Cochrane Methods Training Event 2016:
Statistical methods training for
statisticians supporting CRGs

Birmingham, UK, 17-18 March 2016

Estimating and Interpreting Heterogeneity and Summary Effects

Wolfgang Viechtbauer
Maastricht University
http://www.wvbauer.com

Topics

- outcome measures for meta-analysis
- review of the random-effects model
- estimation of and inference for μ
- estimation of and inference for τ^2
- inference about (the distribution of) θ_i
- measures of heterogeneity (τ^2, I^2, and H^2)

Outcome Measures for Meta-Analysis

- commonly used outcome measures:
  - raw or standardized mean difference
  - (log) ratio of means ('response ratio')
  - risk difference, (log) risk/odds ratio
  - correlation (raw or Fisher r-to-z transformed)
  - raw mean, (logit) proportion
  - ...

Observed vs. True Outcomes

- y_i = observed outcome in the ith study
- θ_i = true outcome in the ith study
- assumption: $E[y_i] = \theta_i$ (i.e., unbiasedness)
- bias adjustments may be necessary:
  - standardized mean difference (Hedges, 1981)
  - log risk/odds ratio (Haldane, Anscombe, Gart, …)
  - correlation coefficient (Olkin & Pratt, 1958)
  - …

Sampling Variance

- variability in the estimates if one were to repeat the study (repeatedly sample) under identical circumstances (with constant θ_i)
- equations for the sampling variance of the various outcome measures can be derived

Assumptions

- normal sampling distribution
  - often only true asymptotically
- known sampling variance
  - often based on an asymptotic approximation
  - often we need to substitute observed values into the equation, so we really only get an estimate
- sampling variance is independent of θ_i

Example: Standardized Mean Difference

- standardized mean difference:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{SD_p} \text{ is an estimate of } \theta = \frac{\mu_1 - \mu_2}{\sigma}$$

- bias correction:

$$g = \left[1 - \frac{3}{4(n_1 + n_2) - 9}\right] d$$

- asymptotic sampling variance:

$$v = \frac{1}{n_1} + \frac{1}{n_2} + \frac{\theta^2}{2(n_1 + n_2)}$$

8

Example: Standardized Mean Difference

- standardized mean difference:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{SD_p} \text{ is an estimate of } \theta = \frac{\mu_1 - \mu_2}{\sigma}$$

- bias correction:

$$g = \left[1 - \frac{3}{4(n_1 + n_2) - 9}\right] d$$

- estimated sampling variance:

$$v = \frac{1}{n_1} + \frac{1}{n_2} + \frac{g^2}{2(n_1 + n_2)}$$

9

Example: Log Odds Ratio

- log odds ratio:

$$y = \ln\left[\frac{a/b}{c/d}\right] \text{ is an estimate of } \theta = \ln\left[\frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)}\right]$$

- bias correction:

$$y = \ln\left[\frac{(a+\frac{1}{2})/(b+\frac{1}{2})}{(c+\frac{1}{2})/(d+\frac{1}{2})}\right]$$

- asymptotic sampling variance:

$$v = \frac{1}{\pi_1 n_1} + \frac{1}{(1-\pi_1)n_1} + \frac{1}{\pi_2 n_2} + \frac{1}{(1-\pi_2)n_2}$$

10

Example: Log Odds Ratio

- log odds ratio:

$$y = \ln\left[\frac{a/b}{c/d}\right] \text{ is an estimate of } \theta = \ln\left[\frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)}\right]$$

- bias correction:

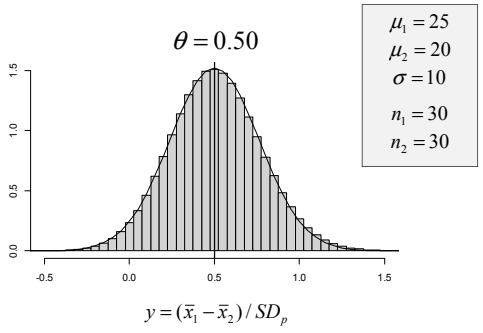
$$y = \ln\left[\frac{(a+\frac{1}{2})/(b+\frac{1}{2})}{(c+\frac{1}{2})/(d+\frac{1}{2})}\right]$$

- estimated sampling variance:

$$v = \frac{1}{a+\frac{1}{2}} + \frac{1}{b+\frac{1}{2}} + \frac{1}{c+\frac{1}{2}} + \frac{1}{d+\frac{1}{2}}$$

11

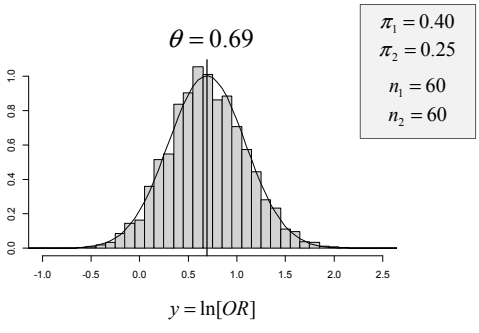
Example: Standardized Mean Difference



$y = (\bar{x}_1 - \bar{x}_2) / SD_p$

12

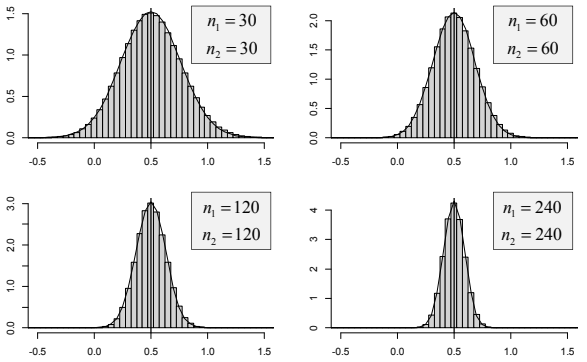
Example: Log Odds Ratio



$y = \ln[OR]$

13

Example: Standardized Mean Difference



14

Meta-Analytic Data

- $i = 1, \ldots, k$ studies
- have y_i and corresponding v_i
- in essence, we assume:

$$y_i \mid \theta_i \sim N(\theta_i, v_i)$$

- and independence of the estimates
- approx. 95% CI for θ_i: $y_i \pm 1.96\sqrt{v_i}$

