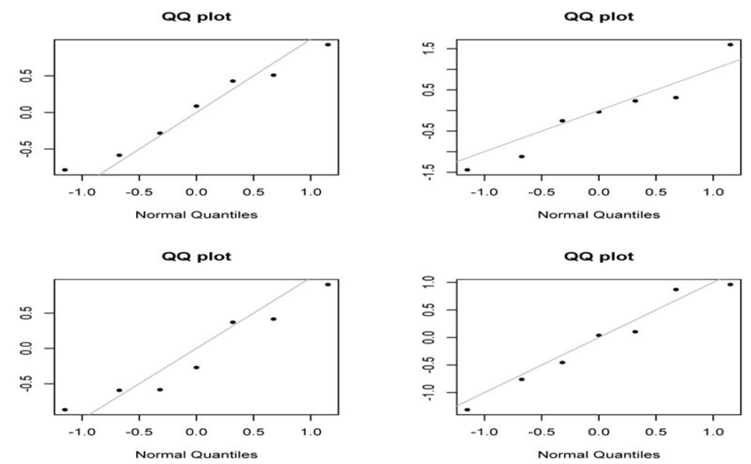
Plotting quantiles

Plotting the K ordered studentized differences $\frac{\hat{\theta}_k - \hat{\theta}}{s_k}$ as a function of the normal quantiles $z_{i/(K+1)}$ is a good way to see patterns. This is called the quantile-quantile (QQ) plot.

If the biases b_k were zero or constant, one would expect the points to wind along the line with slope 1 and intercept 0.

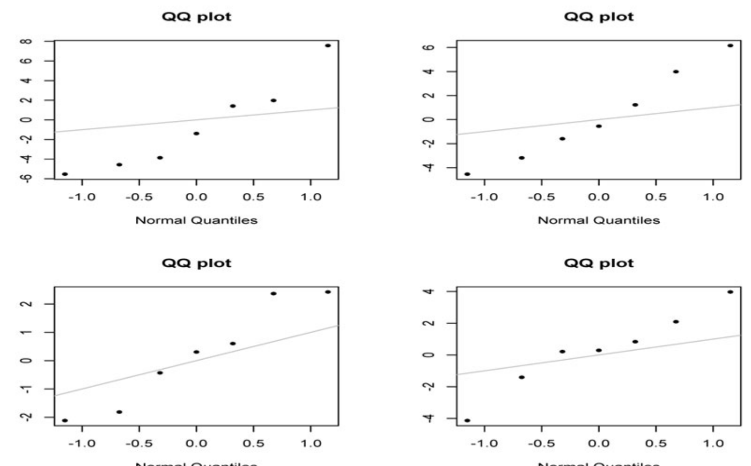
If the biases vary (around 0 or another constant value), the slope increases because there is extra variation.

QQ plots for K=7: no bias or constant bias



Outline
Biases in ecology and conseration
Diagnosing bias
Quantile (QQ) plots
Funnel plots
Sensitivity analysis
Example: Valkama et al. (2008)

QQ plots for K=7: random bias (centered or not)



Elena Kulinskaya
Bias in meta-analysis

The funnel plot

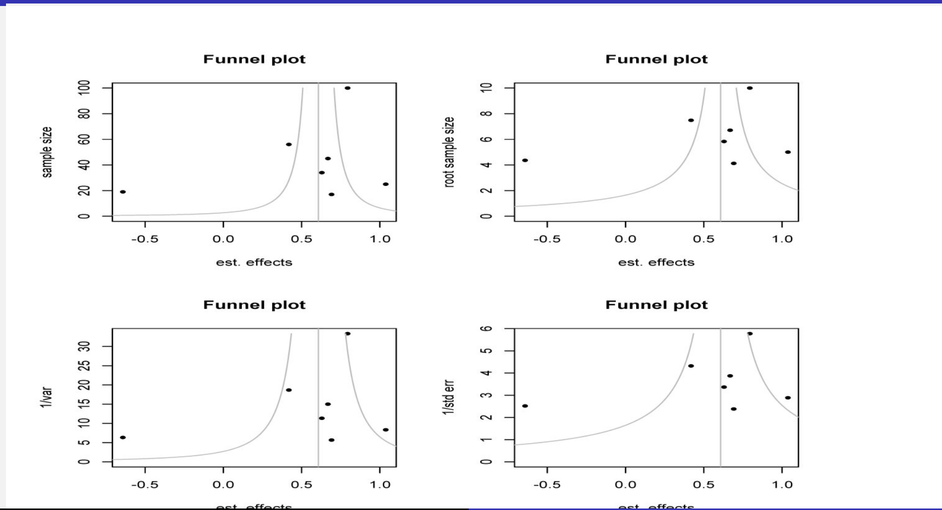
In a funnel plot, one sees the estimated effects together with the sample sizes of the K trials. If the standard deviation of the underlying effect measure were constant in each study, the standard error would be proportional to $1/\sqrt{n_k}$, which gives the plot its name.

