Meta-analysis

Fiona Callaghan, MA MS PhD,
Amgen Inc., South San Francisco, CA

Disclosure: Opinions/statements in this course reflect personal opinions, not policy or opinions of Amgen, Inc. or FDA.

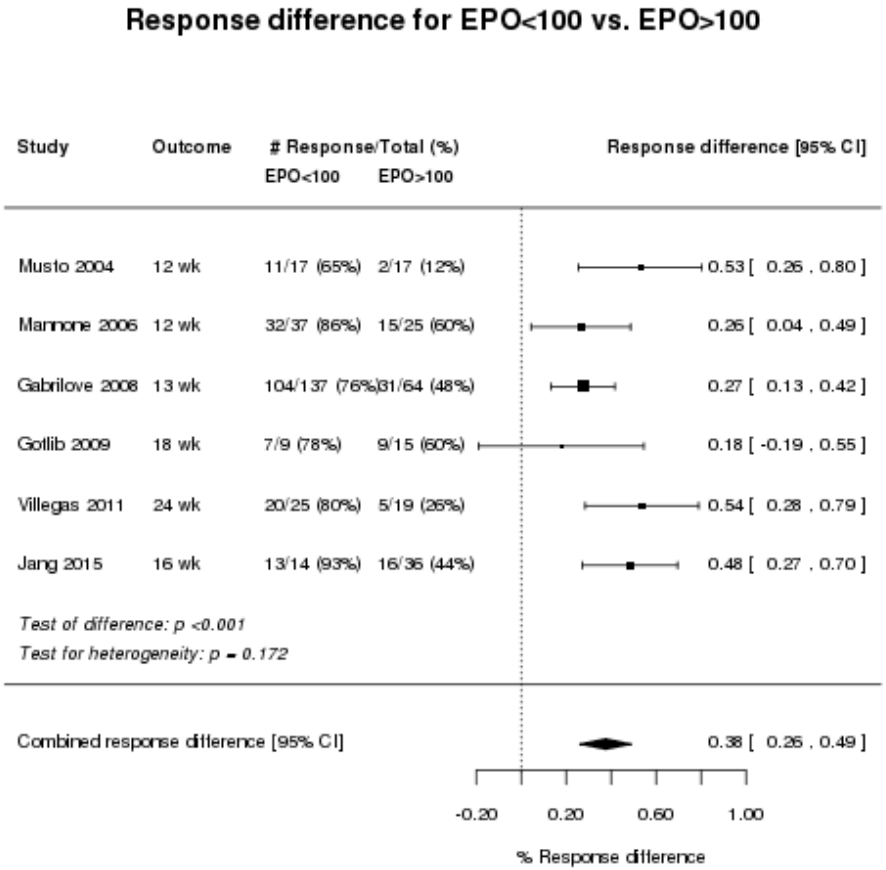
11 February 2016

Section 1

What is a meta-analysis?

What is meta-analysis?

- Combine the results from similar studies, stratified on study.



Why do a meta-analysis?

1. Many studies with conflicting results, but possibly converging on an answer, or
2. Many studies trending in the same direction but few are significant $\rightarrow$ combining the results may have more power, and generate a conclusion

Rofecoxib and myocardial infarction: randomized trials

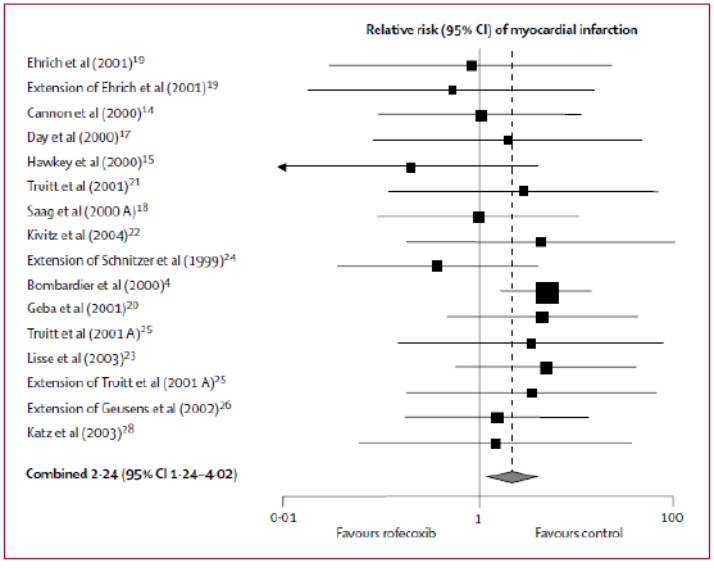


Figure 2: **Meta-analysis** of randomised trials comparing rofecoxib with control

Jüni et al. (2004)

How do you do a meta-analysis?

- Pre-specify (before collecting data or looking at your results):
  - The plan/protocol for the systematic review
  - Statistical analysis plan
  - Data abstraction form
- An iterative process: a change in one can affect the others
- Deciding which studies get in often decides the result.

Systematic review*

- Define the primary objective and any secondary objectives,
  - Is there an increased risk of X for exposed vs non-exposed in cohort Y?
  - Does this hold in the subgroups: gender, dose, disease severity etc.
- Document inclusion/exclusion criteria
- Document your search, e.g., PubMed terms, date of search etc. Make it reproducible.

