Advanced Topics in meta analysis

Shayan Mostafaei
Assistant Professor of Biostatistics

Contents

What is a meta-analysis?

- A statistical analysis of the results from independent studies, which generally aims to produce a single estimate of the treatment effect

Egger et al, 2001

انواع واریانس‌ها در مطالعات متاانالیز

- واریانس درون مطالعات Within studies variation

این واریانس حاصل از تفاوت بین داده‌های هر مطالعه اولیه براساس پارامتر مورد بررسی در بین آنها است. این واریانس به دلیل ماهیت تصادفی بودن نمونه‌های در هر مطالعه اولیه است.

- واریانس بین مطالعات Between studies variation

این وارایانس حاصل از تفاوت بین نتایج مطالعات اولیه براساس پارامتر مورد بررسی در بین آنها است. این واریانس ممکن است در اثر منشا بالینی و یا متدولوژیک باشد.

Types of heterogeneity

- Clinical:

1. Heterogeneity between Participants
2. Heterogeneity between interventions (Intensity / dose / duration)

- Methodological:

1. Design
2. Conduct (Allocation concealment/blinding/analysis methods)
3. Multiple outcomes measurement tools

- Statistical: Statistical heterogeneity is the observed variation in effect sizes that cannot be explained by chance or random error alone.

Identifying and testing of heterogeneity

- Visual:

1. Forest plot
2. Funnel plot (bias or systematic heterogeneity)
3. Baujat plot (to detect studies overly contributing to the heterogeneity of a meta-analysis)

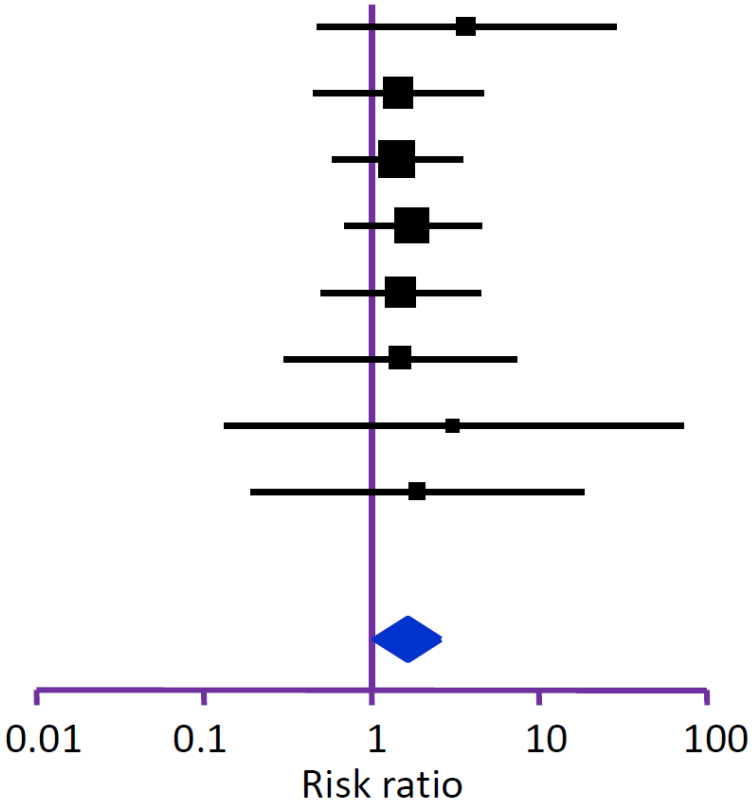
- Statistics:

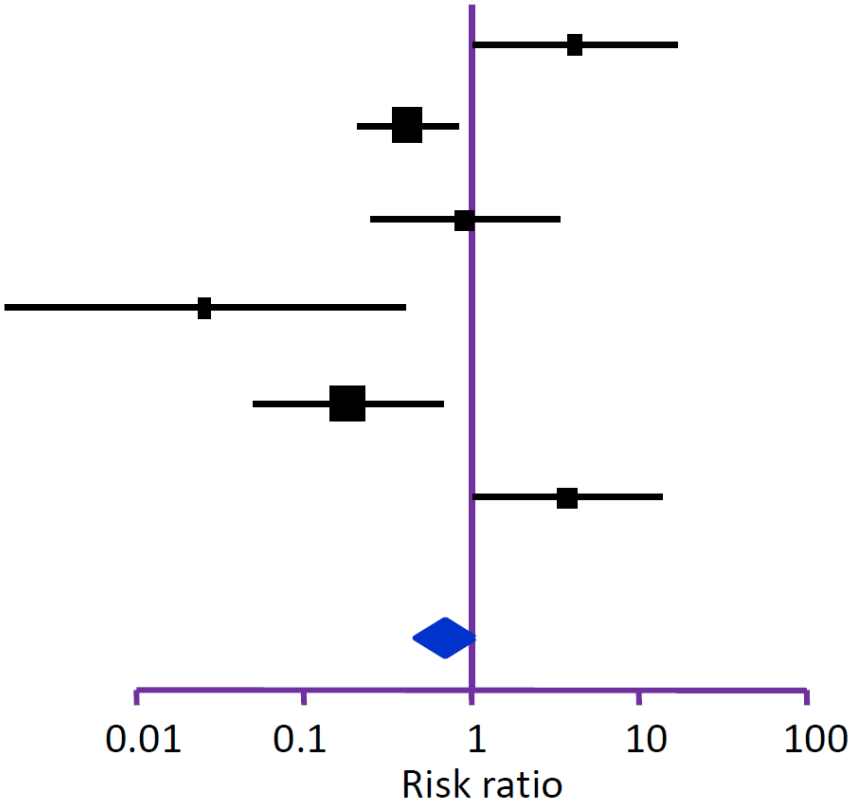
1. I^2
2. Tau^2

- Testing:

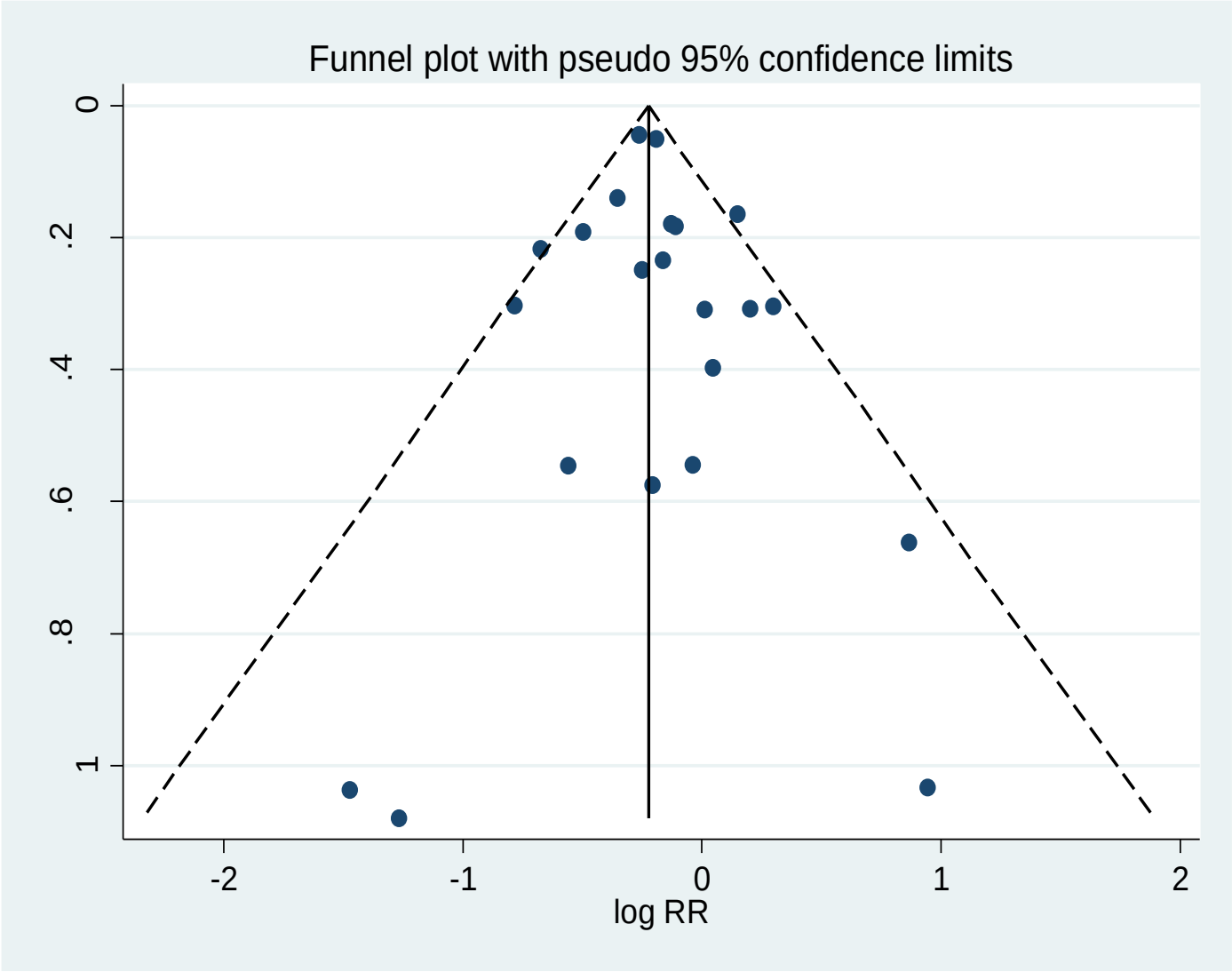
1. Q test based on the chi-square test

Forest plot

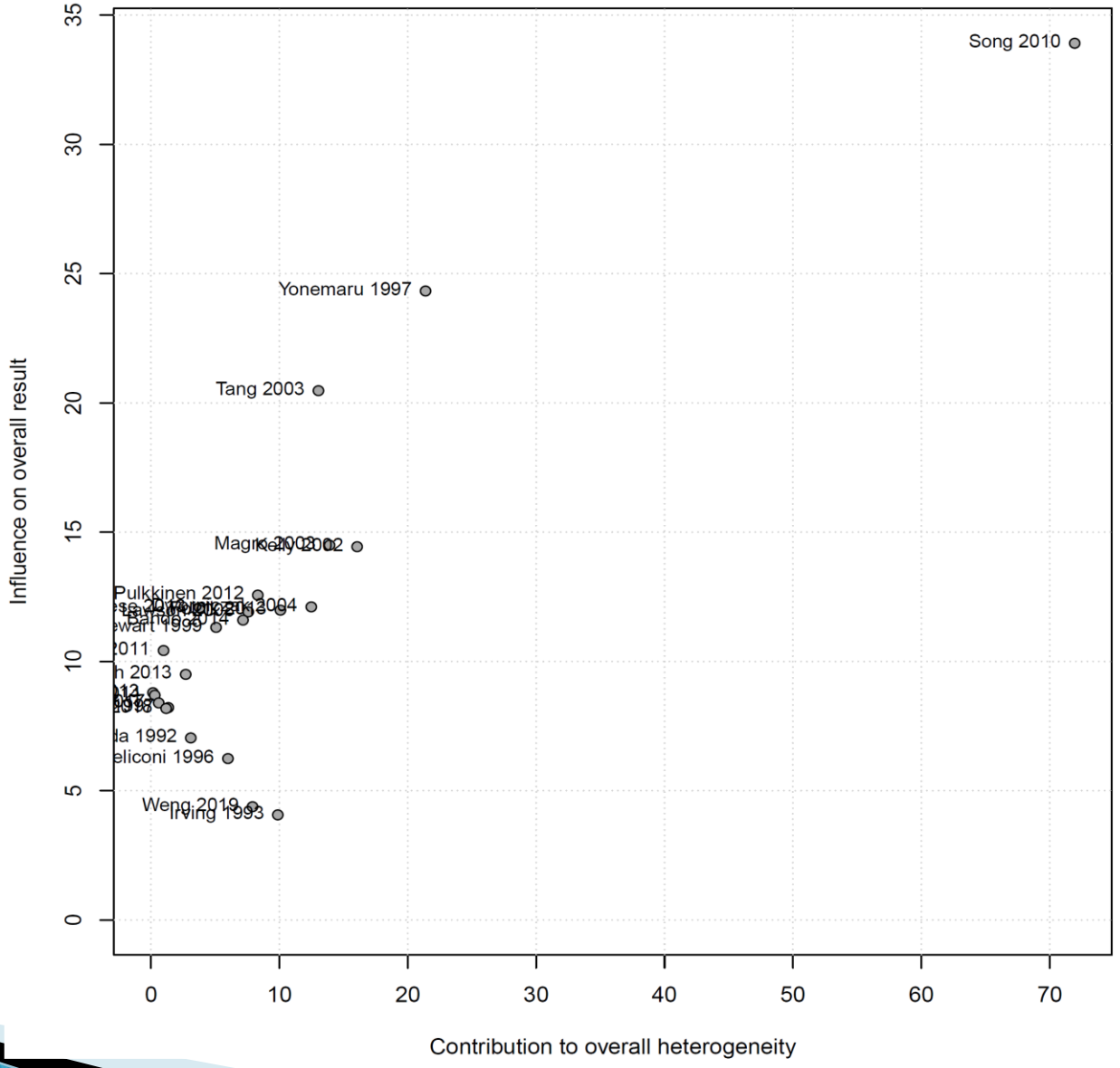




Funnel plot


Funnel plot with pseudo 95% confidence limits
0
.2
.4
.6
.8
1
-2
-1
0
1
2
log RR

Baujat plot


Influence on overall result
Contribution to overall heterogeneity
0
5
10
15
20
25
30
35
0
10
20
30
40
50
60
70
Song 2010
Yonemaru 1997
Tang 2003
Kelly 2002
Pulkkinen 2012
Irving 1993