15

Example: BCG Vaccine

- BCG: Bacillus Calmette-Guérin (BCG)
- BCG is a vaccine against tuberculosis (TB)
- effectiveness study: compare proportion of TB positive cases in a vaccinated and a non-vaccinated group

Camille Guérin

Albert Calmette

BCG Vaccine

16

Example: BCG Vaccine

	Tuberculosis		
	Positive	**Negative**	
Vaccinated	4	119	123
Not Vaccinated	11	128	139

$$p_T = 4/123 = .0325$$
$$p_C = 11/139 = .0791$$
$$RR = \frac{4/123}{11/139} = .41$$

$$y = \ln[RR] = \ln\left[\frac{4/123}{11/139}\right] = -.89$$
$$v = \frac{1}{4} - \frac{1}{123} + \frac{1}{11} - \frac{1}{139} = .326$$

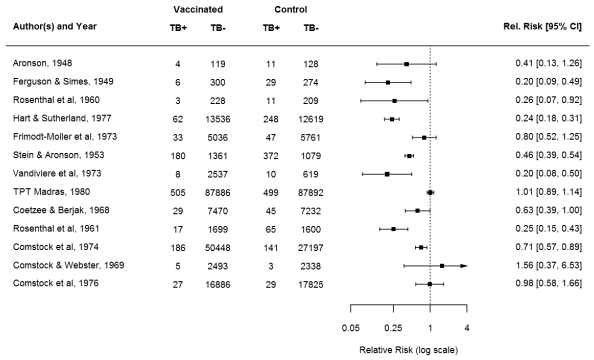
17

Example: BCG Vaccine

Study	Year	*RR*	*y* = ln(*RR*)	*v*	*Allocation*	Latitude
1	1948	0.41	-0.89	.326	random	44
2	1949	0.20	-1.59	.195	random	55
3	1960	0.26	-1.35	.415	random	42
4	1977	0.24	-1.44	.020	random	52
5	1973	0.80	-0.22	.051	alternate	13
6	1953	0.46	-0.79	.007	alternate	44
7	1973	0.20	-1.62	.223	random	19
8	1980	1.01	0.01	.004	random	13
9	1968	0.63	-0.47	.056	random	27
10	1961	0.25	-1.37	.073	systematic	42
11	1974	0.71	-0.34	.012	systematic	18
12	1969	1.56	0.45	.533	systematic	33
13	1976	0.98	-0.02	.071	systematic	33

18

Example: BCG Vaccine



Testing for Heterogeneity

$$H_0 : \theta_1 = \theta_2 = \ldots = \theta_k$$

$$\hat{\theta} = \frac{\sum w_i y_i}{\sum w_i} \quad \text{with} \quad w_i = 1/v_i$$

$$Q = \sum w_i (y_i - \hat{\theta})^2$$

if the effect sizes are really homogeneous, then Q follows a chi-square distribution with $k - 1$ degrees of freedom

Example: BCG Vaccine

Q = 152.23

critical value (for α = .05 and df = 12): 21.03

p-value: <.0001

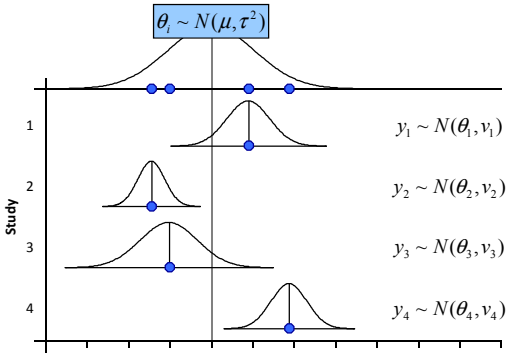
reject H_0: $\theta_1 = \theta_2 = \ldots = \theta_{13}$

conclusion: the true effects are heterogeneous

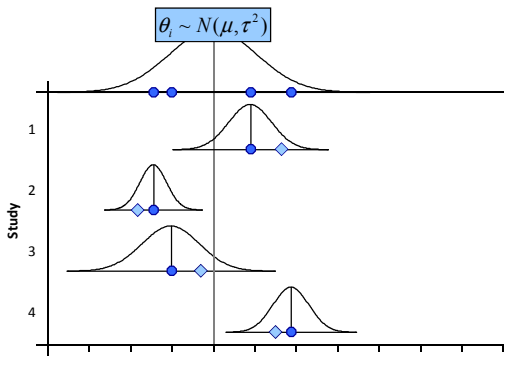
Note

- $Q \overset{H_0}{\sim} \chi^2_{k-1}$ only when assumptions are fulfilled
- in practice, not true (sometimes barely!)
- so this may be a rough approximation at best
- see also Hoaglin (2016) in Stat Med

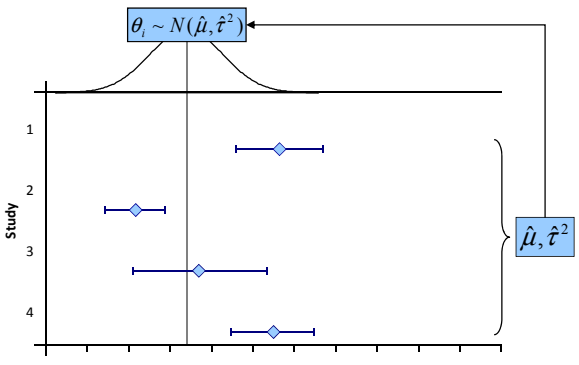
Random-Effects Model



Random-Effects Model



Random-Effects Model



Random-Effects Model

Model	$y_i = \overbrace{\mu + u_i}^{\theta_i} + \varepsilon_i$	$\varepsilon_i \sim N(0, v_i)$ $u_i \sim N(0, \tau^2)$
Parameter Estimate	$\hat{\mu} = \frac{\sum w_i y_i}{\sum w_i}$	$w_i = \frac{1}{v_i + \hat{\tau}^2}$
Var and SE of the Estimate	$Var[\hat{\mu}] = \frac{1}{\sum w_i}$	$SE[\hat{\theta}] = \sqrt{\frac{1}{\sum w_i}}$
Inference	$z = \frac{\hat{\mu}}{SE[\hat{\mu}]}$	$\hat{\mu} \pm 1.96 SE[\hat{\mu}]$