*Adapted from presentation by C Poole

Systematic review cont'd*

- Create statistical analysis plan (SAP)/Data abstraction form
- Extract the study characteristics and results.
  - 2+ independent abstractors
- Which collections of studies are similar enough to compare?
  - Inclusion/ exclusion criteria?
  - Demographics?
  - Same outcome/exposure measurements?
- Contact investigators for missing information?
- Create the meta-analytic data set.

*Adapted from presentation by C Poole

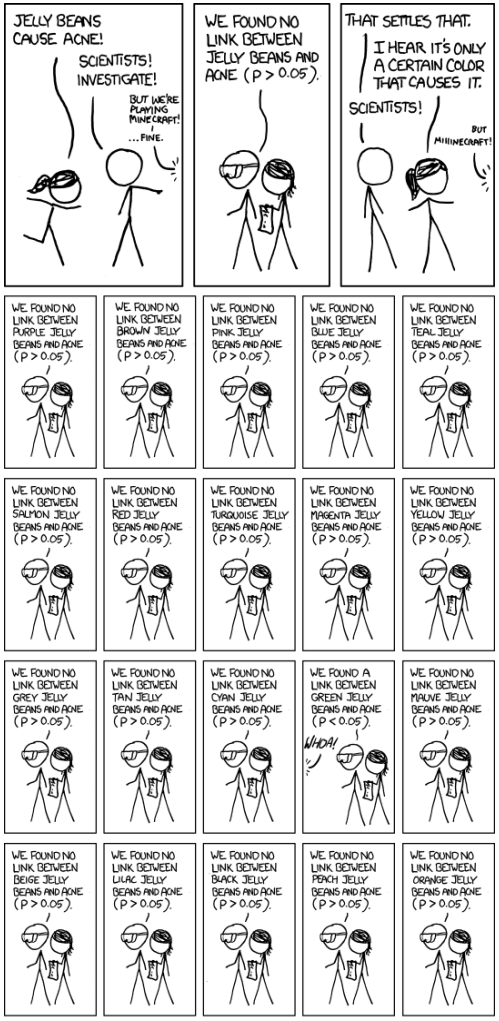
Statistical analysis plan SAP

- Primary objective: what you will win or lose on
- Secondary objectives
- What are you going to estimate, e.g. RR, OR, risk difference
- Inclusion/exclusion criteria
- Analytical methods, e.g., Fixed or random effects
- Check assumptions, e.g., homogeneity
- Sensitivity analyses
    - Alternative analyses that anticipate the critiques of your study

Sensitivity analyses

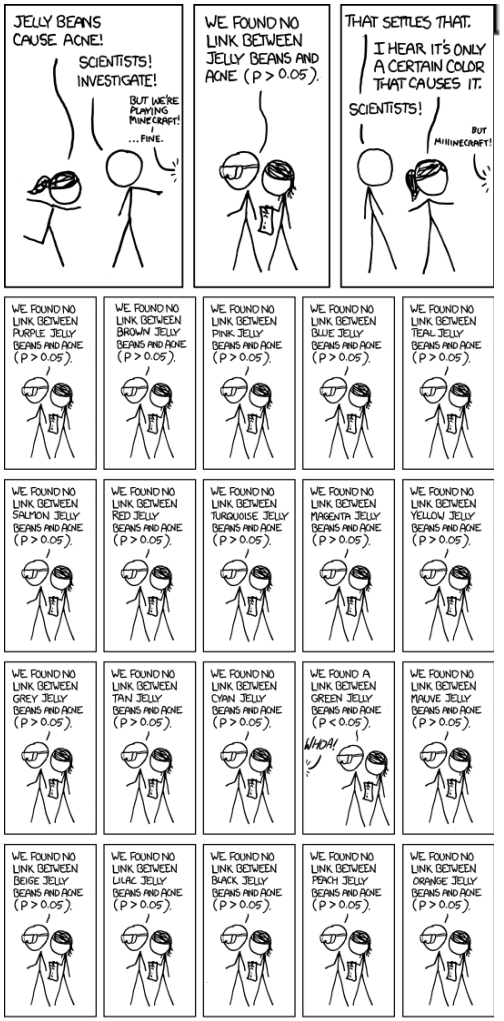
- Alternate analyses to answer questions such as "Would your results change if....?"
- Different analysis (fixed effects/random effects), outcomes, measures of the outcomes, inclusion/exclusion criteria...
- Tests how "robust" your conclusions are to your assumptions
- If all analyses point the same way, then you have a strong conclusion