Variants of the plot replace n_k by $\sqrt{n_k}$ or by $1/s_k$ or $1/v_k$ or $-s_k$, etc. These choices lead to different shapes of the funnel, but otherwise have no significant visual effect.

The outer lines indicate the region within which 95% of unbiased studies are expected to lie (estimated combined effect $\pm 1.96\times$ its standard error). The solid vertical line corresponds to the estimated combined effect but may be drawn elsewhere.

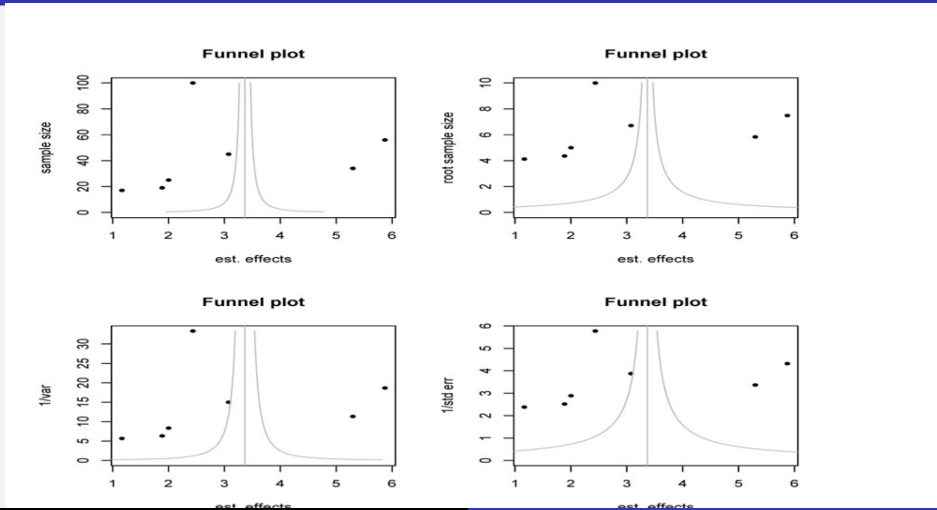
Outline
Biases in ecology and conservation
Diagnosing bias
Quantile (QQ) plots
Funnel plots
Sensitivity analysis
Example: Valkama et al. (2008)

Examples of funnel plots



Elena Kulinskaya Bias in meta-analysis

Examples of funnel plots (biased)



Outline
Biases in ecology and conservation
Diagnosing bias
Quantile (QQ) plots
Funnel plots
Sensitivity analysis
Example: Valkama et al. (2008)

Diagnosing selection bias

If all the trials ever performed were made available, the funnel plot ought to fill out and the point cloud should be in the shape of a funnel.

