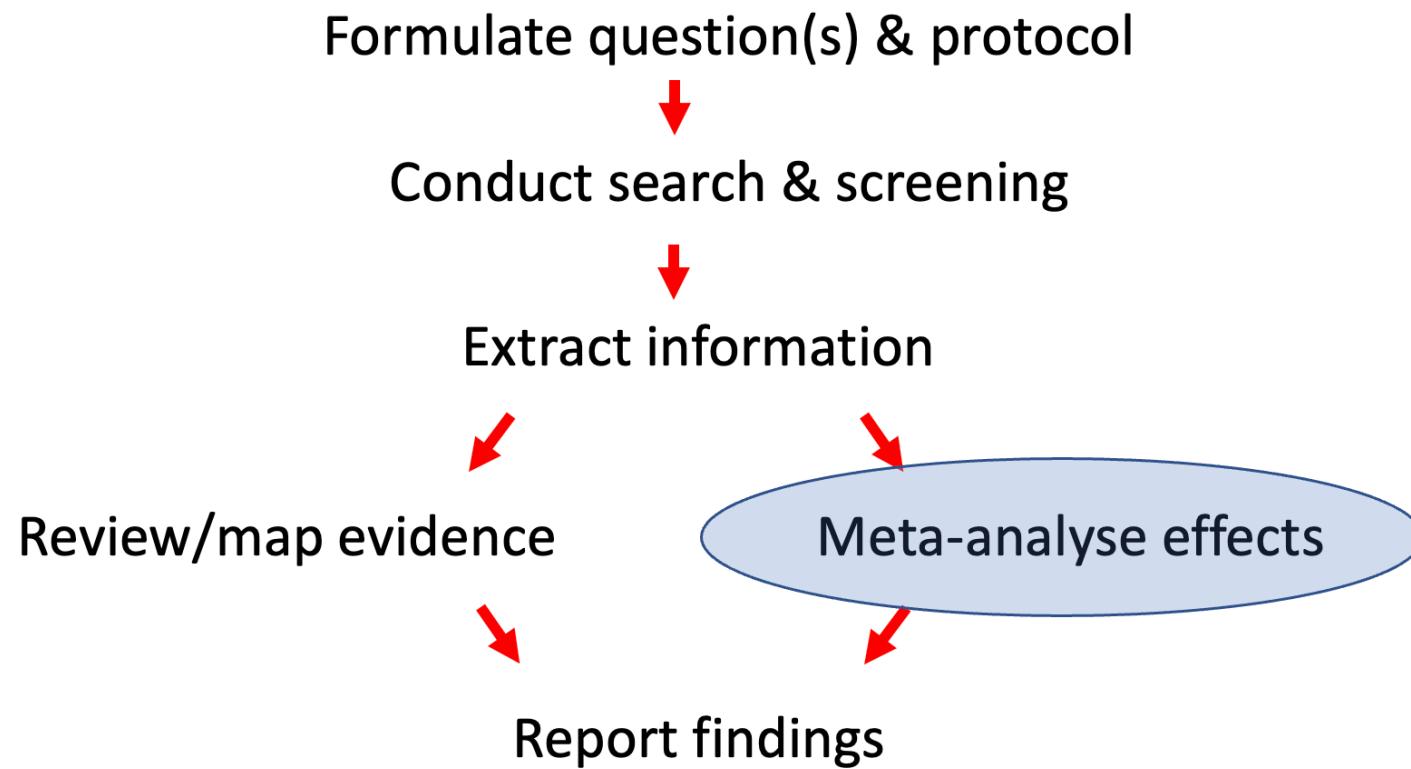


# **SOLES codeRs**

Meta-analysis 2: models & analyses

Thomas White

# The evidence synthesis pipeline



# Extracting and calculating effects



# What is an effect size?

To synthesise a diverse evidence base in the context of a meta-analysis, we need to quantify the **outcomes** of individual studies in a common *numerical* currency.

**Effect sizes are that currency.**

# What is an effect size?

*An **effect size** is a number describing a quantity or relationship. It describes strength and (often) direction of an outcome.*

May be *descriptive* if it's a sample-based estimate of a quantity. Or *comparative* if it estimates the relationship between two variables in a population, or an estimate of a quantity.

Also known as a **outcome measure**, **treatments effect**, or **effect** depending on the field.

# Selecting effects

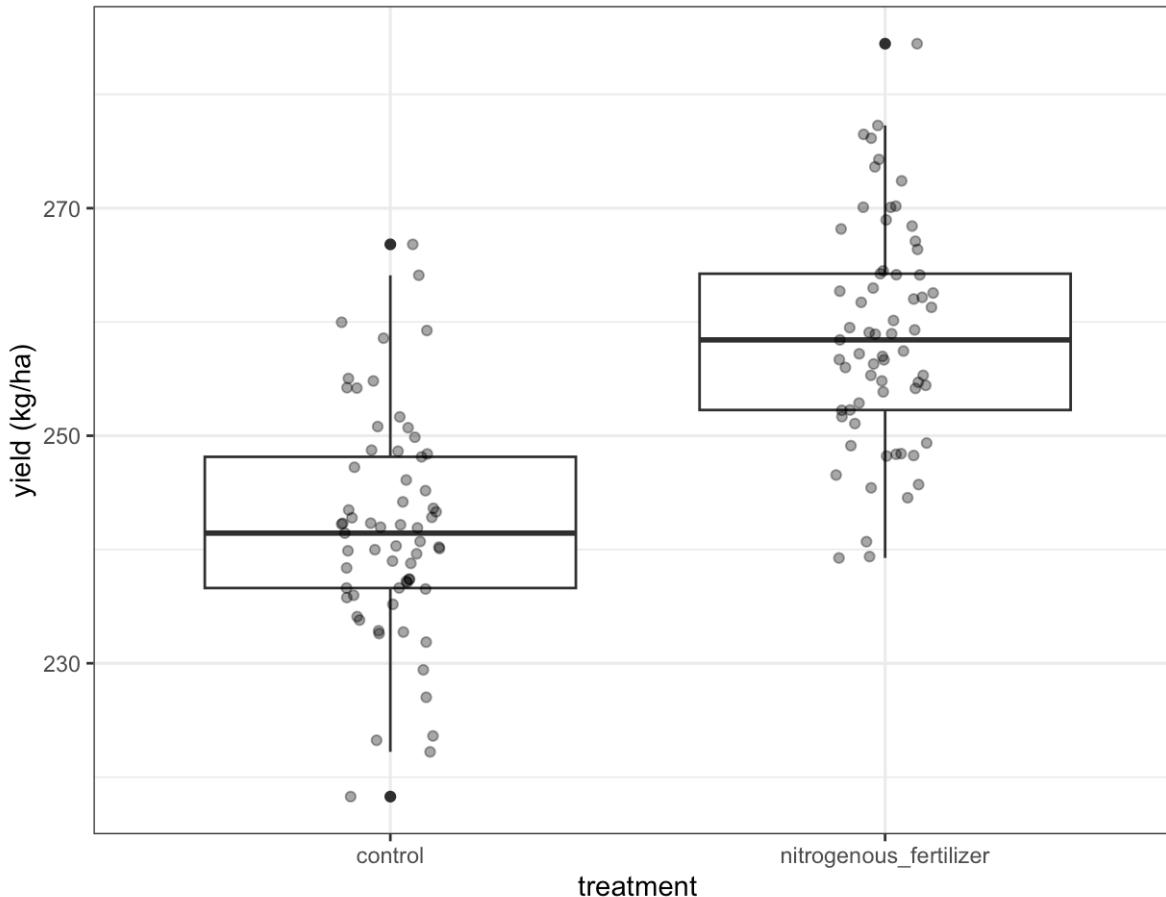


# What are we trying to measure?

Three common cases

# What are we trying to measure?

1: A difference in the **means** of groups.