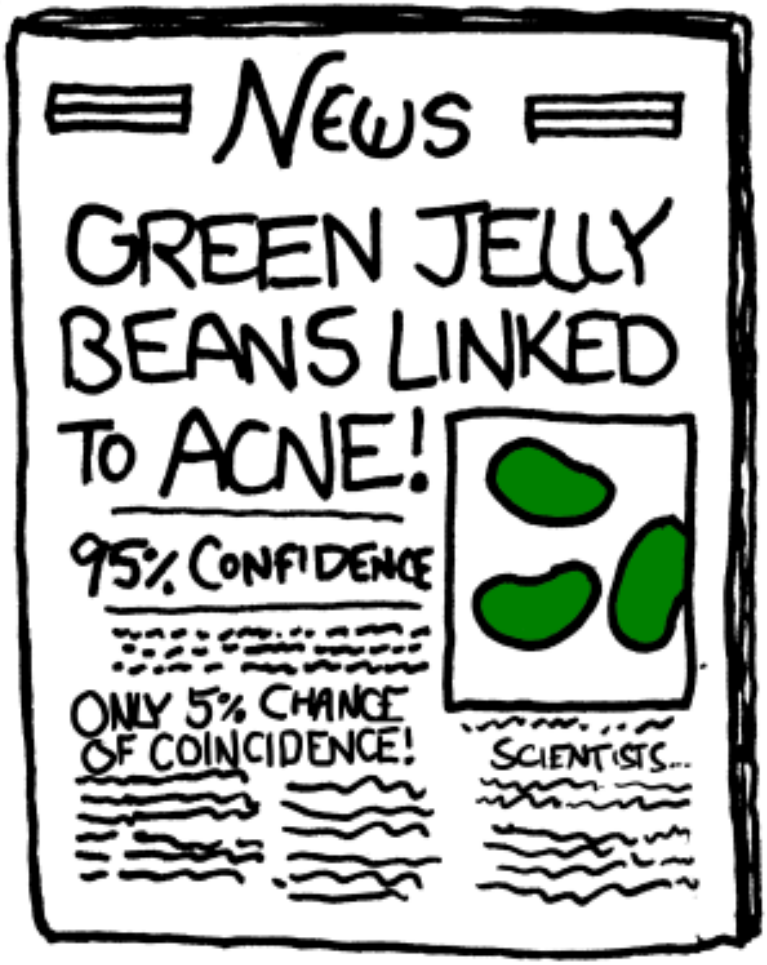
Pre-specified hypotheses vs. bait and switch



http://xkcd.com/882/

Pre-specified hypotheses vs. bait and switch





http://xkcd.com/882/

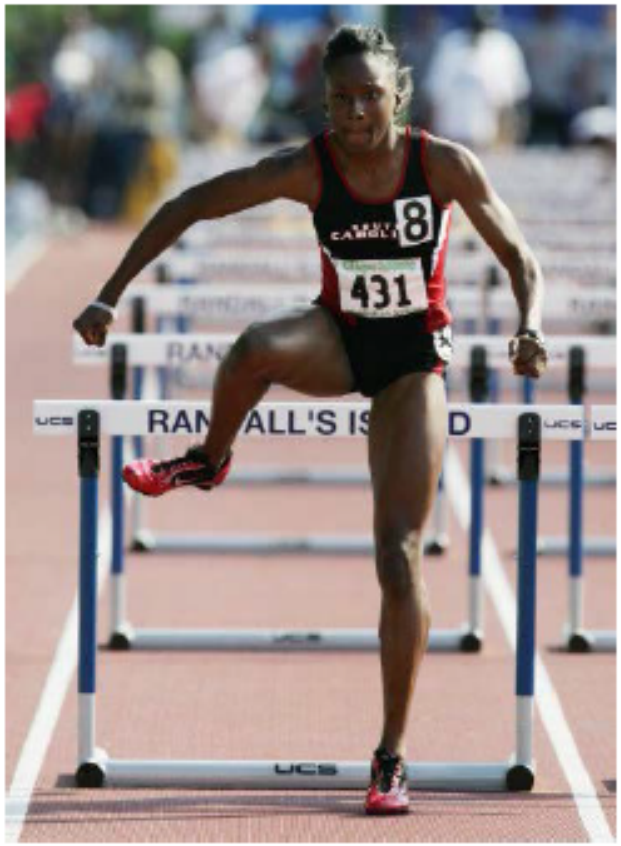
Four hurdles to meta-analysis*

1. Common quantitative footing
   - Can we say that the studies are estimating the same parameter?
2. Funnel plot symmetry
   - File-drawer bias check
3. Overall homogeneity
   - Do the estimates differ more than we would expect if only natural random variation were at play?
4. Study population and characteristics
   - Do the studies have different designs/methods, or different populations?

*Adapted from presentation by C Poole

1. The first hurdle: common quantitative footing

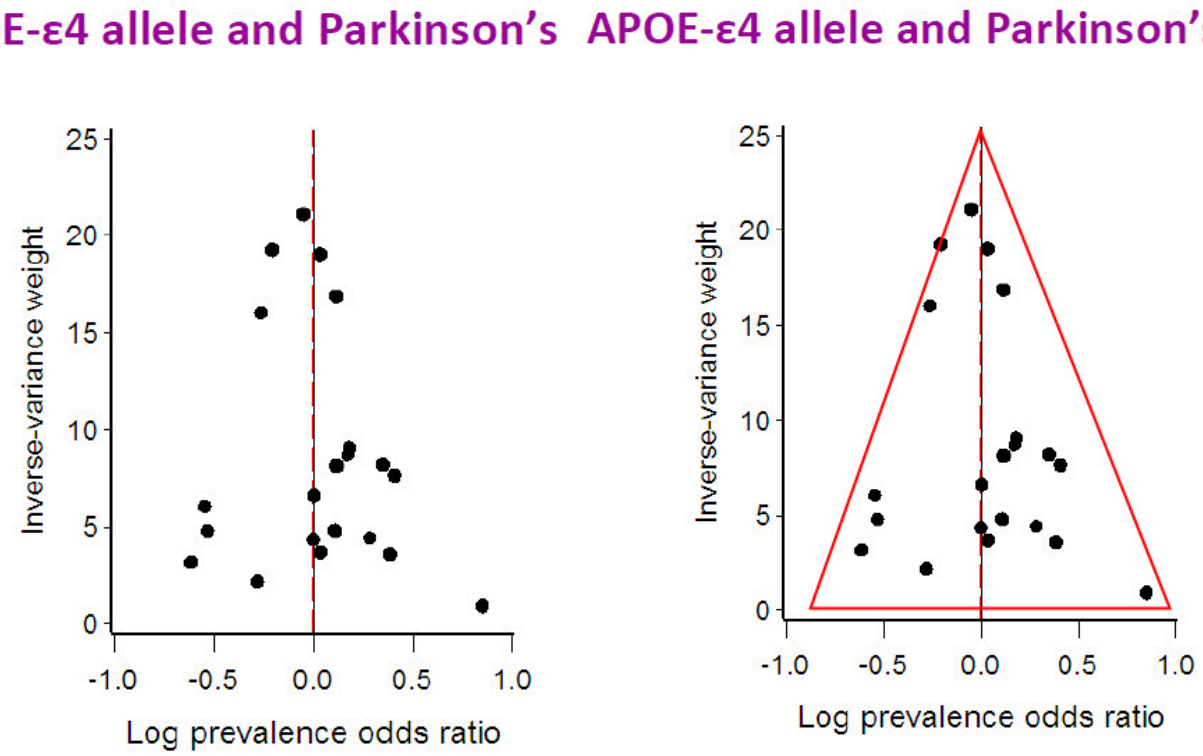
- To be meta-analyzable the studies must be estimating the same parameter
- Important to abstract raw count data if possible – maybe an alternative outcome can be calculated