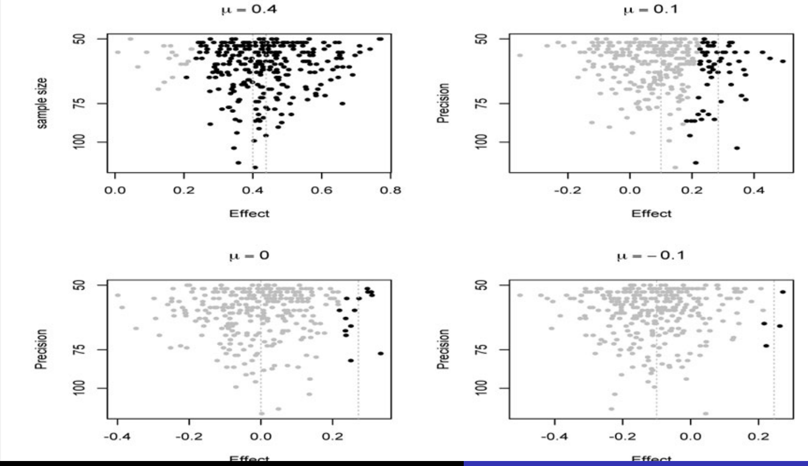
If only selected trials are available for a meta analysis, and if the selection was not done in a random manner, the funnel plot may exhibit blank spaces. Ignoring such selection bias will give a biased and misleading meta-analysis.

An example of this kind is the publication bias, where the trials that did not result in a significant effect are thinned or are missing completely. In the funnel plot, this implies that the left-hand part of the funnel is underrepresented, which if ignored will lead to an overestimation of the combined or meta effect.

Elena Kulinskaya Bias in meta-analysis

Publication bias: equal fixed effect μ, selection $\hat{\mu}_k > 0.2$

Estimated effects move to right from the true effects



The case of the misleading funnel plot

Lau,J, Ioannidis, J.P.A., Terrin,N., Schmid, CH and Olkin,I.
BMJ, 2006 Sep 16; 333(7568): 597–600.
doi: 10.1136/bmj.333.7568.597

"Some methods have become an integral part of systematic reviews and meta-analyses, with reviewers, editors, instructional handbooks, and guidelines encouraging their routine inclusion. However, the evidence for using these methods is sometimes lacking, as the reliance on funnel plots shows."

Possible sources of asymmetry in funnel plots (Sterne at al., BMJ 2011)

Reporting biases

- Publication bias:

Delayed publication (also known as time lag or pipeline) bias
Location biases (eg, language bias, citation bias, multiple publication bias)

- Selective outcome reporting
- Selective analysis reporting

Poor methodological quality leading to spuriously inflated effects in smaller studies

- Poor methodological design
- Inadequate analysis
- Fraud

Possible sources of asymmetry in funnel plots, continued

True heterogeneity

- Size of effect differs according to study size (eg, because of differences in the intensity of interventions or in underlying risk between studies of different sizes)

Artefactual

- In some circumstances, sampling variation can lead to an association between the intervention effect and its standard error

Chance

- Asymmetry may occur by chance, which motivates the use of asymmetry tests

Sterne at al., BMJ 2011;342:d4002 doi: 10.1136/bmj.d4002

Statistical tests for funnel plot asymmetry

A test for funnel plot asymmetry examines whether the association between estimated intervention effects and a measure of study size is greater than might be expected to occur by chance. These tests typically have low power, so even when a test does not provide evidence of asymmetry, bias cannot be excluded.

For outcomes measured on a continuous scale a test based on a weighted linear regression of the effect estimates on their standard errors is straightforward. (Command regtest in metafor.)

Alternatively, nonparametric correlation can be used, Begg and Mazumdar (1994). Command ranktest in metafor uses Kendall's τ.

Limitations of statistical tests for funnel plot asymmetry

When outcomes are binary and the effect measures are odds ratios or relative risks, regression test corresponds to an inverse variance weighted linear regression of the log OR or logRR on its standard error.

There are statistical problems both for regression and correlation tests because the standard error of the log odds ratio is mathematically linked to the size of the odds ratio.
This also applies to the SMD and to the response ratio.

Use sample size not standard error or variance in these plots and tests!

Robust alternatives to funnel plot and related tests

In regtest, specify predictor="ni" for the total sample size, and predictor="ninv" for inverse sample size.

For correlation test, use cor.test(yi,ni,method="kendall") instead.