Heterogeneity test and statistics

$$Q = \sum w_i \left(\hat{y}_i - \hat{\theta}\right)^2$$

$$I^2 = \frac{Q-k+1}{Q} * 100\%$$

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

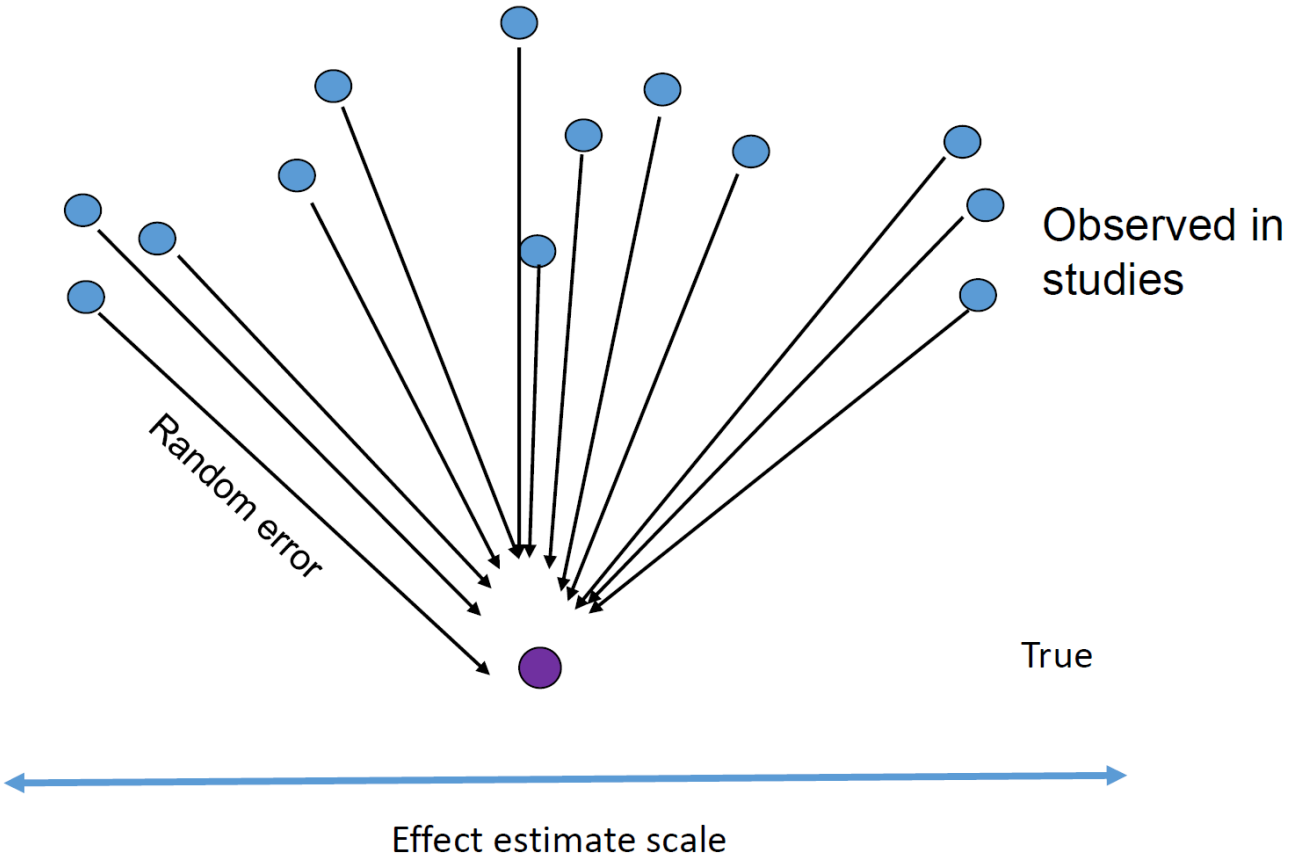
Interpretation of I-square statistic

- 0% to 40%: might not be important;
- 30% to 60%: may represent moderate heterogeneity*;
- 50% to 90%: may represent substantial heterogeneity*;
- 75% to 100%: considerable heterogeneity*.