2. The second hurdle: funnel plot symmetry

- Non-significant studies tend not to be published: "file-drawer" bias
- Small studies tend to have more variation and have more positive and negative results
- Typically, if the small studies with negative effect sizes are missing then bias might exist.
- A funnel plot is a plot of variation vs. effect size

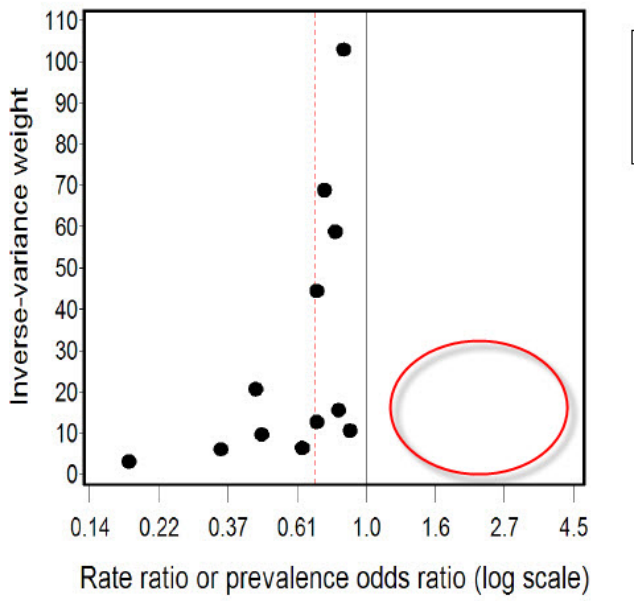
"Good" funnel plot*



- Shows symmetrical triangular, "funnel" shape
- From Huang et al 2004.

*Adapted from presentation by C Poole

"Bad" funnel plot*



- Small studies showing positive relationship between coffee and Parkinson's missing
- From Hernán et al 2004.

*Adapted from presentation by C Poole

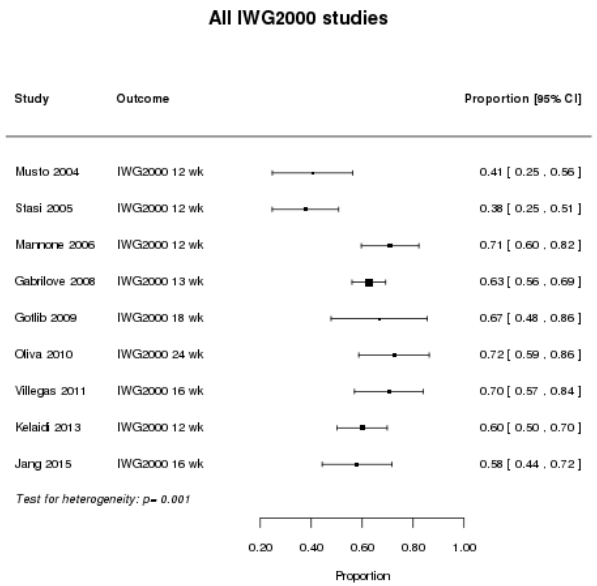
3. The third hurdle: Quantitative heterogeneity

- Inconsistent estimates: a single estimate is not a accurate representation of the truth

*Adapted from presentation by C Poole

3. The third hurdle: Quantitative heterogeneity

- % MDS patients responding to darbepoetin alfa treatment
- Results have wide variation, ranging from 40-70%



*Adapted from presentation by C Poole

Test of homogeneity

- Cochran's Q
- $Q = \sum_i \frac{(\hat{\mu}_i - \hat{\mu}_{\text{total}})^2}{\sigma_i^2} \sim \chi^2_{k-1}$
- k=number of studies
- Limitations:
  - If sample size is too small there is no power to detect variation. Can be seen incorrectly as a green flag to do analysis.
  - Hurdle to clear, but a low hurdle (necessary but not sufficient)
  - Plot is usually more informative
  - "Random effects" methodology does not solve the heterogeneity problem

4. The fourth hurdle: Study characteristics

- Combining different kinds of studies makes no sense
- In two studies with different patient populations the control will be different leading to different meanings of, say, the odds ratio.