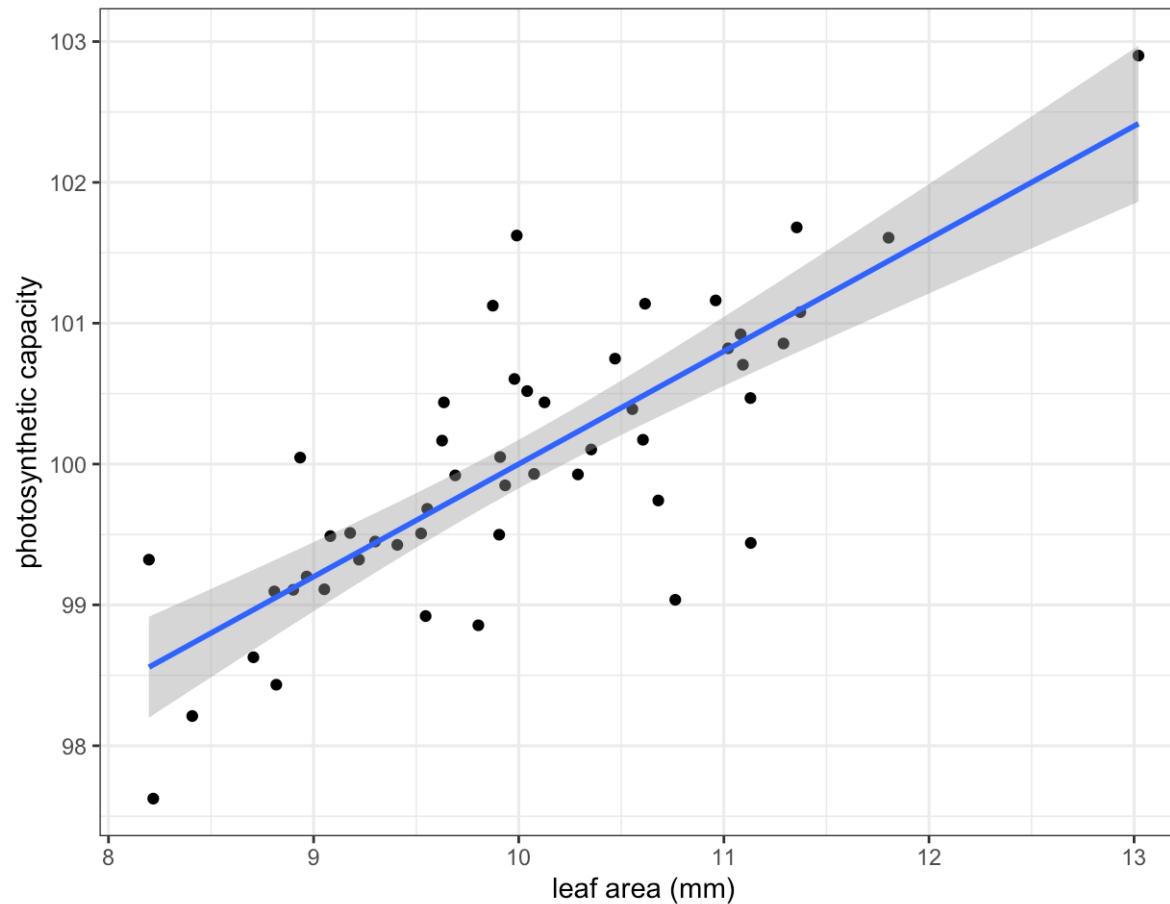
# What are we trying to measure?

2: A difference in **binary outcomes** between groups.

	dead	alive
pesticide	29	21
control	8	42

# What are we trying to measure?

3: A **continuous association** between groups.



# Common effect size measures

1. A difference in the **means** of groups.

- Standardized mean difference (Hedge's  $d$  or  $g$ )
- Raw (unstandardized) mean difference ( $D$ )
- Response ratios ( $RR$ )

2. A difference in **binary outcomes** between groups.

- Odds ratio ( $OR$ )  $\leftrightarrow$  (Log odds ratio)
- Risk ratio ( $RR$ )
- Risk difference ( $RD$ )

3. A **continuous association** between groups.

- Correlation coefficient (Pearson's  $r$ )  $\leftrightarrow$  (Fisher's  $z$ )

N.B Samples are always only samples, so measures always have *error*

# Common effect size measures

1. A difference in the **means** of groups.

- **Standardized mean difference (Hedge's  $d$  or  $g$ )**
- Raw (unstandardized) mean difference ( $D$ )
- Response ratios ( $RR$ )

2. A difference in **binary outcomes** between groups.

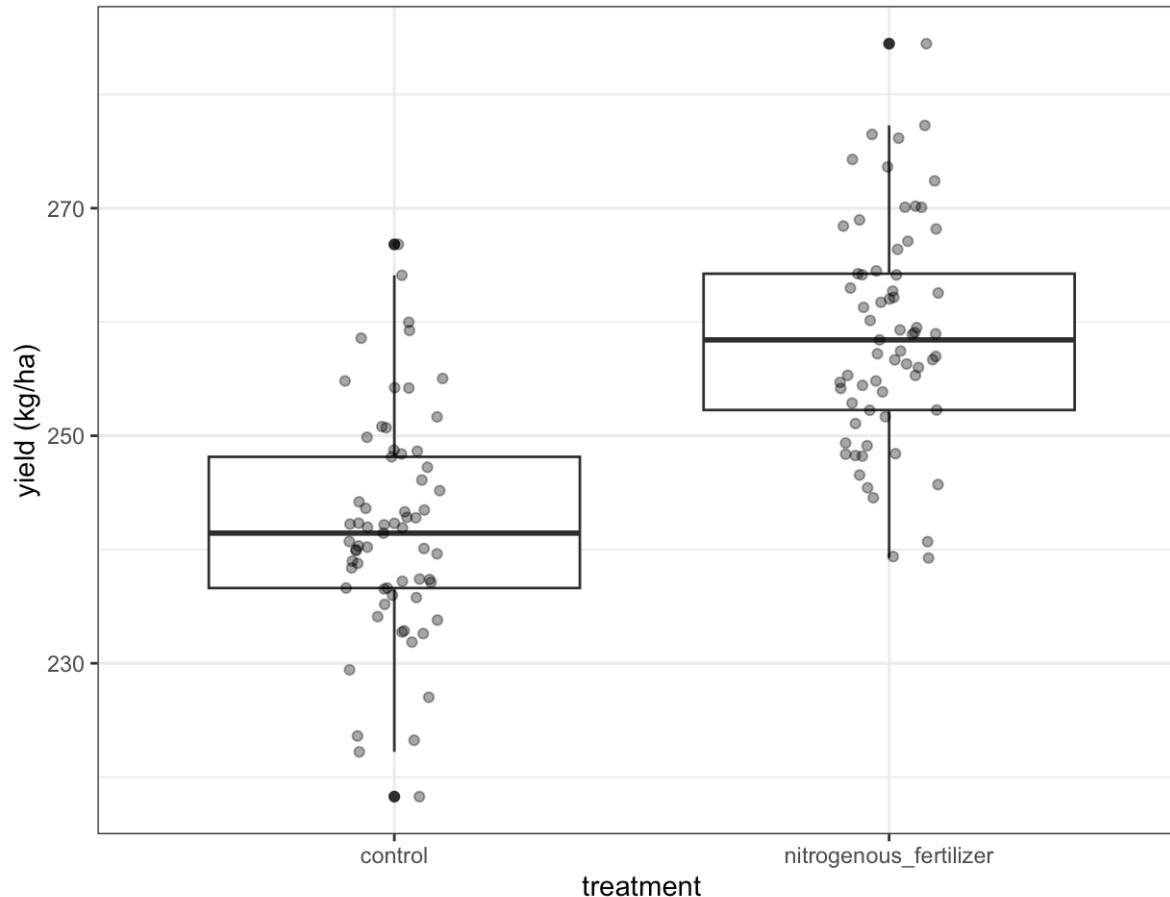
- **Odds ratio ( $OR$ )**  $\leftrightarrow$  (Log odds ratio)
- Risk ratio ( $RR$ )
- Risk difference ( $RD$ )

3. A **continuous association** between groups.

- **Correlation coefficient (Pearson's  $r$ )**  $\leftrightarrow$  (Fisher's  $z$ )

# The Standardized Mean Difference

Useful when quantifying a difference in the **means** of groups.