30

DerSimonian-Laird Estimator for τ^2

- method of moments estimator

$$\hat{\tau}^2 = \frac{Q - (k-1)}{\sum w_i - \frac{\sum w_i^2}{\sum w_i}} \qquad w_i = \frac{1}{v_i}$$

- if estimate is negative, then set it equal to 0
- semi-parametric in nature

31

Example: BCG Vaccine

$$\hat{\tau}^2 = \frac{Q - (k-1)}{\sum w_i - \frac{\sum w_i^2}{\sum w_i}} = \frac{152.23 - (13-1)}{609.7007 - \frac{94820.58}{609.7007}} = .3088$$

estimated variance in the true log relative risks

32

Example: BCG Vaccine

$\hat{\mu} = -.7141$
(estimated average log relative risk)

$(e^{-.7141} \approx .49)$
(estimated average relative risk)

$SE[\hat{\mu}] = .178742$

$z = -4.00$

$95\%\ \text{CI} : (-1.0644, -.3638)$
(95% CI for the true average log relative risk)

$(e^{-1.0644} \approx .34, e^{-.3638} \approx .70)$
(95% CI for the true average relative risk)

33

Note

- exponentiation is a non-linear transformation
- if $E[\theta_i] = \mu$, then $E[g(\theta_i)] \neq g(\mu)$
- $e^{\hat{\mu}}$ estimates the **median** true relative risk
- if you really want to estimate the mean:

$$e^{\hat{\mu} + \hat{\tau}^2/2}$$

- but nobody does that … ☹

34

Credibility/Prediction Interval

- interval where approximately 95% of the true outcomes are estimated/predicted to fall:

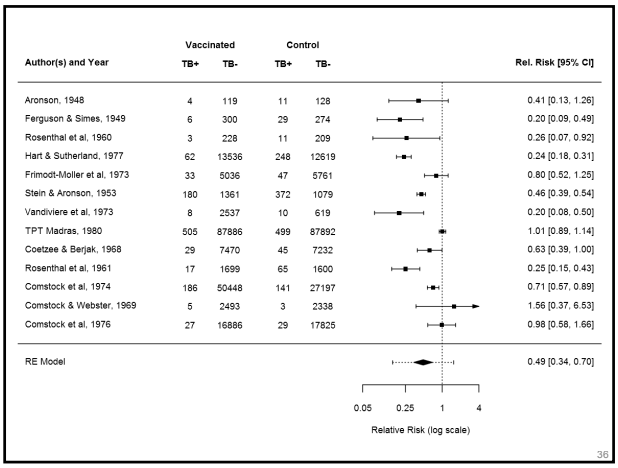
$$\hat{\mu} \pm 1.96\sqrt{\hat{\tau}^2}$$

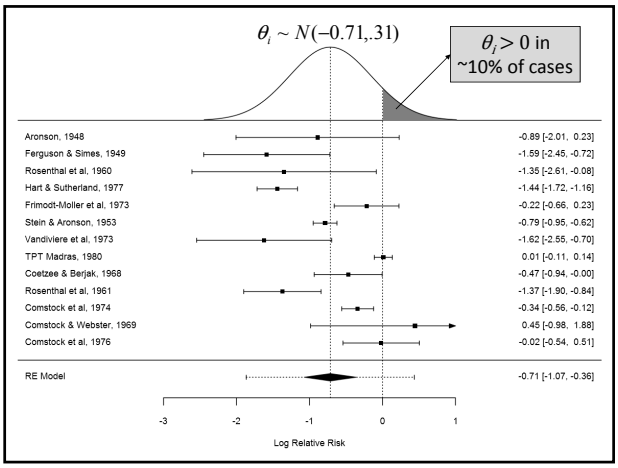
- example:

$$-0.71 \pm 1.96\sqrt{.31} = -1.80 \text{ to } 0.37$$

back-transformed: 0.16 to 1.45

35





Credibility/Prediction Interval

- interval ignores uncertainty in $\hat{\mu}$ (i.e., $Var[\hat{\mu}]$)
- an improved 95% interval:

$$\hat{\mu} \pm 1.96\sqrt{\hat{\tau}^2 + Var[\hat{\mu}]}$$

- example:

$$-0.71 \pm 1.96\sqrt{.31 + .032} = -1.86 \text{ to } 0.43$$

back-transformed: 0.16 to 1.54

Empirical Bayes Estimates

- also called best linear unbiased predictions (BLUPs) for the true effects
- optimally combines information from y_i and $\hat{\mu}$ to estimate θ_i (minimum MSE)

$$\lambda_i = \frac{\hat{\tau}^2}{\hat{\tau}^2 + v_i}$$

$$\hat{\theta}_i^{EB} = \lambda_i y_i + (1-\lambda_i)\hat{\mu}$$

Example: BCG Vaccine

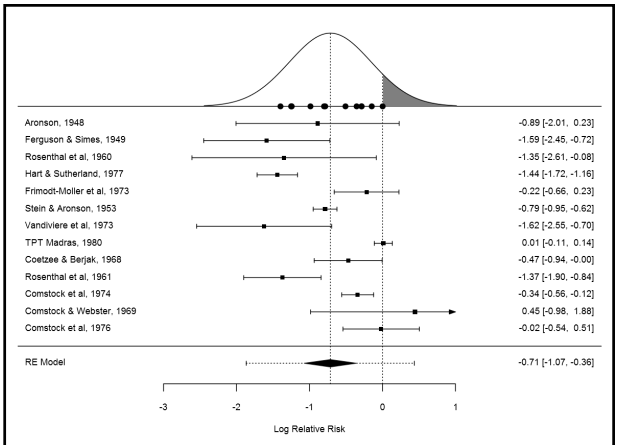
$RR_1 = .41$
$y_1 = -.89$
$v_1 = .326$

$\exp[\hat{\mu}] = .49$
$\hat{\mu} = -.71$
$\hat{\tau}^2 = .309$

$$\lambda_1 = \frac{\hat{\tau}^2}{\hat{\tau}^2 + v_1} = \frac{.309}{.309 + .326} = .487$$

$$\hat{\theta}_1^{EB} = (.487)(-.89) + (1-.487)(-.71) = -.80$$

$$\exp[\hat{\theta}_1^{EB}] = \exp[-.80] = .45$$



Meta-Analysis with R (*metafor*)