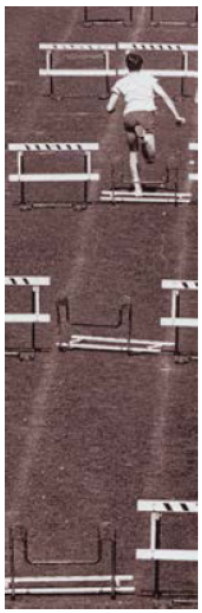
*Adapted from presentation by C Poole

Four hurdles on the track to meta-analysis aggregation

Common footing | **Funnel plot asymmetry** | **Overall heterogeneity** | **Study characteristics**

*Adapted from presentation by C Poole

Some studies do not clear any of the hurdles...



Study characteristics

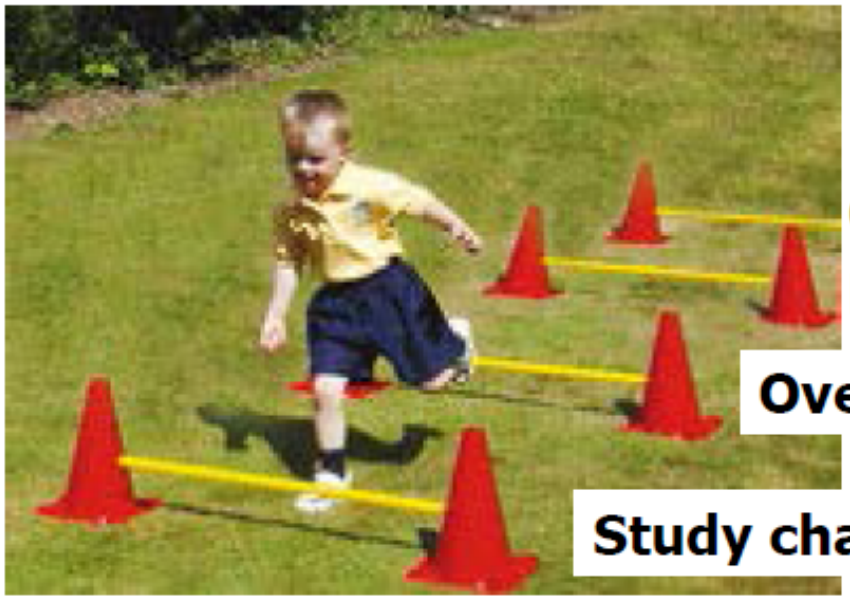
Overall heterogeneity

Funnel plot asymmetry

Common footing

*Adapted from presentation by C Poole

...and some meta-analyses set the bar too low



*Adapted from presentation by C Poole

Section 2

Statistical framework

Statistical framework

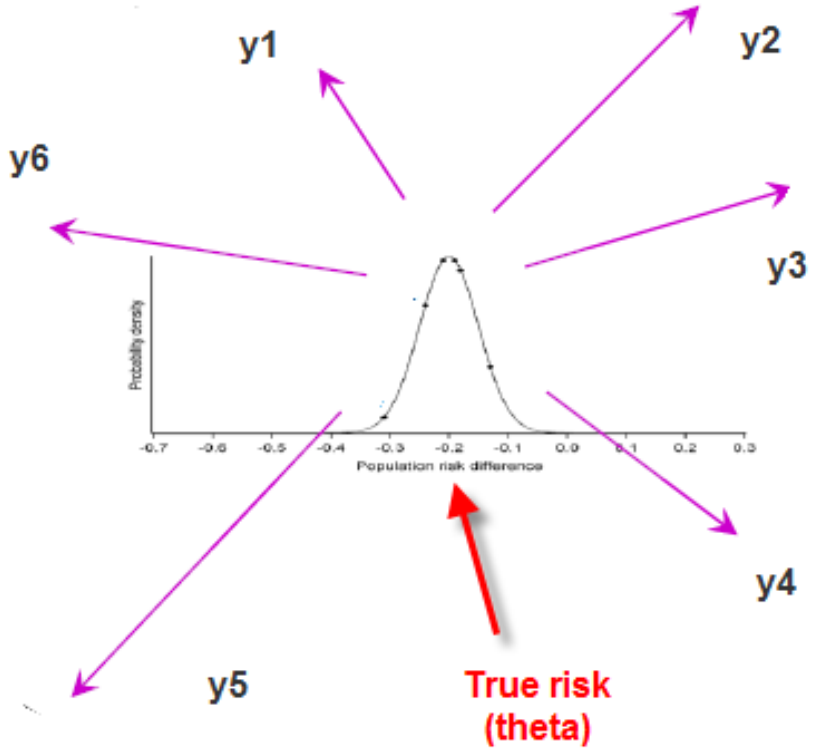
- What are we really doing when we combine the results from the studies?
- Each study generates an estimate y_i
  - e.g., proportion, risk difference, odds ratio
- Call the estimates $y_1, y_2, y_3, ...$
- We really want to know the true value of the RR, OR etc. Call it θ.
- How do the y_i's relate to θ?

What if we had patient-level data?

- Would we pool all patients?
- Still stratify on study.
- Why? Preserve randomized comparisons
- Can pool if you think studies are virtually identical (but strong assumption).