You can also plot a funnel plot with sample sizes on the Y axis using R command plot(x,y).

plot(xi, ni)

The Fail-safe number (Nfs)

- Fail-safe number (Nfs) estimates the number of non-significant unpublished studies required to eliminate a significant overall effect size
- A fail-safe number of 5N + 10 is considered to provide evidence of a robust average effect size
- Example: N=25, 5N + 10 = 135, Nfs=150 -undetected unpublished studies are unlikely to change the significance of the detected effect size.
- Does not take into account studies reporting effects of opposite direction.
- In metafor: fsn(yi, vi, type="Rosenthal" or "Orwin" or "Rosenberg")

The trim and fill method

This is an iterative non-parametric method attempting to "correct" the asymmetry of a funnel plot. It is recommended to be used as a form of sensitivity analysis of the pooled estimate (Duval & Tweedie, 2000)

The key assumption of the trim and fill method is that studies with the most extreme effect sizes are suppressed. Assuming studies with extreme effects on the left-hand side of the funnel plot are suppressed, the method works by estimating k0, the number of asymmetric studies on the right-hand side of the funnel, i.e. the number of studies that have no counterpart in the left-hand side of the funnel plot.

The k0 asymmetric studies are then trimmed from the right-hand side of the funnel and a pooled estimate is calculated using the remaining symmetrical studies. The trimmed studies and their left-hand side counterparts (i.e. the missing studies) are replaced and a pooled estimate is calculated. Iterations continue until the number k0 and the estimates are stable.

Limitations of the trim and fill method

"When there is large between-study heterogeneity the trim and fill method can underestimate the true positive effect when there is no publication bias.

However, when publication bias is present the trim and fill method can give estimates that are less biased than the usual meta-analysis models. ...

Because we do not know whether funnel plot asymmetry is truly caused by publication bias, and because there is great variability in the performance of different trim and fill estimators and models in various meta-analysis scenarios, we recommend use of the trim and fill method as a form of sensitivity analysis"
(Peters et al, Statist. Med. 2007; 26:4544–4562)

Publication bias in ecology: Møller and Jennions 2002

- 40 published meta-analyses in ecology assessed by using trim and fill method.
- 38% of data sets had a significant number of 'missing'studies after correcting for publication bias,
- 21% of weighted mean effects were no longer significantly different from zero.
- The mean correlation between sample size and standardized effect size was also significantly negative (rs = -0.20, $p < 0.0001$).
- Publication bias may affect the main conclusions of at least 15-21% of meta-analyses in ecology.

Dataset 4: The impact of reed management on wildlife, Valkama et al. (2008)

The authors found that reed management modifies the structure of re-growing reed stands: reed stems were shorter and denser in managed sites than in unmanaged sites.

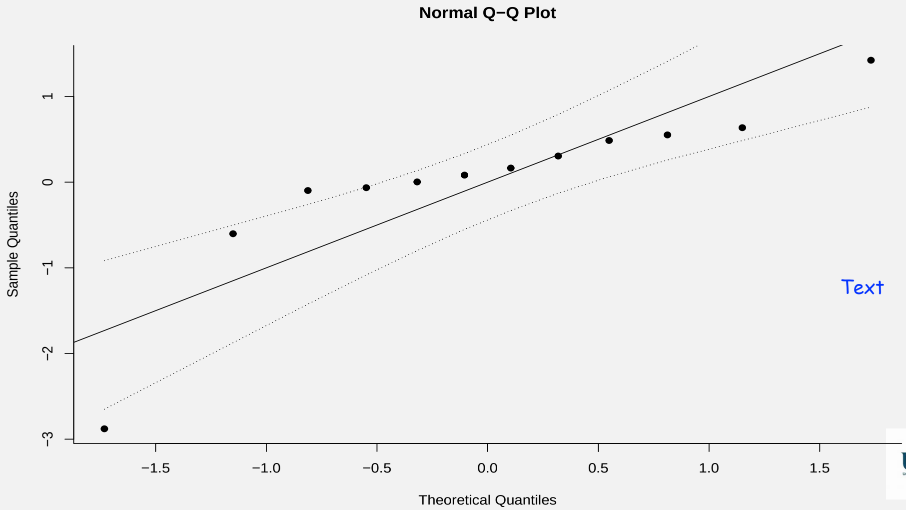
We have performed meta analysis of the stem height of re-growing reed in the session on methodology of MA.