- install with: `install.packages("metafor")`
- (only need to do this once, or after reinstalling R, or to upgrade to a new package version)
- load package with: `library(metafor)`
- (have to do this each time you (re)start R)
- **put your commands in a script file!!!**
- if you are new to R, consider using RStudio
- comments start with `#` (use them!)

Loading External/Internal Datasets

- for an external dataset, change the working directory to where the data file is stored
  - Windows: File → Change Dir
  - MacOS: Misc → Change Working Directory
  - RStudio: Session → Set Working Directory

```
dat <- read.table("data_bcg.txt", header=TRUE)
```

- for an internal dataset:

```
dat <- get(data(dat.bcg))
```



```
> ### load BCG vaccine data
> dat <- get(data(dat.bcg))
> dat
```

	treated	tpos	tneg
	control	cpos	cneg

	trial	author	year	tpos	tneg	cpos	cneg	ablat	alloc
1	1	Aronson	1948	4	119	11	128	44	random
2	2	Ferguson & Simes	1949	6	300	29	274	55	random
3	3	Rosenthal et al	1960	3	228	11	209	42	random
4	4	Hart & Sutherland	1977	62	13536	248	12619	52	random
5	5	Frimodt-Moller et al	1973	33	5036	47	5761	13	alternate
6	6	Stein & Aronson	1953	180	1361	372	1079	44	alternate
7	7	Vandiviere et al	1973	8	2537	10	619	19	random
8	8	TPT Madras	1980	505	87886	499	87892	13	random
9	9	Coetzee & Berjak	1968	29	7470	45	7232	27	random
10	10	Rosenthal et al	1961	17	1699	65	1600	42	systematic
11	11	Comstock et al	1974	186	50448	141	27197	18	systematic
12	12	Comstock & Webster	1969	5	2493	3	2338	33	systematic
13	13	Comstock et al	1976	27	16886	29	17825	33	systematic

Computing Observed Outcomes

- can of course use external software for data management and preparations
- to compute outcomes: `escalc()` command
- basic syntax:

```
dat <- escalc(measure="", ..., data=dat)
```

- to specify the outcome measure (RD, RR, OR, MD, SMD, ROM, ...)
- to specify the variables needed to compute the observed outcomes
- name of data frame containing the variables


```
> ### calculate log relative risks and sampling variances
> dat <- escalc(measure="RR", ai=tpos, bi=tneg,
                              ci=cpos, di=cneg, data=dat)
> dat
```

	trial	author	year	...	yi	vi
1	1	Aronson	1948	...	-0.8893	0.3256
2	2	Ferguson & Simes	1949	...	-1.5854	0.1946
3	3	Rosenthal et al	1960	...	-1.3481	0.4154
4	4	Hart & Sutherland	1977	...	-1.4416	0.0200
5	5	Frimodt-Moller et al	1973	...	-0.2175	0.0512
6	6	Stein & Aronson	1953	...	-0.7861	0.0069
7	7	Vandiviere et al	1973	...	-1.6209	0.2230
8	8	TPT Madras	1980	...	0.0120	0.0040
9	9	Coetzee & Berjak	1968	...	-0.4694	0.0564
10	10	Rosenthal et al	1961	...	-1.3713	0.0730
11	11	Comstock et al	1974	...	-0.3394	0.0124
12	12	Comstock & Webster	1969	...	0.4459	0.5325
13	13	Comstock et al	1976	...	-0.0173	0.0714

log relative risks and sampling variances

Random-Effects Model

- basic syntax:

```
res <- rma(yi, vi, method="DL", data=dat)
```

- name of variable for the observed outcomes
- name of variable for the corresponding sampling variances
- to select the τ^2 estimator (DL, ML, REML, PM, EB, ...)
- name of data frame containing the variables

- to print results, type: `res`
- or use: `print(res, digits=2)`

Random-Effects Model

- default is `method="REML"`
- use `predict()` to get credibility/prediction interval (and back-transformation if applicable)

```
predict(res, digits=2)
predict(res, transf=<>, digits=2)
```

  - for exponentiation: `transf=exp`
  - for z-to-r transformation: `transf=transf.ztor`
- to obtain BLUPs: `blup()`

48

```
> ### fit random-effects model
> res <- rma(yi, vi, method="DL", data=dat)
> res

Random-Effects Model (k = 13; tau^2 estimator: DL)

tau^2 (estimated amount of total heterogeneity): 0.3088
tau (square root of estimated tau^2 value):      0.5557
I^2 (total heterogeneity / total variability):   92.12%
H^2 (total variability / sampling variability):  12.69

Test for Heterogeneity:
Q(df = 12) = 152.2330, p-val < .0001

Model Results:

estimate       se     zval     pval    ci.lb    ci.ub
 -0.7141   0.1787  -3.9952   <.0001  -1.0644  -0.3638

> ### estimated average relative risk (and 95% CI/CR)
> predict(res, transf=exp, digits=2)
 pred ci.lb ci.ub cr.lb cr.ub
 0.49  0.34  0.70  0.16  1.54
```

cr.lb/cr.ub = bounds of a 95% credibility/prediction interval

49

```
> ### best linear unbiased predictions of true log ORs
> blup(res, digits=2)

    pred   se pi.lb pi.ub
1  -0.80 0.41 -1.60  0.00
2  -1.25 0.35 -1.94 -0.56
3  -0.99 0.43 -1.84 -0.13
..   ...  ...   ...   ...
13 -0.15 0.24 -0.62  0.33

> ### best linear unbiased predictions of true ORs
> blup(res, transf=exp, digits=2)

   pred pi.lb pi.ub
1  0.45  0.20  1.00
2  0.29  0.14  0.57
3  0.37  0.16  0.87
..  ...   ...   ...
13 0.86  0.54  1.39
```