Fixed effects assumptions

- Assumption: Each study is trying to estimate the same θ, e.g., the true , fixed, underlying OR
  - We know the estimates from each study $y_1, y_2, y_3, ...$
  - We assume all the studies are unbiased estimators for θ, $E(y_1) = E(y_2) = ... = \theta$
- We take a mean of the y_is, after weighting each study



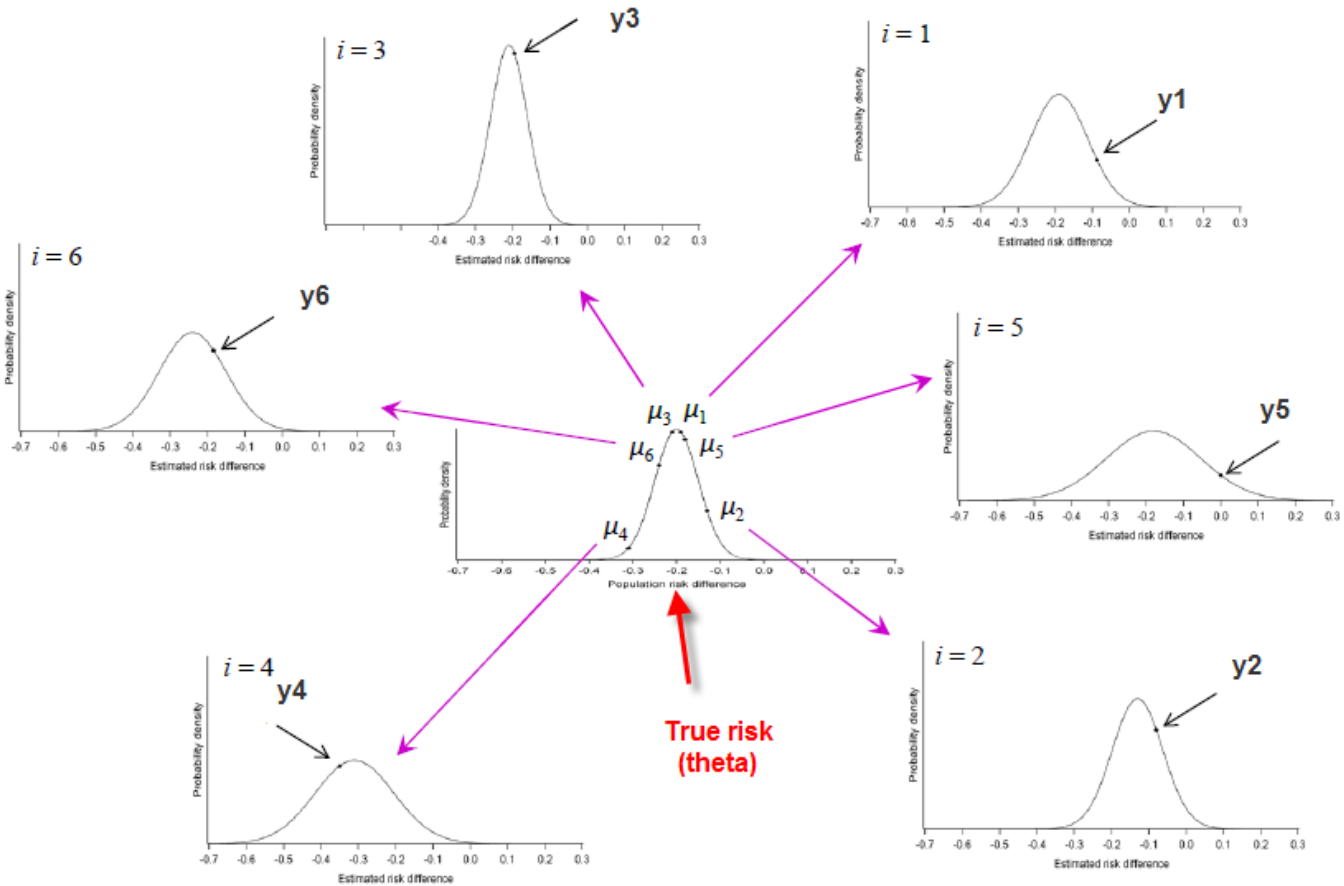
Implications: Fixed effects assumptions

- Assumption: the true difference between treatment and control θ is the same for every study
  - Are the studies really exactly the same, except for natural random variation?

Alternative: Random effects

- What if studies are not assumed to be estimating the same quantity?
- Assumption 1: Each study estimates a different true mean:
  - Study 1 estimates θ_1, study 2 estimates θ_2, etc.
  - What is the point? How do we find out the true value θ?
- Assumption 2: The θ_i's come from the same distribution centered around what we really want to know, θ
  - i.e, $\theta_i \sim N(\theta, \tau^2)$

Alternative: Random effects



Implications: Random effects (RE) assumptions

- Interpretation: The studies are drawn from a population of study designs/protocols, as well as simple random variation.
- Not assuming studies are the same, perhaps more realistic but...
  - If studies are different, what does combining them mean?
  - We still want to combine "apples with apples", i.e., there is still a homogeneity assumption
  - The studies need to be different in "degree" not "kind"
- In practice, RE method generally gives wider confidence intervals.