# The Standardized Mean Difference

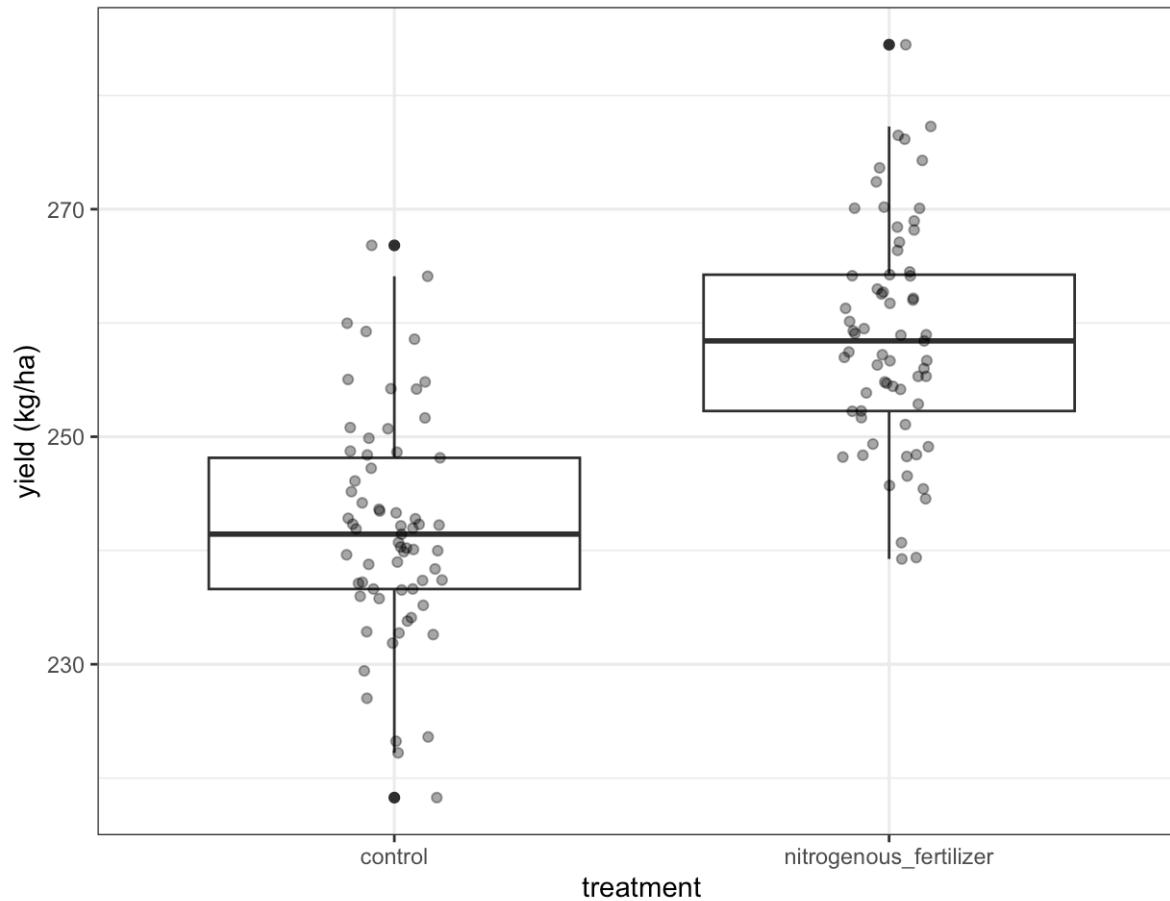
$$g = \frac{Y_1 - Y_2}{S_{\text{pooled}}} J$$

*Seeks to capture the difference between the two distributions, and how each represents a distinct cluster of scores, even if they do not measure exactly the same outcome*

*Measures the distance between two distributions in units of pooled standard-deviations*

# The Standardized Mean Difference

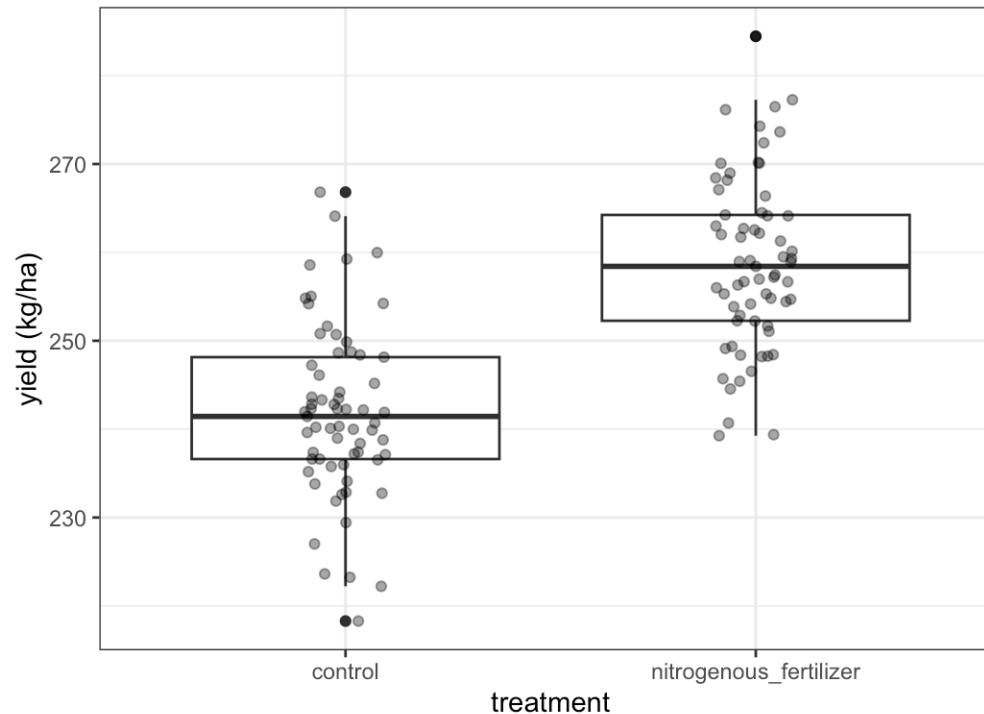
How does the addition of nitrogen-rich fertilizer affect the yield of wheat?



```
dat_yield %>%  
  group_by(treatment) %>%  
  summarise(mean = mean(yield),  
           sd = sd(yield),  
           n = n())
```

# The Standardized Mean Difference

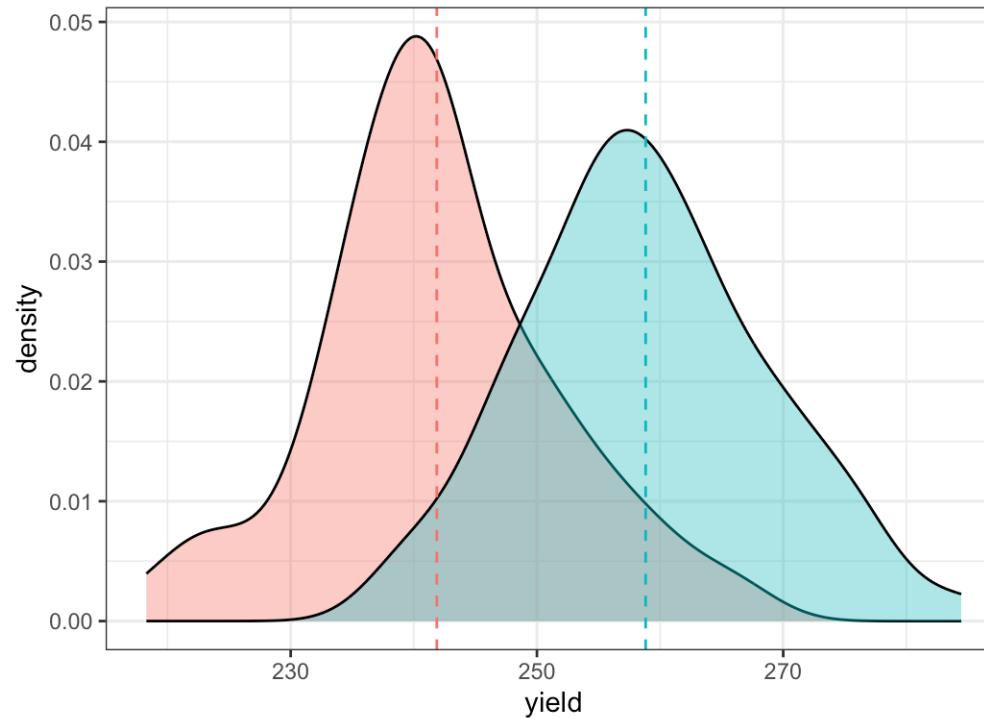
How does the addition of nitrogen-rich fertilizer affect the yield of wheat?



$$g = 1.74$$

# The Standardized Mean Difference

How does the addition of nitrogen-rich fertilizer affect the yield of wheat?



$$g = 1.74$$

# The Standardized Mean Difference

How does the addition of nitrogen-rich fertilizer affect the yield of wheat?

$$g = 1.74$$

Interpretation. ~ish. Not really. (Cohen 1988):

- $0.2$  = ‘small’
- $0.5$  = ‘medium’
- $0.8$  = ‘large’

# The Odds Ratio

For measuring a difference in **binary outcomes** between groups.