50

Exercises

- look at: **`exercises.r`**
- example 1:
  - effects of diuretics in pregnancy on the risk of any form of pre-eclampsia (Collins et al., 1985)
  - outcome measure: (log) odds ratio
- example 2:
  - effectiveness of self-management education and regular medical review for adults with asthma on the mean number of days off work/school (Gibson et al., 2002)
  - outcome measure: standardized mean difference
- save your syntax! (will come back to examples later)

51

Estimators for τ^2

- DerSimonian-Laird estimator
- Hedges (& Olkin) (Cochran) estimator
- Hunter-Schmidt estimator
- Sidik-Jonkman estimator
- maximum likelihood estimator
- restricted maximum likelihood estimator
- empirical Bayes / Paule-Mandel estimator
- generalized Q-statistic estimator
- …

53

Generalized Q-statistic Estimator

- define:

  $Q_{gen} = \sum w_i (y_i - \hat{\theta})^2$ with $\hat{\theta} = \sum w_i y_i / \sum w_i$

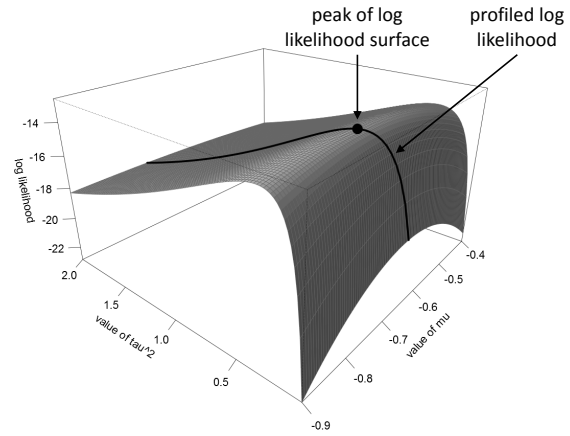
  for any fixed but arbitrary weights
- MoM: find $E[Q_{gen}]$ and then solve for τ^2
  - DL estimator: $w_i = 1/v_i$
  - HE estimator: $w_i = 1$
  - another interesting option: $w_i = 1/\sqrt{v_i}$

54

ML and REML Estimation

- model implies: $y_i \sim N(\mu, \tau^2 + v_i)$
- can easily write down the log likelihood (ll)
- for given τ^2, MLE of μ:
  $\hat{\mu} = \sum w_i y_i / \sum w_i$ where $w_i = 1/(\tau^2 + v_i)$
- so profile out μ from ll and optimize over τ^2
- REML takes linear combination of the $y_i's$, so that resulting data are independent of μ (and then we again optimize over τ^2)

55



56

Paule-Mandel Estimator

- define:
  $Q_{gen} = \sum w_i (y_i - \hat{\mu})^2$ with $\hat{\mu} = \sum w_i y_i / \sum w_i$
  and use weights $w_i = 1/(\tau^2 + v_i)$
- then $Q_{gen} \sim \chi^2_{k-1}$ (so then: $E[Q_{gen}] = k - 1$)
- so find that value of τ^2, so that $Q_{gen} = k - 1$
- identical to the empirical Bayes estimator (Morris, 1983; Berkey et al., 1995)

57

Bias and Efficiency

- bias (before truncation of negative values):
  - HS and ML are negatively biased
  - SJ is positively biased (esp. for small τ^2)
  - HE, DL, REML, EB/PM, GENQ are (approx.) unbiased
- efficiency:
  - HS and ML are most efficient (usually)
  - HE is least efficient (usually)
  - DL, SJ, REML, EB/PM somewhere in between
  - for GENQ, depends on the weights used

58

Example: BCG Vaccine

Estimator	$\hat{\tau}^2$	*SE*	I^2	H^2
DL	0.31	0.230	92.1	12.7
HE	0.33	0.207	92.6	13.4
HS	0.23	0.128	89.6	9.6
SJ	0.35	0.150	92.9	14.1
ML	0.28	0.144	91.4	11.6
REML	0.31	0.166	92.2	12.9
EB/PM	0.32	0.174	92.3	13.0

59

Bias in Estimators

- when model assumptions break down, even unbiased estimators can become biased
- problem stems (in part) from poor estimates of the sampling variances (when sample sizes are small and/or the event is rare)
  - DL and HS based on $w_i = 1/v_i$
  - HE is unweighted
  - SJ based on $w_i = 1/(\hat{\tau}_0^2 + v_i)$
  - ML, REML, EB/PM based on $w_i = 1/(\hat{\tau}^2 + v_i)$

60

Estimators for τ^2 in metafor

- specify via `method=""` (DL, HE, HS, SJ, ML, REML, EB, PM, GENQ)
- for GENQ, must specify weights; for example:

```
rma(yi, vi, weights=1/vi, method="GENQ")
```

- ML, REML, EB, and PM are iterative (see progress with `verbose=TRUE`)
- non-convergence can be solved

61

Confidence Interval for τ^2

- the precision of $\hat{\tau}^2$ depends mostly on k
- k is often small in applications
- so τ^2 is usually estimated imprecisely
- may want to obtain a CI for τ^2

62

Confidence Interval for τ^2

- Wald-type CI: $\hat{\tau}^2 \pm 1.96 \times SE[\hat{\tau}^2]$ (don't!)
- Q-profile
- GENQ method
- profile likelihood
- …