Section 3

Other topics

Meta-regression

- The goal is *not* "What is the true risk?". Instead the question is:
  - "Does the risk increase/decrease with variable X?" (continuous predictor)
  - "Do studies of Type A have a different risk estimate to studies of Type B?" (binary predictor)
- Perform regression on 1 study-level predictor

Meta-regression: Vaccination example

- There was conflicting evidence on whether the BCG vaccination for tuberculosis worked

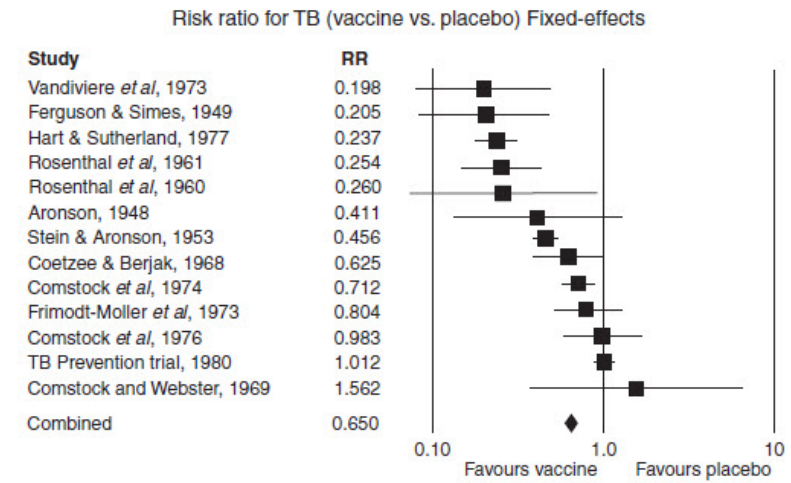


Figure 20.1 Fixed-effect model – forest plot for the BCG data.

Berkey CS, et al. *A random-effects regression model for meta-analysis.*Statistics in Medicine. 1995 Feb 28;14(4):395-411.

Borenstein M, et al. *Introduction to meta-analysis*, 2009, ISBN: 978-0-470-05724-7.

Meta-regression: Vaccination example

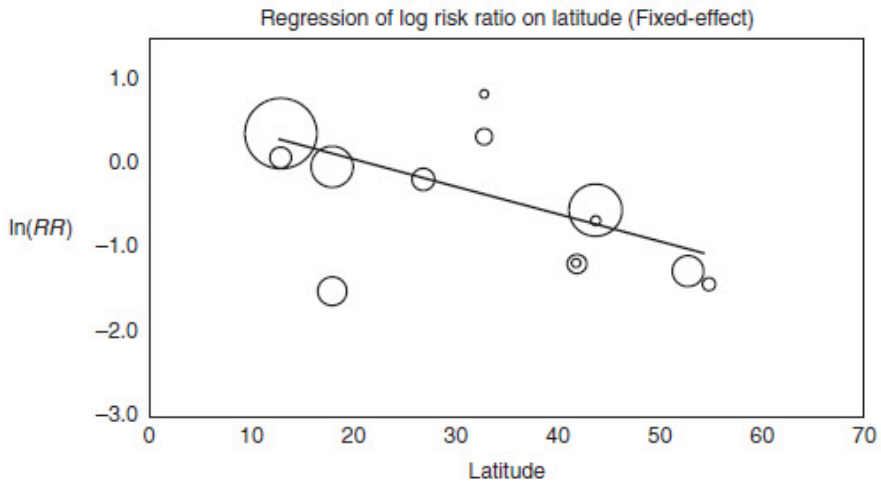


Figure 20.2 Fixed-effect model – regression of log risk ratio on latitude.

- The larger the circle the bigger the study
- With meta-regression found that vaccination worked for locations further away from the equator
- Not all locations had sufficient refrigeration facilities and vaccination was neutralized

Software

- All statistical software will perform meta-analysis, and get the same answers
- Some have more functionality than others
- R is free and has "metafor" package.

R demonstration

- Load the package, read in the data (an Excel file saved as a comma delimited file, .csv), and look at the data:

```
>library("metafor")
>mds<-read.csv(file="C:/Users/fcallagh/Documents/mds.csv",sep=",",header=T)
>mds
               Study     PrimaryResponse     N_reported       Response      Year
1          Musto 2005                  15             37  IWG2000 12 wk     2005
2          Stasi 2005                  20             53  IWG2000 12 wk     2005
3       Mannone 2006                   44             62  IWG2000 12 wk     2006
4     Gabrilove 2008                  129            206  IWG2000 13 wk     2008
5         Gotlib 2009                  16             24  IWG2000 18 wk     2009
6          Oliva 2010                  29             40  IWG2000 24 wk     2010
7  Nilsson-Ehle 2011                  20             36 PR/CR at 16 wk     2011
8      Villegas 2011                   31             44  IWG2000 16 wk     2011
9       Kelaidi 2013                   57             95  IWG2000 12 wk     2013
10         Jang 2015                  29             50  IWG2000 16 wk     2015
```

R demonstration

- Tell R what kind of data it is, e.g., Odds ratios , proportions,..
- R adds the appropriate estimates of risk (y_i) and variance (v_i) for each study

```
>dat <- escalc(measure="PR", xi=PrimaryResponse, ni=N_reported, data=mds)
>dat
             Study   PrimaryResponse   N_reported       Response   Year        yi          vi
1         Musto 2005                15           37  IWG2000 12 wk     2005 0.4054054 0.006514915
2         Stasi 2005                20           53  IWG2000 12 wk     2005 0.3773585 0.004433190
3       Mannone 2006                44           62  IWG2000 12 wk     2006 0.7096774 0.003323151
4     Gabrilove 2008               129          206  IWG2000 13 wk     2008 0.6262136 0.001136263
5        Gotlib 2009                16           24  IWG2000 18 wk     2009 0.6666667 0.009259259
6         Oliva 2010                29           40  IWG2000 24 wk     2010 0.7250000 0.004984375
7  Nilsson-Ehle 2011                20           36 PR/CR at 16 wk     2011 0.5555556 0.006858711
8      Villegas 2011                31           44  IWG2000 16 wk     2011 0.7045455 0.004730935
9       Kelaidi 2013                57           95  IWG2000 12 wk     2013 0.6000000 0.002526316
10         Jang 2015                29           50  IWG2000 16 wk     2014 0.5800000 0.004872000
```