# The Odds Ratio

For measuring a difference in **binary outcomes** between groups.

- The **odds** of an event are the number of events that produce an outcome divided by the number of events that do not. Or is the probability of ‘success’ divided by the probability of ‘failure’.

$$O = \frac{1}{6-1} = 0.2$$

- The **odds ratio** estimates the odds of an event happening in one group relative to the odds of the same event happening in the other group.

# The Odds Ratio

*How effective is pesticide X?*

	dead	alive
pesticide	29	21
control	8	42

$$\text{OR} = \frac{29/21}{8/42} = \frac{1.38}{0.19} = 7.25$$

# The Odds Ratio

*How effective is pesticide X?*

$$\text{OR} = \frac{29/21}{8/42} = \frac{1.38}{0.19} = 7.25$$

Insects in patches that received a pesticide treatment had **7.25 times the odds of dying** than those in the pesticide-free control.

N.B. No mention of “risk,” “likely/likelihood,” or “probability”

# The Odds Ratio

*How effective is pesticide X?*

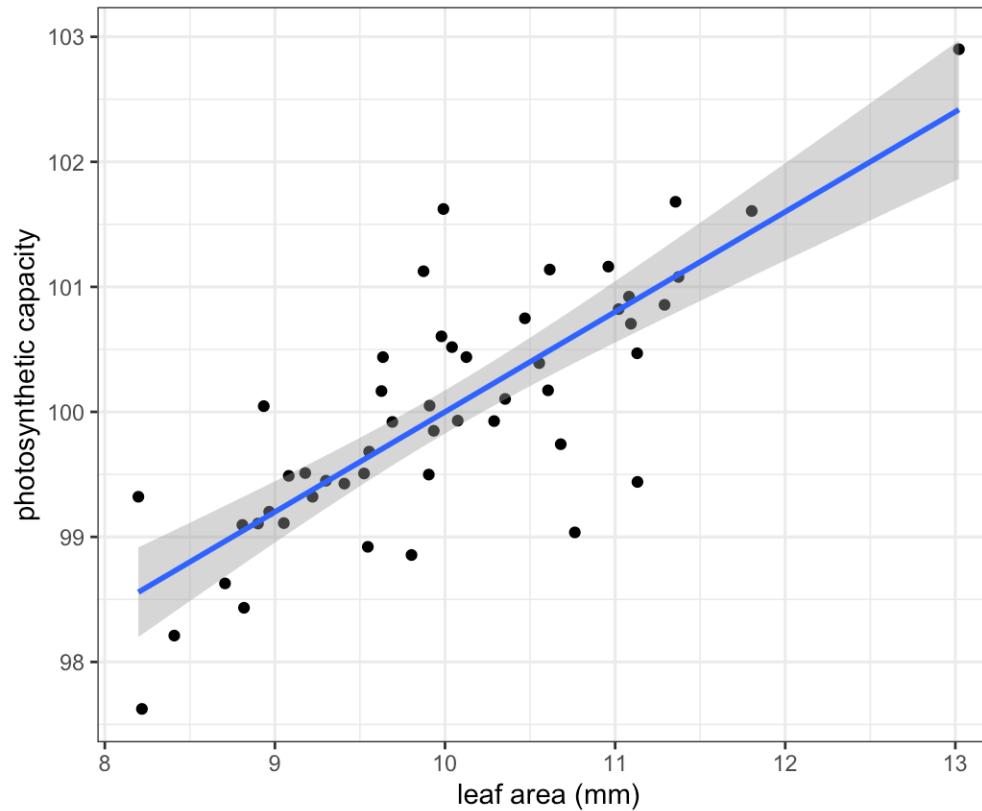
$$\text{OR} = \frac{29/21}{8/42} = 7.25$$

Interpretation:

- $0 \leq \text{OR} \geq \infty$
- $\text{OR} > 1$ : greater odds of B occurring in the presence of A.
- $\text{OR} = 1$ : equal odds (i.e. no association). Odds of A occurring are unrelated to odds of B occurring.
- $\text{OR} < 1$ : reduced odds of B occurring in the presence of A.

# The Correlation Coefficient

For quantifying the **continuous association** between groups.



# The Correlation Coefficient

## Pearson's $r$

It's the *covariance* between two variables, normalised by their standard deviation(s)

$$r = \frac{\text{COV}(X, Y)}{\sigma_X \sigma_Y} = \frac{0.8}{1*1} = 0.8$$

Interpretation:

- $-1 \leq r \geq 1$
- $r = 1$ : strong positive association between A and B. As A increases, B increases.
- $r = 0$ : no association between A and B.
- $r = -1$ : strong negative association between A and B. As A increases, B decreases.

# Extracting effects



# Extracting effects

From best to worst-case:

- Raw data are provided
  - Calculate directly yourself
- Rich summary data are provided or can be estimated (e.g. means, errors, sample sizes)
  - Calculate directly yourself
- Raw data are plotted
  - Snatch the data from the plot(s), then calculate yourself
- Effects are reported
  - Look at them, write them down
- Summary statistics are provided (e.g. test statistics, degrees of freedom)
  - Impute and/or approximate from available information
- Some haphazard combination of the above
  - Do your best, success not guaranteed
- Nothing is provided, everything is awful
  - Email the authors, probably be ignored

# Calculating and converting among effects

in R

- `metafor::escalc()`: Calculate various effect sizes or outcome measures (and the corresponding sampling variances) that are commonly used in meta-analyses. When you have rich **summary data**.
- Package `compute.es`: For calculating the most widely used effect sizes (ES), along with their variances, confidence intervals and p-values. When you need to **convert among effects** or impute from **test statistics**.

# Calculating effects

Example: from data to  $g$

....the addition of nitrogen-rich fertilizer had a strong positive effect on yield (control =  $242 \pm 9.73$ , treat =  $259 \pm 9.72$  kg/ha).

```
metafor::escalc('SMD', m1i = 259, sd1i = 9.73, n1i = 65,
                 m2i = 242, sd2i = 9.72, n2i = 65)
```

```
##  
##      yi      vi  
## 1 1.7378 0.0424
```

# Converting among effects

Example: from  $r$  to  $g$

...across the surveyed farms, we found a strong correlation between nitrogen content and crop yield ( $r = 0.62$ ,  $n = 33$ ).

# Converting among effects

```
compute.es::res(r = 0.62, n = 33)
```

```
## Mean Differences ES:
```

```
##  
## d [ 95 %CI] = 1.58 [ 0.7 , 2.46 ]  
## var(d) = 0.2  
## p-value(d) = 0  
## U3(d) = 94.3 %  
## CLES(d) = 86.81 %  
## Cliff's Delta = 0.74  
##
```

```
## Correlation ES:
```

```
##  
## r [ 95 %CI] = 0.62 [ 0.35 , 0.79 ]  
## var(r) = 0.01  
## p-value(r) = 0  
##
```

```
## z [ 95 %CI] = 0.73 [ 0.37 , 1.08 ]
```

```
## var(z) = 0.03  
## p-value(z) = 0  
##
```

```
## Odds Ratio ES:
```

```
##  
## OR [ 95 %CI] = 17.58 [ 3.54 , 87.23 ]
```

# Meta-analytic (statistical) models



# The story so far

We have:

1. Formulated a question and designed a search
2. Scoured and collated the existing literature
3. Extracted numerical outcome measures from every study
4. **Now what?**

# Now what?

Two goals:

1. Estimate overall effect
2. Assess and explore heterogeneity between studies



# Now what?

Two goals:

1. **Estimate overall effect**
2. Assess and explore heterogeneity between studies