63

Q-Profile Method

$$Q = \sum \frac{(y_i - \hat{\theta})^2}{v_i} \overset{H_0}{\sim} \chi^2_{k-1}$$

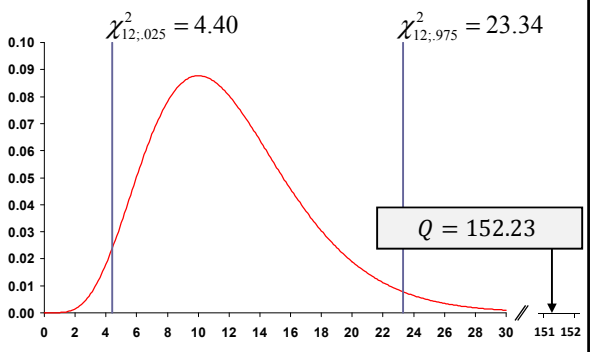
$$Q_{(\tau^2)} = \sum \frac{(y_i - \hat{\mu}_{(\tau^2)})^2}{v_i + \tau^2} \sim \chi^2_{k-1}$$

95% Confidence Interval for τ^2

- let $\hat{\tau}^2_{LB} = \tilde{\tau}^2$: $Q_{(\tilde{\tau}^2)} = \chi^2_{k-1;.975}$
- let $\hat{\tau}^2_{UB} = \tilde{\tau}^2$: $Q_{(\tilde{\tau}^2)} = \chi^2_{k-1;.025}$

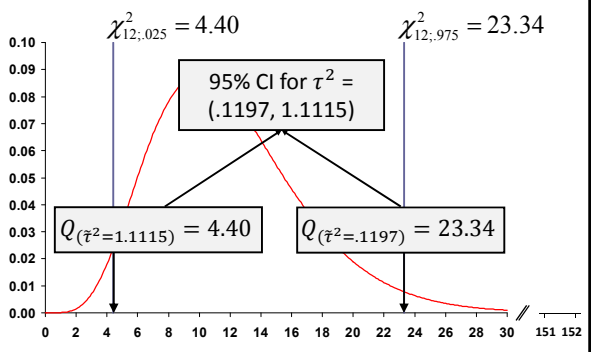
64

Example: BCG Vaccine



65

Example: BCG Vaccine



68

Software

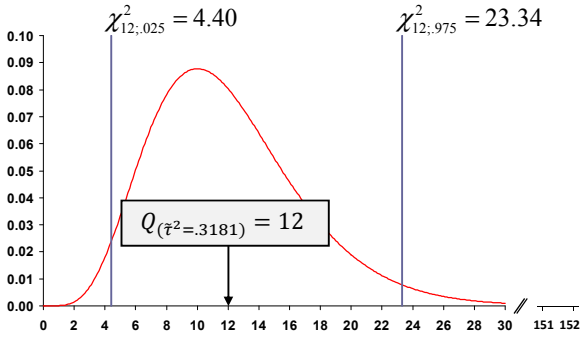
`confint(res)`

res = an object that is a fitted RE model from the rma() function

```
> res <- rma(yi, vi, method="DL", data=dat)
> confint(res)

       estimate   ci.lb   ci.ub
tau^2    0.3088  0.1197  1.1115
tau      0.5557  0.3460  1.0543
I^2(%)  92.1173 81.9177 97.6781
H^2     12.6861  5.5303 43.0680
```

69

Example: BCG Vaccine



70

GENQ Method

- Biggerstaff & Jackson (2008) & Jackson (2013) derived the exact distribution of Q_{gen}

$$Q_{gen} \sim \sum \lambda_i \chi^2_1$$

where λ_i depends on τ^2

- can use this to derive an exact CI for τ^2

71

Software

`confint(res)`

res = an object that is a fitted RE model from the rma() function using method="GENQ"

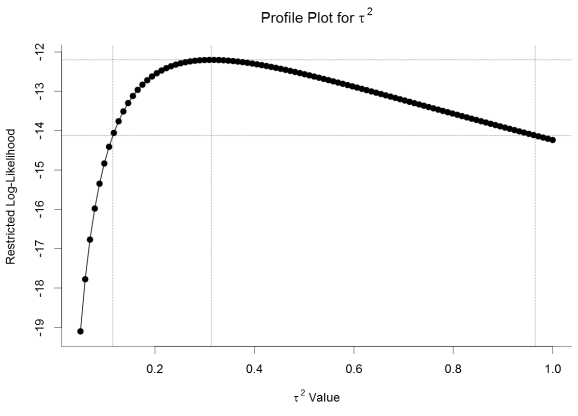
```
> res <- rma(yi, vi, weights=1/vi,
             method="GENQ", data=dat)
> confint(res)

       estimate   ci.lb   ci.ub
tau^2    0.3088  0.1011  1.6361
tau      0.5557  0.3179  1.2791
I^2(%)  92.1173 79.2728 98.4108
H^2     12.6861  4.8246 62.9228
```

72

Profile Likelihood Method

- $LRT = -2(ll_{\tau^2=0} - ll_{\hat{\tau}^2}) \overset{H_0}{\sim} \chi^2_1$
- find all τ^2 values that would not be rejected
- can do this with ML and REML

73



74

Software

```
res <- rma.mv(yi, vi, random = ~ 1 | trial, data=dat)
confint(res)
```

```
> res <- rma.mv(yi, vi, random = ~ 1 | trial, data=dat)
> confint(res)

        estimate  ci.lb  ci.ub
sigma^2   0.3132 0.1152 0.9647
sigma     0.5597 0.3395 0.9822
```


Note

- method consistency:
  - Q-profile CI can exclude $\hat{\tau}^2$ when not using PM
  - so should use PM together with Q-profile
  - GENQ CI will contain $\hat{\tau}^2$ (GENQ)
  - profile likelihood CI will contain $\hat{\tau}^2$ (ML/REML)
- 95% CI for τ^2 may include 0, but Q-test rejects at $\alpha = .05$ (a 90% CI will be consistent)

Quantifying Heterogeneity

- the raw estimate of τ^2 is difficult to interpret (is the value small/large?)
- cannot compare τ^2 estimates across different effect size or outcome measures (e.g., log risk ratios, standardized mean differences, correlations, ...)