Output for the reed data, REM

```
> summary(valkama_r2)
Random-Effects Model (k = 12; tau^2 estimator: REML)
  logLik  Deviance       AIC       BIC
 -8.2109   16.4217   20.4217   21.2175
tau^2 (estimate of total amount of heterogeneity): 0.2049 (SE = 0.0903)
tau (sqrt of the estimate of total heterogeneity): 0.4527
I^2 (% of total variability due to heterogeneity): 99.68%
H^2 (total variability / sampling variability):    311.90

Test for Heterogeneity:
Q(df = 11) = 1016.4447, p-val < .0001

Model Results:
estimate       se      zval     pval     ci.lb    ci.ub
 -0.2146   0.1329   -1.6145   0.1064   -0.4752   0.0459
```

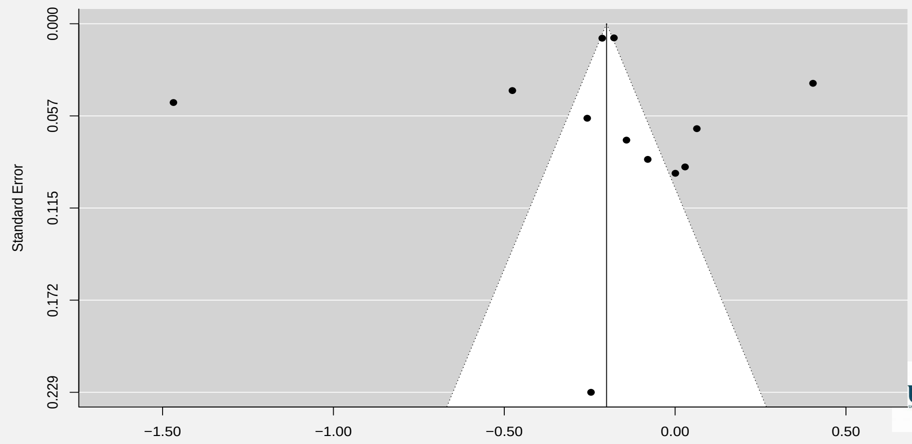

Outline
Biases in ecology and conservation
Diagnosing bias
Quantile (QQ) plots
Funnel plots
Sensitivity analysis
Example: Valkama et al. (2008)

QQ plot for the reed data qqnorm(valkama_r2)



Elena Kulinskaya Bias in meta-analysis

Funnel plot for the reed data



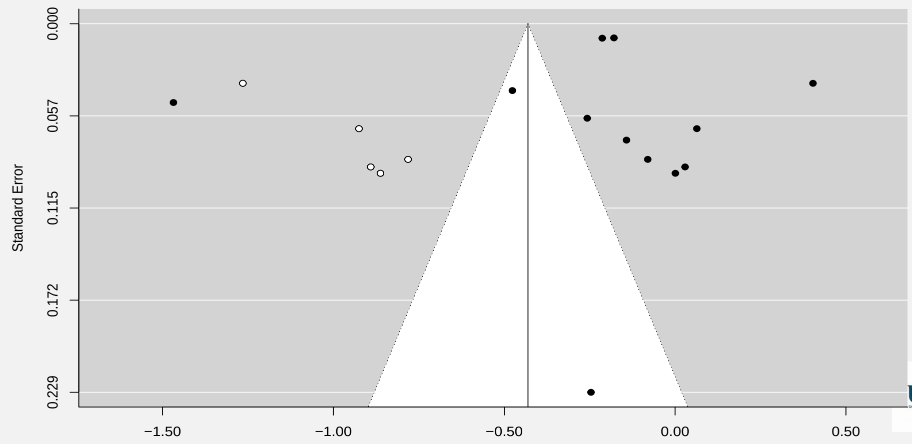
Outline
Biases in ecology and conseration
Diagnosing bias
Quantile (QQ) plots
Funnel plots
Sensitivity analysis
Example: Valkama et al. (2008)