# Meta-analytic models

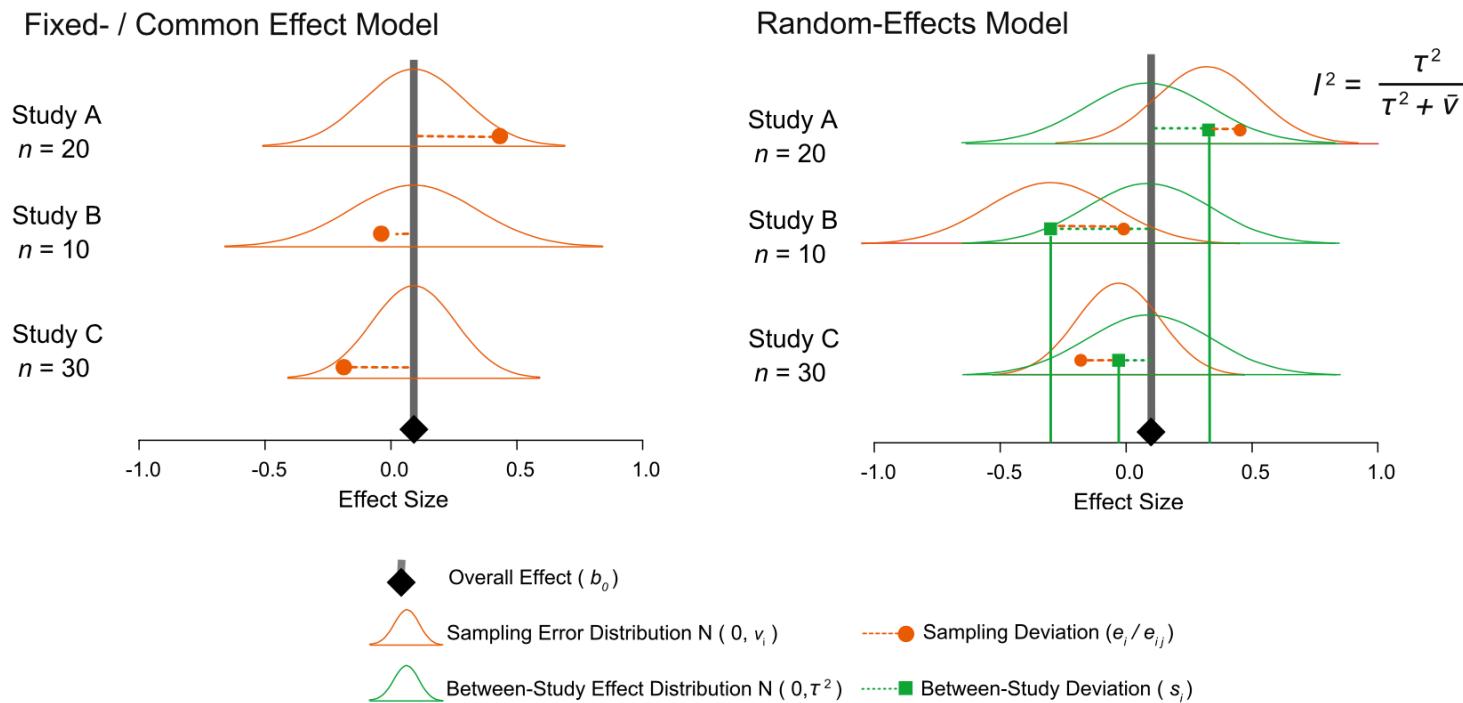
Two types of statistical model

- Fixed or **Common-effects** model
- **Random-effects** model

The decision is informed by the question and nature of the constituent studies

# Meta-analytic models

- The **common-effects** model assumes there is **one true effect** that underlies all the studies in the analysis. All differences in observed effects among individual studies are due to random sampling error alone.
- The **random-effects** model assumes the **true effect differs from study to study**. Differences in observed effects among studies are due to random sampling error, as well as between-study heterogeneity in true effects. You'll almost always use this.



# The 'metafor' package

## metafor: A Meta-Analysis Package for R

license [GPL](#) R-CMD-check [passing](#) codecov [63%](#) CRAN [2.4-0](#) devel [2.5-69](#) downloads [17K/month](#) downloads [520K](#)

- A comprehensive (or near enough) package for meta-analysis modelling in R
- Extensive, extremely helpful package documentation, as well as a website:  
<http://www.metafor-project.org/doku.php>
- Created by Wolfgang Viechtbauer

# The 'metafor' package

## Some key functions

- `escalc()`: For effect size calculation
- `rma()`: The workhorse for linear meta-analytic modelling (fixed/random/mixed-effects)
- `rma.mv()`: For multilevel linear modelling (fixed/random/mixed-effects)

# Some data

Are aposematic signals honest?



- Aposematic signals = ‘warning’ signals
- Prediction: signals are *honest* guides to strength of chemical defences
- Data: correlations (Pearson’s r) between measures of colour signal expression and toxicity (112 effects/22 studies)
- `meta_warning` in your datasets folder

# Running a random-effects model

Inspect the data

```
names(dat_warning)
```

```
## [1] "author"      "year"        "obs"         "study"       "group"       "study_type"   "class"       "subclass"
## [9] "order"        "suborder"     "family"      "genus"       "species"     "tox_measure" "col_var"     "n"
## [17] "r"
```

# Running a random-effects model

Inspect the data

```
dat_warning[, 16:17]
```

```
## # A tibble: 122 × 2
##       n      r
##   <dbl> <dbl>
## 1     39  0.19
## 2     36  0.22
## 3     36  0.12
## 4     36  0.38
## 5     37  0.03
## 6     37  0.22
## 7     39 -0.09
## 8     36  0.3
## 9     36  0.21
## 10    36 -0.14
## # i 112 more rows
```

# Running a random-effects model

Convert our correlation coefficient (Pearson's r) to the more suitable Fisher's z

```
dat_warning <- escalc(measure = 'ZCOR', ri = r, ni = n, data = dat_warning)
```

# Running a random-effects model

Have another look at the data

```
dat_warning[, 16:19]
```

```
##  
##      n      r      yi      vi  
## 1  39  0.19  0.1923 0.0278  
## 2  36  0.22  0.2237 0.0303  
## 3  36  0.12  0.1206 0.0303  
## 4  36  0.38  0.4001 0.0303  
## 5  37  0.03  0.0300 0.0294  
## 6  37  0.22  0.2237 0.0294  
## 7  39 -0.09 -0.0902 0.0278  
## 8  36  0.30  0.3095 0.0303  
## 9  36  0.21  0.2132 0.0303  
## 10 36 -0.14 -0.1409 0.0303  
## 11 37 -0.11 -0.1104 0.0294  
## 12 37 -0.19 -0.1923 0.0294  
## 13 30  0.41  0.4356 0.0370  
## 14 17 -0.03 -0.0300 0.0714  
## 15 25 -0.19 -0.1923 0.0455  
## 16 22  0.41  0.4356 0.0526  
## 17 27  0.61  0.7089 0.0417  
## 18 22  0.16  0.1614 0.0526
```

# Running a random-effects model

```
m_random <- rma(yi = yi, vi = vi, method = "REML", weighted = TRUE, data = dat_warning)
```

- `yi`: The vector containing our effect sizes
- `vi`: The vector containing the associated variance for each effect size
- `method`: REML for random-effects estimation (many options, but that's the default & a good choice)
- `weighted`: Should effect sizes be weighted during estimation? (Default is `TRUE`, but we'll be explicit).
- `data`: the dataset containing all this