Quantifying Heterogeneity

I^2 estimates (in %) how much of the total variability in the effect size estimates is due to heterogeneity among the true effects

$$I^2 = 100\% \times \frac{\hat{\tau}^2}{\hat{\tau}^2 + \tilde{v}} \qquad \tilde{v} = \frac{(k-1)\sum w_i}{(\sum w_i)^2 - \sum w_i^2} \qquad w_i = 1/v_i$$

$$= 100\% \times \frac{Q-(k-1)}{Q}$$ (when estimating τ^2 with the DL estimator)

Example: BCG Vaccine

$k = 13 \qquad Q = 152.23 \qquad \hat{\tau}^2 = .3088$

$$\left.\begin{aligned} I^2 &= 100\% \times \frac{.3088}{.3088 + .0264} \\ &= 100\% \times \frac{152.23-(13-1)}{152.23} \end{aligned}\right\} = 92\% \qquad \tilde{v} = .0264$$

Quantifying Heterogeneity

H^2 estimates the ratio of the total amount of variability (heterogeneity plus sampling variance) to the amount of sampling variance

$$H^2 = \frac{\hat{\tau}^2 + \tilde{v}}{\tilde{v}} \qquad \tilde{v} = \frac{(k-1)\sum w_i}{(\sum w_i)^2 - \sum w_i^2} \qquad w_i = 1/v_i$$

$$= \frac{Q}{k-1}$$ (when estimating τ^2 with the DL estimator)

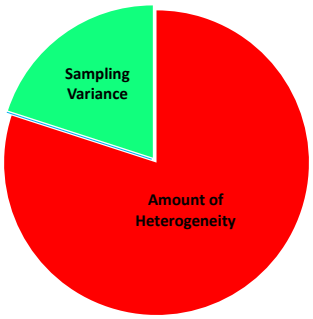
Example: BCG Vaccine

$k = 13$ $\quad Q = 152.23$ $\quad \hat{\tau}^2 = .3088$

$$H^2 = \frac{.3088 + .0264}{.0264} = \frac{152.23}{13-1} \Bigg\} = 12.69 \qquad \widetilde{v} = .0264$$

81

Quantifying Heterogeneity



$$I^2 = \frac{\square}{\square + \square}$$

$$H^2 = \frac{\square + \square}{\square}$$

82

CI for I^2 and H^2

- can plug bounds of CI for τ^2 into equations for I^2 and H^2 to obtain CIs for these measures
- exact for Q-profile and GENQ method
- intervals are often very wide

83

Software

`confint(res)`

res = an object that is a fitted RE model from the rma() function

```
> res <- rma(yi, vi, method="DL", data=dat)
> confint(res)

       estimate   ci.lb   ci.ub
tau^2    0.3088  0.1197  1.1115
tau      0.5557  0.3460  1.0543
I^2(%)  92.1173 81.9177 97.6781
H^2     12.6861  5.5303 43.0680
```

84

Refined Tests/CIs for μ

- inference for μ assumes known sampling variances (v_i) and amount of heterogeneity (τ^2)
- → incorrect Type I error rate / CI coverage
- refined methods have been developed that account for the uncertainty in τ^2

85

Knapp and Hartung Method

- estimate μ with weights $w_i = 1/(v_i + \hat{\tau}^2)$
- then compute: $s^2 = \frac{\sum w_i (y_i - \hat{\mu})^2}{k-1}$
- $Var[\hat{\mu}] = \frac{s^2}{\sum w_i}$ and $SE[\hat{\mu}] = \sqrt{Var[\hat{\mu}]}$
- $t = \hat{\mu}/SE[\hat{\mu}]$ and $\hat{\mu} \pm t_{crit} SE[\hat{\mu}]$
- use t-distribution with $k - 1$ df
- use **knha=TRUE** in **rma()**

86

```
> ### fit random-effects model with Knapp & Hartung method
> res <- rma(yi, vi, method="DL", data=dat, knha=TRUE)
> res

Random-Effects Model (k = 13; tau^2 estimator: DL)

tau^2 (estimated amount of total heterogeneity): 0.3088
tau (square root of estimated tau^2 value):      0.5557
I^2 (total heterogeneity / total variability):   92.12%
H^2 (total variability / sampling variability):  12.69

Test for Heterogeneity:
Q(df = 12) = 152.2330, p-val < .0001

Model Results:

estimate       se     tval     pval    ci.lb    ci.ub
 -0.7141   0.1807  -3.9520   0.0019  -1.1078  -0.3204

> ### estimated average relative risk (and 95% CI/CR)
> predict(res, transf=exp, digits=2)
 pred ci.lb ci.ub cr.lb cr.ub
 0.49  0.33  0.73  0.14  1.75
```

87

Credibility/Prediction Interval

- 95% interval with K&H method:

$$\hat{\mu} \pm t_{crit}\sqrt{\hat{\tau}^2 + Var[\hat{\mu}]}$$

- use $k - 1$ df for consistency
- Riley, Higgins, & Deeks (2011) use $k - 2$ df

88

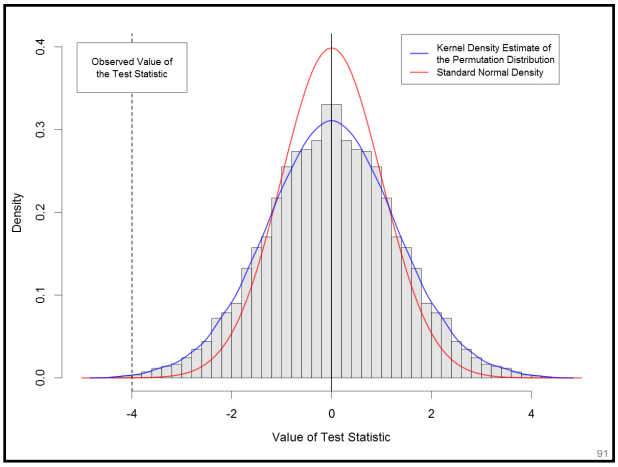
Permutation Test

- if $H_0: \mu = 0$ is true, then sign of y_i is arbitrary
- compute $z = \hat{\mu}/SE[\hat{\mu}]$ in the usual manner
- then randomly permute signs of the y_i values, refit model, and compute test statistic
- repeat m times: $z_1, z_2, \ldots, z_m$
- p-value: 2 × proportion of times that z is as extreme or more extreme than $z_1, z_2, \ldots, z_m$
- exact test requires 2^k permutations (note: $p \leq .05$ only possible with $k \geq 6$)
- see Follman & Proschan (1999)