Trimfill for the reed data

```
> trimfill(valkama_r2)
Estimated number of missing studies on the left side: 5

Random-Effects Model (k = 17; tau^2 estimator: REML)
tau^2 (estimate of total amount of heterogeneity): 0.2652 (SE = 0.0961)
tau (sqrt of the estimate of total heterogeneity): 0.5150
I^2 (% of total variability due to heterogeneity): 99.68%
H^2 (total variability / sampling variability):    308.18
Test for Heterogeneity:
Q(df = 16) = 2073.3078, p-val < .0001

Model Results:
estimate       se     zval     pval    ci.lb    ci.ub
 -0.4305   0.1265  -3.4032   0.0007  -0.6785  -0.1826      ***
```

Elena Kulinskaya Bias in meta-analysis

Funnel plot for the trimfilled reed data



Outline
Biases in ecology and conseration
Diagnosing bias
Quantile (QQ) plots
Funnel plots
Sensitivity analysis
Example: Valkama et al. (2008)

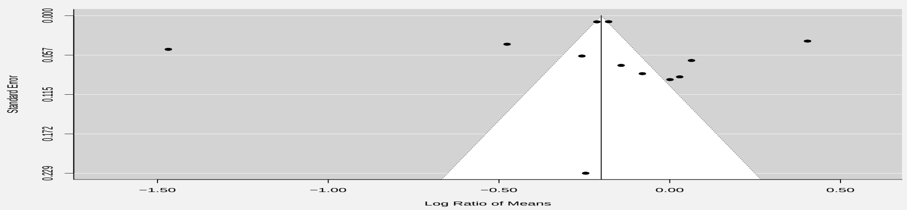
Regression and rank tests for the reed data

```
> regtest(valkama_r2)

Regression Test for Funnel Plot Asymmetry

model:     mixed-effects meta-regression model
predictor: standard error

z = 0.2077, p = 0.8355

> ranktest(valkama_r2)

Rank Correlation Test for Funnel Plot Asymmetry

Kendall's tau = 0.0606, p = 0.8406
```

Correct regression and rank tests for the reed data

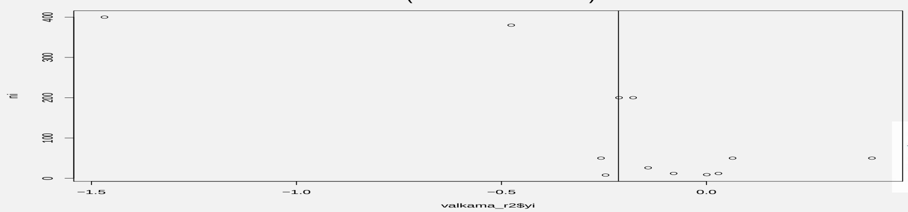
```
> ni<-ni[(valkama$Variable.2=="stem height")
        &(valkama$Species.2=="reed")]
> ni  [1] 200 200   9 400  12  12  26  50  50  50   8 380
> valkama_r2$ni<-ni
> regtest(valkama_r2,predictor="ni")
Regression Test for Funnel Plot Asymmetry
model:     mixed-effects meta-regression model
predictor: total sample size
z = -3.7165, p = 0.0002
> cor.test(valkama_r2$yi,ni,method="kendall")
        Kendall's rank correlation tau
data:  valkama_r2$yi and ni
z = -1.5986, p-value = 0.1099
alternative hypothesis: true tau is not equal to 0
sample estimates:
       tau -0.3624858
```


Funnel plots for the reed data



plot(valkama_r2$yi,ni)

abline(v=valkama_r2$b)



Conclusions

We have learned to appreciate the difficulties inherent in attempting to combine statistical evidence:

1. one of the major foes of meta-analyses are biases and heterogeneities of various kinds;
2. if we take such biases into account, then the p-value and the confidence interval resulting from the meta-analysis change
3. in an idealized case (distribution assumption and/or knowledge of the selection mechanism), one can reconstruct the hidden influence of publication bias.

It is sometimes said that a meta-analysis may result in a very precise false statement. One must be aware of this danger.