# Interpreting a random-effects model

Take a look at the results

```
m_random
```

```
##  
## Random-Effects Model (k = 122; tau^2 estimator: REML)  
##  
## tau^2 (estimated amount of total heterogeneity): 0.0803 (SE = 0.0191)  
## tau (square root of estimated tau^2 value): 0.2834  
## I^2 (total heterogeneity / total variability): 58.26%  
## H^2 (total variability / sampling variability): 2.40  
##  
## Test for Heterogeneity:  
## Q(df = 121) = 280.1755, p-val < .0001  
##  
## Model Results:  
##  
## estimate se zval pval ci.lb ci.ub  
## 0.2414 0.0361 6.6852 <.0001 0.1706 0.3122 ***  
##  
## ---  
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

N.B. This is still Fisher's z, but we could transform back to r using `metafor::transf.ztor()`

# Interpreting a common-effects model

Are aposematic signals honest?

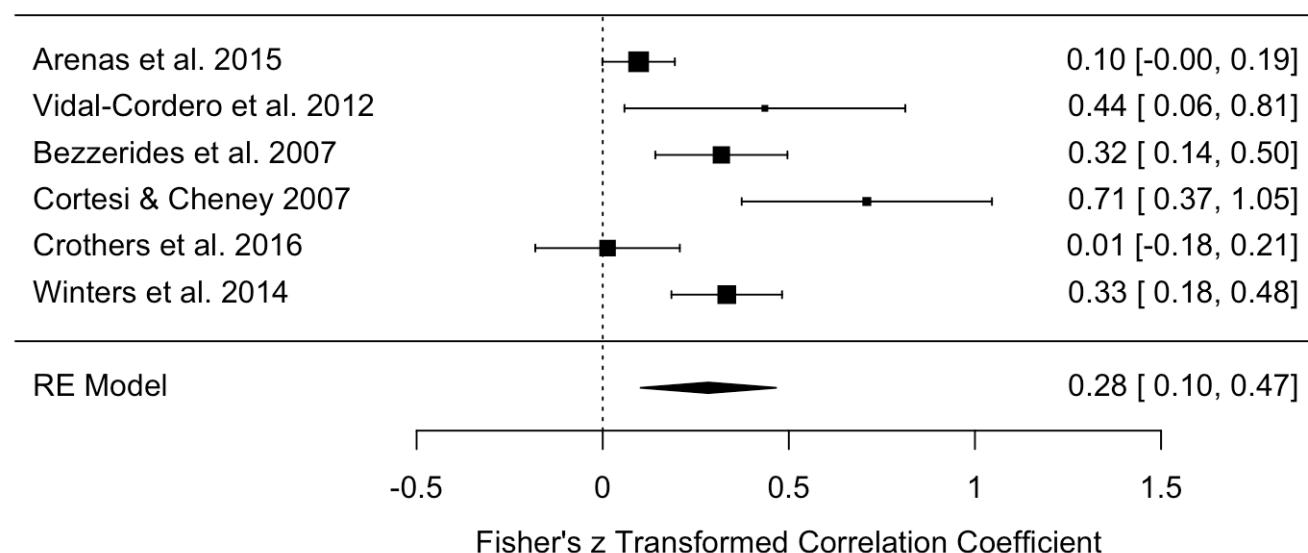


Mean overall correlation between colour signal expression & toxicity under a random-effects model is  **$z = 0.241 \pm 0.037 [0.171, 0.312]$** .

# Visualising your model: the forest plot

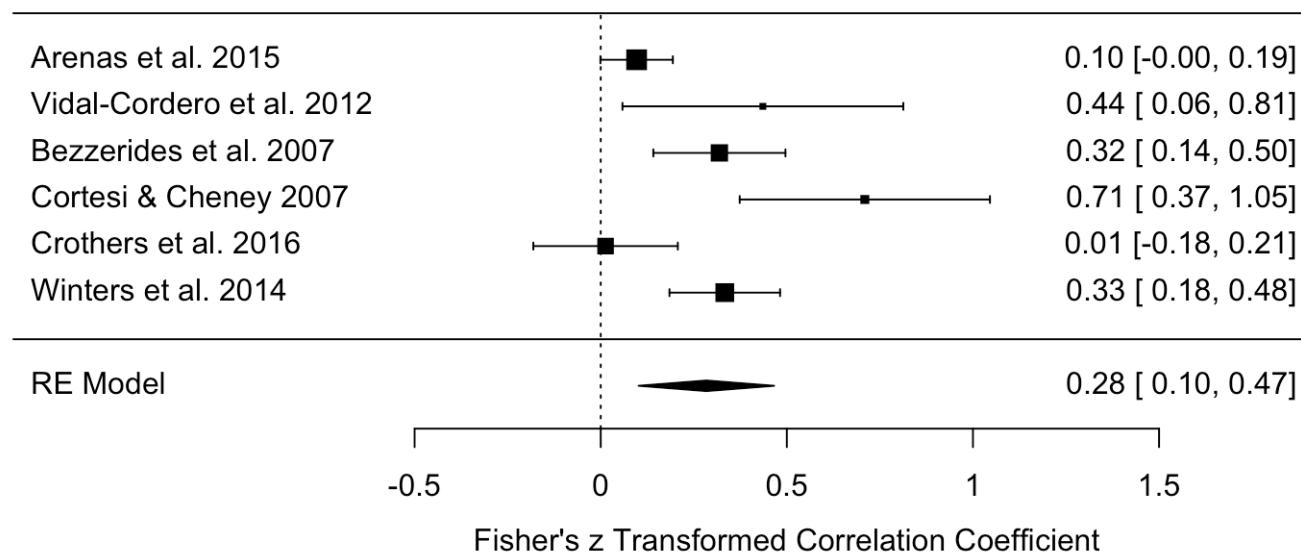
A **forest plot** displays the effect size estimates of individual contributing studies, as well as the overall mean  $\pm$  variance. Can use the `forest()` function from `metafor` for a quick (and fairly ugly) plot.

```
# For a subset only  
forest(m_subset, slab = paste(dat_subset$author, dat_subset$year))
```



# Visualising your model: the forest plot

A **forest plot** displays the effect size estimates of individual contributing studies, as well as the overall mean  $\pm$  variance. Can use the `forest()` function from `metafor` for a quick (and fairly ugly) plot.



# Quantifying heterogeneity



# Now what?

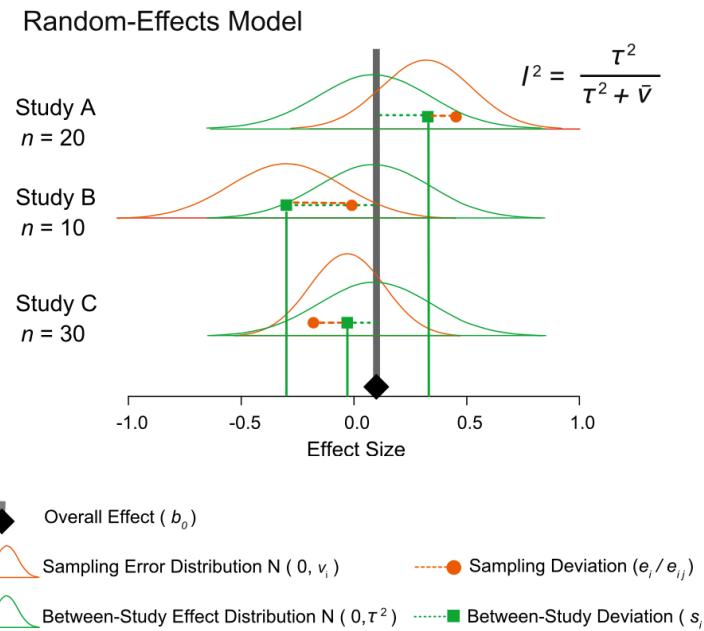
Two goals:

1. Estimate overall effect
2. **Assess and explore heterogeneity between studies**



# What is heterogeneity?

*Heterogeneity is variation in effect sizes, beyond sampling error*



# Why is heterogeneity interesting?

Eg: We synthesise 32 experimental studies examining whether the presence of herbivores can help to prevent the establishment of newly-invasive plants. We find a moderate reduction in the survival of exotic plants in the presence of herbivores ( $SMD = 0.36 \pm 0.08$ ), and again high heterogeneity ( $I^2 = 86\%$ ).

- Do herbivores help?
- Are the effects density dependent?
- Are the differences among specialist & generalist herbivores?
- Are there differences among herbivore taxa (e.g. vertebrate vs invertebrate)?



# The point is

Often heterogeneity is **equally** or **more interesting** than overall mean effects.

Especially in ecology & evolution & ag, where it can be leveraged to test basic and applied theory via *a priori* predictions.



# Quantifying heterogeneity

## Common statistics (and rough definitions)

$\tau$ : The standard deviation of underlying effects across studies. Units = effect size.

$\tau^2$ : The variance of underlying effects across studies. Units = effect size.

$I^2$  : **The amount of variation in effect sizes that cannot be explained by sampling error.**  
**Units = 0 - 100%.**

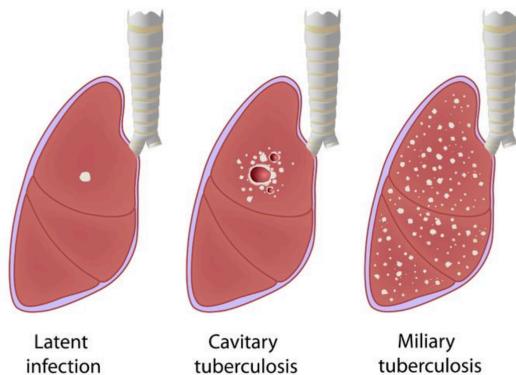
95% CI: The *precision* of the estimated mean effect (confidence interval).

95% PI: The *dispersion* of effects around the mean (prediction interval).

# Quantifying heterogeneity

Colditz et al. (1994) *J. Am. Med. Assoc*

- **Aim:** Examine the overall effectiveness of the BCG vaccine for preventing tuberculosis and to examine moderators that may potentially influence the size of the effect.
- **Summary:** Contains the results from 13 studies examining the effectiveness of the Bacillus Calmette-Guerin (BCG) vaccine against tuberculosis, conducted using different methods of treatment allocation, at different times, and across different study populations & locations.



# Quantifying heterogeneity