R demonstration

- Tell R what kind of analysis, e.g., random effects (method="REML"), and print out results

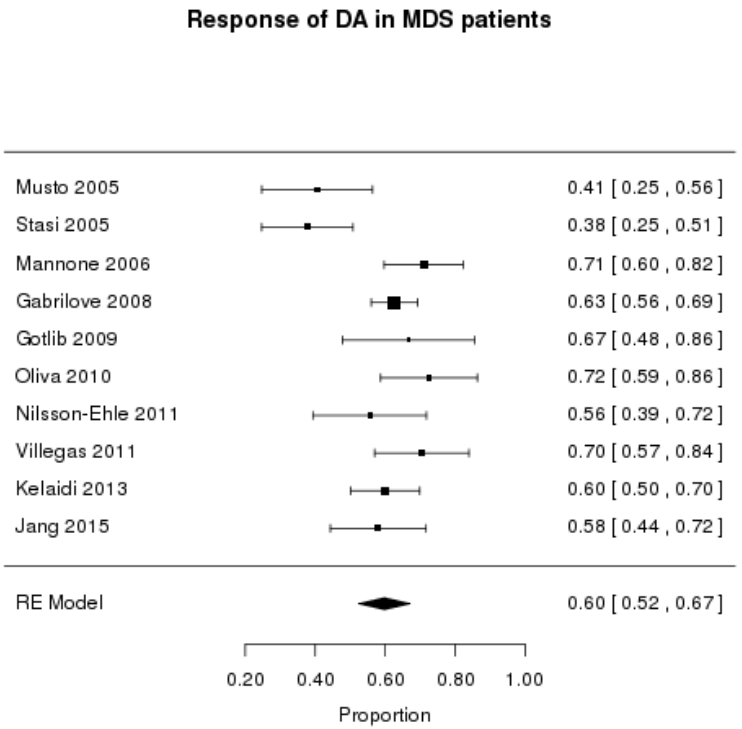
```
>res <- rma(yi, vi, method='REML',data=dat)
>res
Random-Effects Model (k = 10; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0093 (SE = 0.0065)
tau (square root of estimated tau^2 value):      0.0967
I^2 (total heterogeneity / total variability):   71.21%
H^2 (total variability / sampling variability):  3.47

Test for Heterogeneity:
Q(df = 9) = 27.3751, p-val = 0.0012

Model Results:

estimate       se     zval     pval    ci.lb    ci.ub
  0.5977   0.0372  16.0471   <.0001   0.5247   0.6707      ***

---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

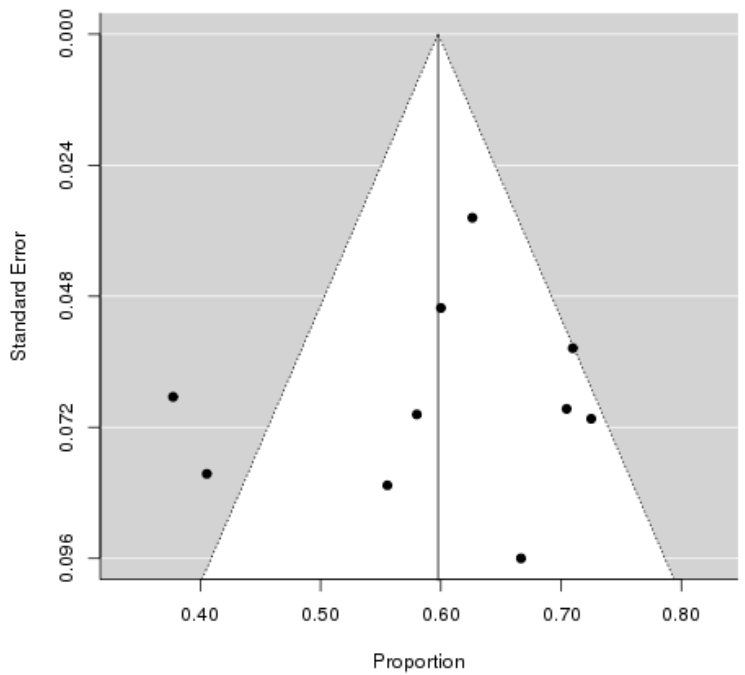
- Combined estimate is 59.77%, 95% CI (52.47%–67.07%)
- But, homogeneity is rejected (cannot combine), p=0.0012

R demonstration

- Create the forest plot* and funnel plot
- * For demonstration purposes, suppose we could combine the studies

```
>forest(res,slab=dat$Study, main="Response of DA in MDS patients", refline=NA)
>funnel(res)
```





Meta-analysis standards/ references

- Guidelines for meta-analyses and systematic reviews: review before starting:
  - Cochrane Collaboration Handbook and RevMan software: www.cochrane.org
  - Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) http://www.prisma-statement.org/
- References:
  - Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gtzsche PC, Ioannidis JPA, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA Statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: elaboration and explanation. Ann Intern Med 2009;151:W-65-W-94.
  - Greenland, S. Meta-analysis. Chapter 33 in: Rothman KJ, Greenland S, Lash TL. Modern epidemiology. Philadelphia: Lippincott Williams & Wilkins, 2008; 652–682.