89

```
> res <- rma(yi, vi, method="DL", data=dat)
> permutest(res, exact=TRUE)

Running 8192 iterations for exact permutation test.

  |==========================================================| 100%

Model Results:

estimate       se     zval    pval*    ci.lb    ci.ub
 -0.7141   0.1787  -3.9952   0.0017  -1.0644  -0.3638
```

90



91

Should I Use Them?

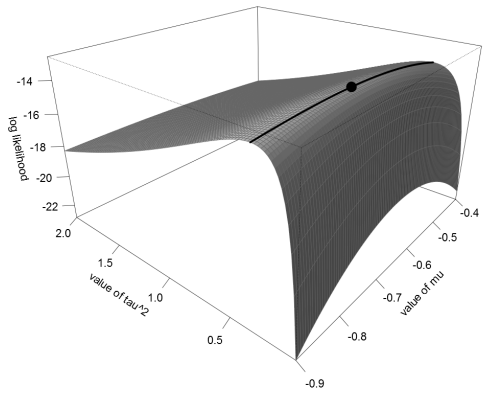
- simple answer: yes
- K&H method should be the default
- with small k, CI can be very wide
- how it should be! (making inferences in the RE model with small k is hard)
- added bonus of K&H method: choice of τ^2 estimator becomes much less important
- permutation test is even better (but not useful for small k & obtaining CI is tricky)

92

Profile Likelihood CI for μ

- can also use profile likelihood method to get CI for μ (Hardy & Thompson, 1996)
- (not implemented in metafor)

93



94

Non-Normal True Effects

- assumed $\theta_i \sim N(\mu, \tau^2)$ throughout
- if not (approximately) true:
  - relatively little effect on inferences for μ
  - more problematic when making inferences for τ^2
  - can totally screw up inferences about the distribution of θ_i (i.e., CR/PI)
- consider models with non-normal θ_i (e.g., Lee & Thompson, 2008; Baker & Jackson, 2008) or mixture distributions (e.g., van Houwelingen et al., 1993, 2002)

95

Literature

Baker, R., & Jackson, D. (2008). A new approach to outliers in meta-analysis. Health Care Management Science, 11(2), 121-131.

Colditz, G. A., Brewer, T. F., Berkey, C. S., Wilson, M. E., Burdick, E., Fineberg, H. V., & Mosteller, F. (1994). Efficacy of BCG vaccine in the prevention of tuberculosis: Meta-analysis of the published literature. Journal of the American Medical Association, 271(9), 698-702.

Collins, R., Yusuf, S., & Peto, R. (1985). Overview of randomised trials of diuretics in pregnancy. British Medical Journal, 290(6461), 17-23.

Follmann, D. A., & Proschan, M. A. (1999). Valid inference in random effects meta-analysis. Biometrics, 55(3), 732-737.

Gibson, P. G., Powell, H., Wilson, A., Abramson, M. J., Haywood, P., Bauman, A., Hensley, M. J., Walters, E. H., & Roberts, J. J. L. (2002). Self-management education and regular practitioner review for adults with asthma. Cochrane Database of Systematic Reviews, Issue 3.

Hartung, J., & Knapp, G. (2001). On tests of the overall treatment effect in meta-analysis with normally distributed responses. Statistics in Medicine, 20(12), 1771-1782.

Hartung, J., & Knapp, G. (2001). A refined method for the meta-analysis of controlled clinical trials with binary outcome. Statistics in Medicine, 20(24), 3875-3889.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. Journal of Educational Statistics, 6(2), 107-128.

Hardy, R. J., & Thompson, S. G. (1996). A likelihood approach to meta-analysis with random effects. Statistics in Medicine, 15(6), 619-629.

98

Hoaglin, D. C. (2016). Misunderstandings about Q and 'Cochran's Q test' in meta-analysis. Statistics in Medicine, 35(4), 485-495

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. Statistics in Medicine, 21(11), 1539-1558.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. British Medical Journal, 327(7414), 557-560.

Lee, K. J., & Thompson, S. G. (2008). Flexible parametric models for random-effects distributions. Statistics in Medicine, 27(3), 418-434.

Olkin, I., & Pratt, J. W. (1958). Unbiased estimation of certain correlation coefficients. Annals of Mathematical Statistics, 29(1), 201-211.

Riley, R. D., Higgins, J. P., & Deeks, J. J. (2011). Interpretation of random effects meta-analyses. British Medical Journal, 342, d549.

Sanchez-Meca, J., & Marin-Martinez, F. (2008). Confidence intervals for the overall effect size in random-effects meta-analysis. Psychological Methods, 13(1), 31-48.

van Houwelingen, H. C., Zwinderman, K. H., & Stijnen, T. (1993). A bivariate approach to meta-analysis. Statistics in Medicine, 12(24), 2273-2284.

van Houwelingen, H. C., Arends, L. R., & Stijnen, T. (2002). Advanced methods in meta-analysis: Multivariate approach and meta-regression. Statistics in Medicine, 21(4), 589-624.

Viechtbauer, W. (2007). Confidence intervals for the amount of heterogeneity in meta-analysis. Statistics in Medicine, 26(1), 37-52.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. Journal of Statistical Software, 36(3), 1-48.

99