```
# Load our library
library(metafor)

# Load data on bcg and calculate effect size (odds ratio)
bcg_data <- escalc(measure = "OR", ai = tpos, bi = tneg, ci = cpos, di = cneg, data = dat.bcg)

# Run a random-effects model
m_bcg <- rma(yi, vi, method = 'REML', data = bcg_data)

# And inspect the result
m_bcg
```

# Quantifying heterogeneity

```
## tau^2 (estimated amount of total heterogeneity): 0.3378 (SE = 0.1784)
## tau (square root of estimated tau^2 value):      0.5812
## I^2 (total heterogeneity / total variability):   92.07%
## H^2 (total variability / sampling variability): 12.61
##
## Test for Heterogeneity:
## Q(df = 12) = 163.1649, p-val < .0001
##
## Model Results:
##
## estimate      se     zval    pval    ci.lb    ci.ub
## -0.7452  0.1860  -4.0057  <.0001  -1.1098  -0.3806  ***
## 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

What is the overall mean effect of the BCG vaccine on tuberculosis prevalence? How heterogeneous are the effects between studies? (Where is our prediction interval?).

# Quantifying heterogeneity

```
# Confidence and prediction intervals
```

```
predict(m_bcg)
```

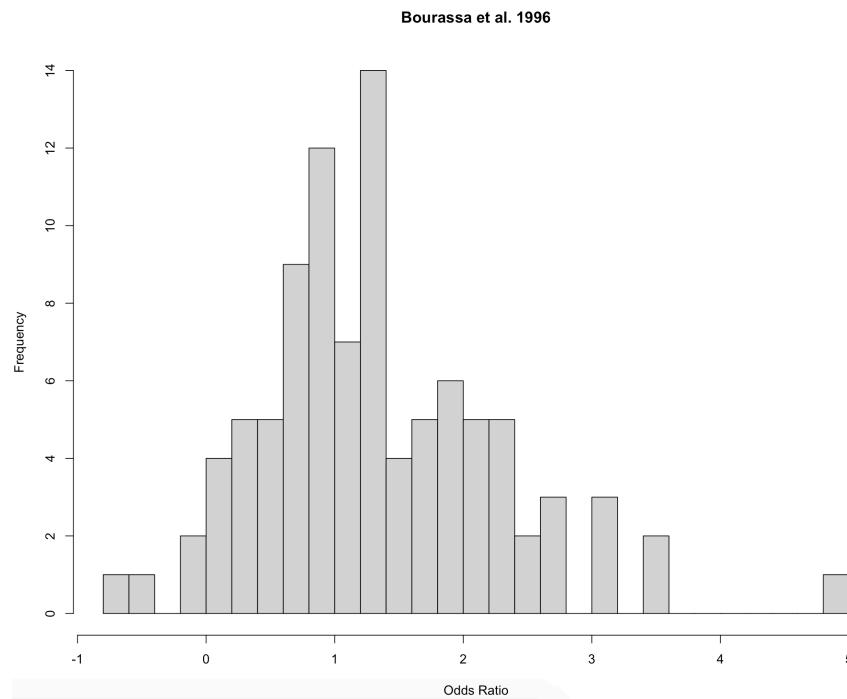
```
##
```

```
##     pred      se    ci.lb    ci.ub    pi.lb    pi.ub
```

```
## -0.7452 0.1860 -1.1098 -0.3806 -1.9412 0.4508
```

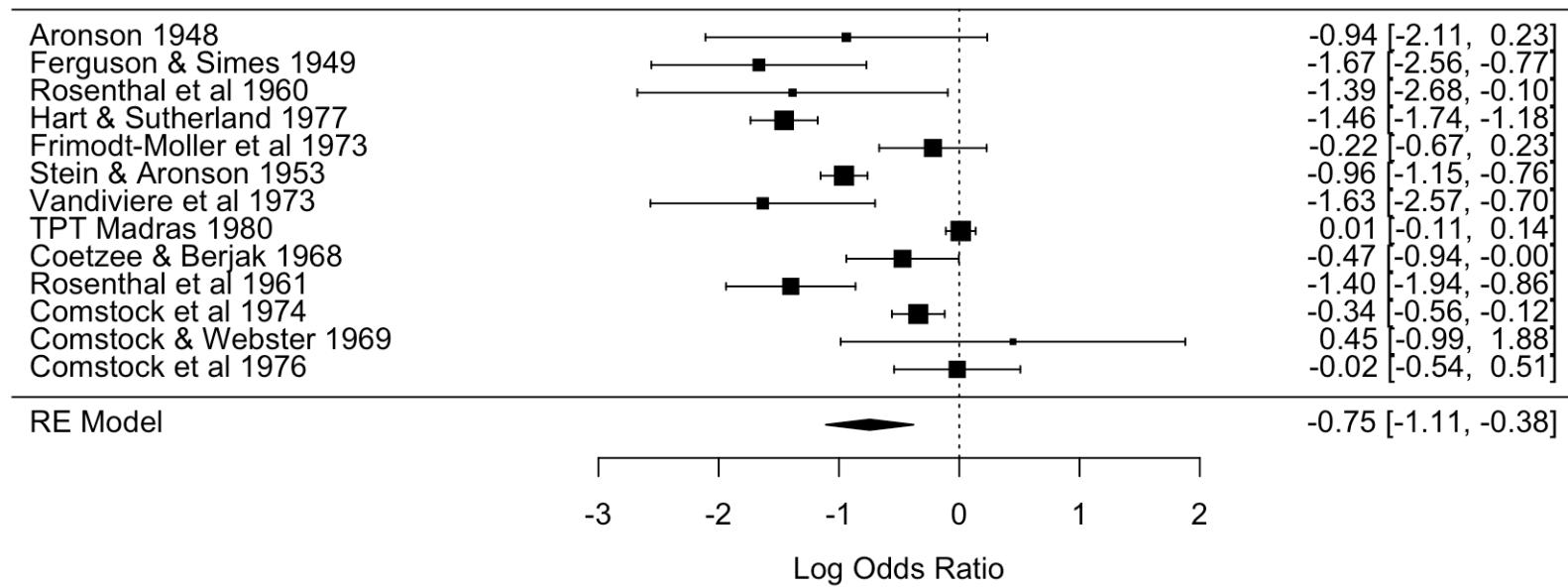
# Meta-regression

- We have our effects
- We know they vary
- We want to know the **source(s)** of this variation
- We have hypotheses as to the causes of such variation
- How do we test them?



# Meta-regression in action

Forest plot for BCG data



# Meta-regression in action

Given all that heterogeneity, we may like to consider the influence of **moderators** via **meta-regression**, by extending our random-effects model to become a mixed-effects model. As part of the dataset, the authors also recorded the following information about each study:

- `ablat`: absolute latitude of the study location (in degrees)
- `alloc`: method of treatment allocation (random, alternate, or systematic assignment)
- `year`: publication year

Hypotheses?

# Meta-regression in action

Lets test our hypothesised moderator using a **mixed-effects model**. In `metafor`, we can specify this via:

```
bcg_me <- rma(yi, vi, mods = ~ ablat - 1, data = bcg_data)
```

And check out the results:

# Meta-regression in action

Mixed-Effects Model (k = 13; tau<sup>2</sup> estimator: REML)

tau<sup>2</sup> (estimated amount of residual heterogeneity): 0.0504 (SE = 0.0449)  
tau (square root of estimated tau<sup>2</sup> value): 0.2246  
I<sup>2</sup> (residual heterogeneity / unaccounted variability): 57.39%  
H<sup>2</sup> (unaccounted variability / sampling variability): 2.35  
R<sup>2</sup> (amount of heterogeneity accounted for): 85.06%

Test for Residual Heterogeneity:

QE(df = 11) = 25.0954, p-val = 0.0088

Test of Moderators (coefficient 2):

QM(df = 1) = 25.2424, p-val < .0001

Model Results:

	estimate	se	zval	pval	ci.lb	ci.ub
intrcpt	0.3010	0.2146	1.4025	0.1608	-0.1197	0.7217
ablat	-0.0315	0.0063	-5.0242	<.0001	-0.0438	-0.0192

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

- R<sup>2</sup>: Model ‘fit’ ~85%.
- intrcpt: LOR is ~0 when ablat = 0 (i.e. treatment less effective/ineffective at the equator)
- ablat: Moderate, negative effect of ablat on LOR estimates. Estimated (average) log odds ratio becomes increasingly negative (i.e. treatment effect is *greater*) for study sites further from the equator.

# Meta-regression in action

- Is the Bacillus Calmette-Guerin (BCG) vaccine effective in preventing tuberculosis infections?

**Yes**, the sum-total evidence to date (well, ~1994) suggests so.



- Is there variation in vaccine efficacy and, if so, what drives it?

**Yes**, A moderate amount of variation is explained by the latitude of the population to whom the vaccine is administered.

Thanks! But continue for 'Assessing publication bias' slides...



# Model checking & publication bias

Non-exhaustive, example sources of bias:

- Indexing/search bias: few search databases, improper terms
- Selection bias: selective, post-hoc inclusion of studies
- Detection bias: differences between groups within studies in how outcomes are determined
- Attrition bias: differences between groups in withdrawals from a study
- Reporting bias: differences between reported and unreported findings
- **Publication bias**: studies with positive and ‘significant’ results are more likely to be published (i.e. the file-drawer effect)

# Assessing publication bias: the funnel plot

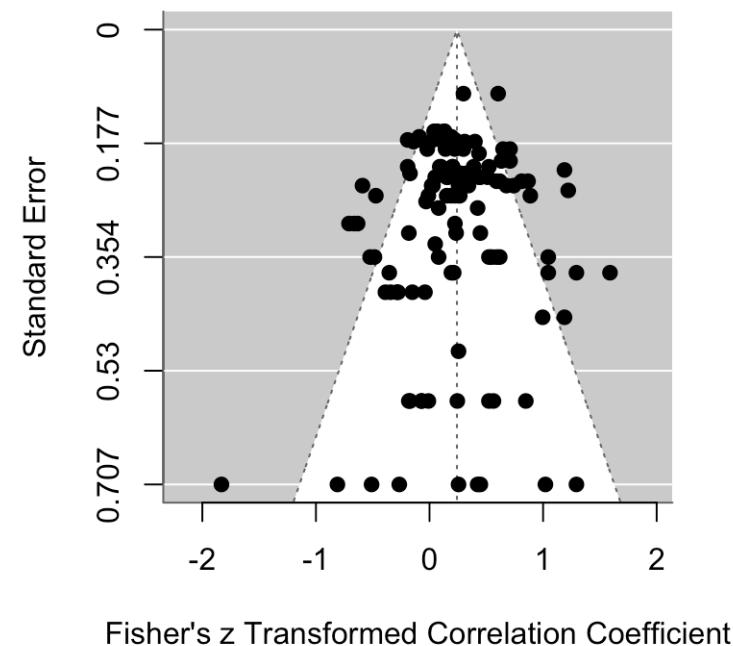
*“The small proportion of results chosen for publication are unrepresentative of scientists’ repeated samplings of the real world.”* - Young et al. *PLoS Medicine*

We can take steps (grey literature, design effective and broad searches), but ultimately can't change the underlying forces. But we have tools at our disposal to explore such possible effects.

# Assessing publication bias: the funnel plot

A **funnel plot** is a scatter plot of each study's effect size vs a measure of the study's **precision** (or size).

```
funnel(m_random)
```

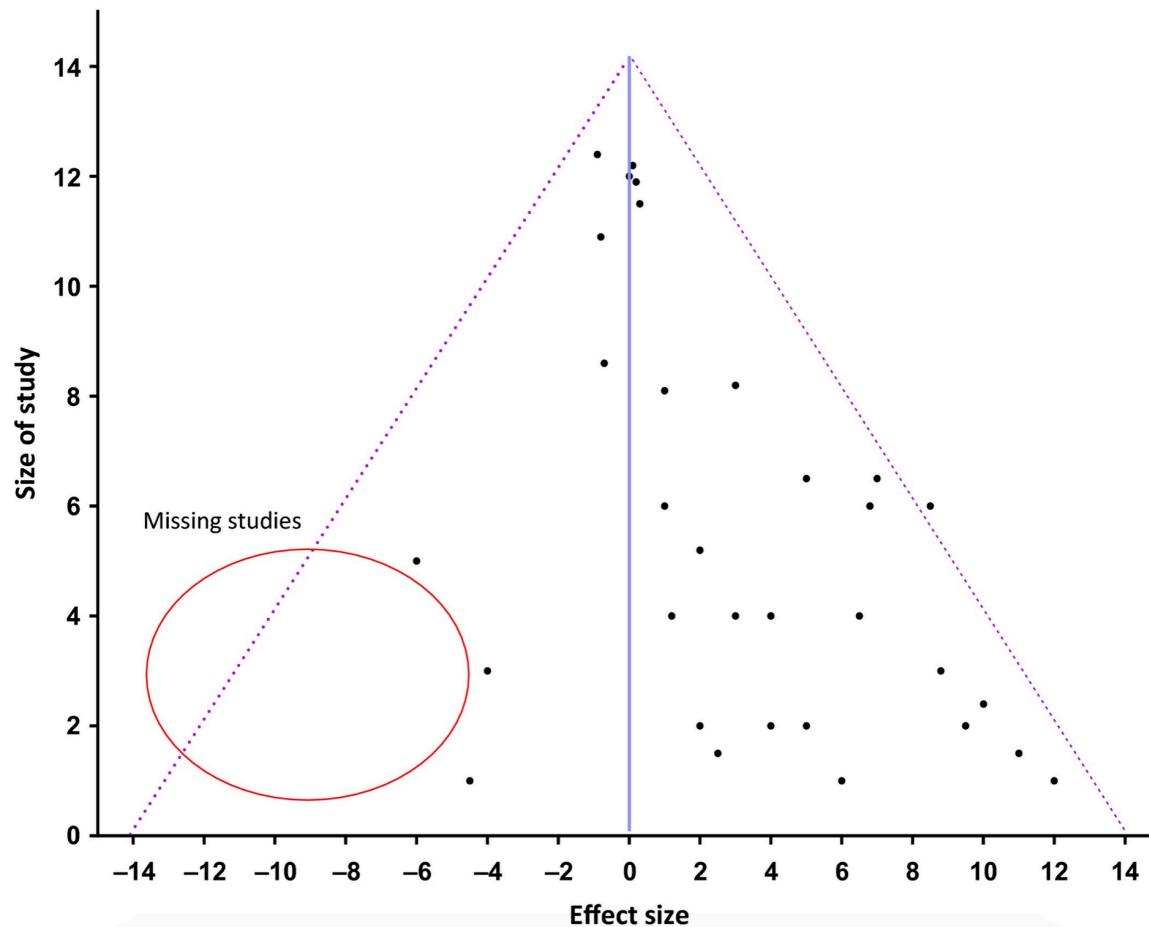


# Assessing publication bias: the funnel plot

The logic:

- Small, underpowered studies tend to report larger effect sizes compared with well-powered studies.
- A plot of effects vs precision should therefore form a ‘funnel’, with larger studies converging on the meta-analytic mean and smaller studies scattered symmetrically around it.
- **Asymmetry** *may* therefore indicate publication bias. If non-significant, imprecisely estimated effects are relegated to the file-drawer, then ‘holes’ will appear in the funnel plot.

# Assessing publication bias: the funnel plot



# Some notes of caution

- Funnel plots are not *definitive* tests of publication bias. There are many forms of publication bias, and several possible causes of funnel plot asymmetry.
- Hence, trim-and-fill analyses do not generate ‘corrected’ or more ‘valid’ estimates of the overall effect. They are a useful way of examining the sensitivity of the results to one particular form of publication bias.

Thanks!