*The importance of the observed value of I^2 depends on (1) magnitude and direction of effects, and (2) strength of evidence for heterogeneity (e.g. P value from the Chi^2 test, or a confidence interval for I^2: uncertainty in the value of I^2 is substantial when the number of studies is small).

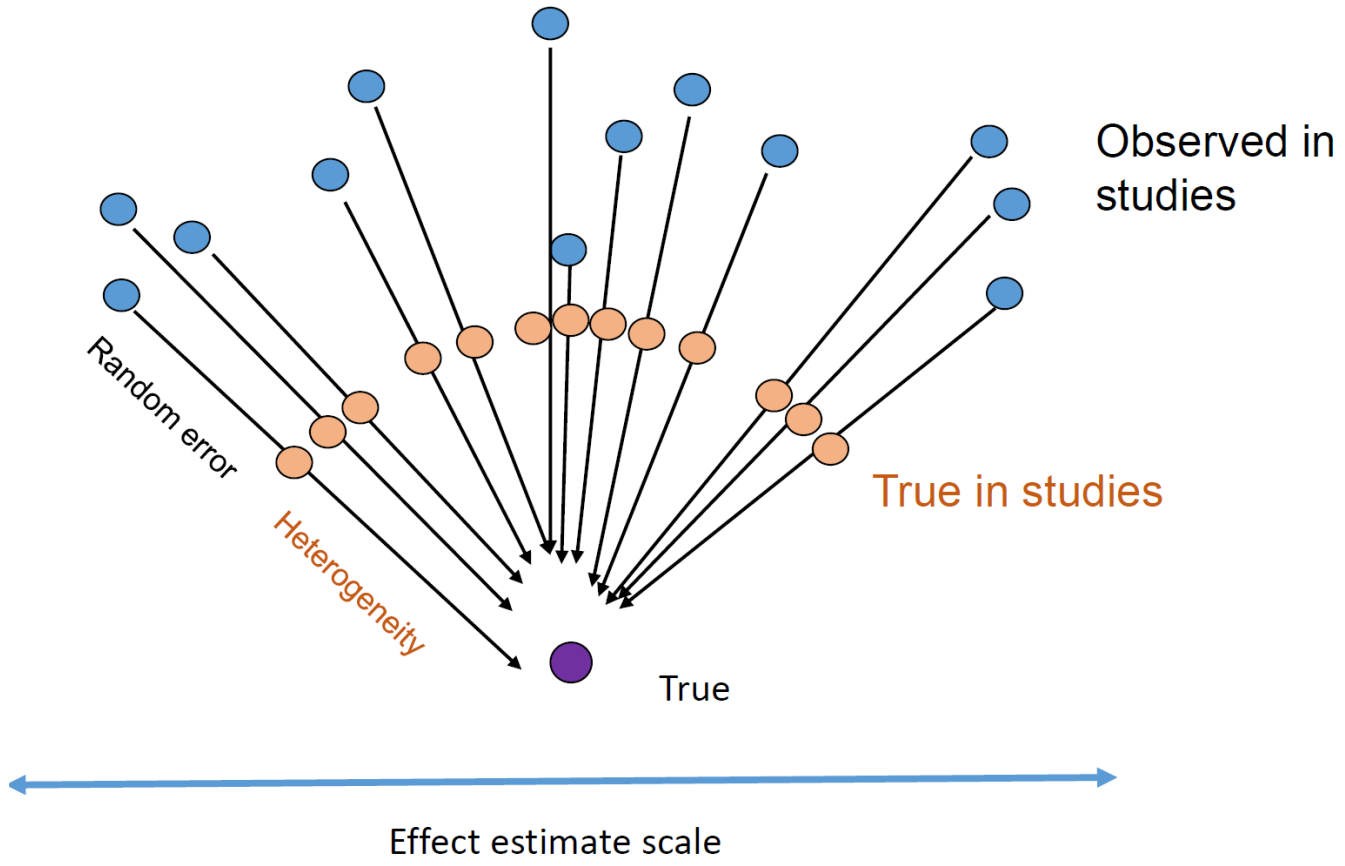
Fixed effect vs. random effects models

The fixed effect assumption



Fixed effect vs. random effects models

The random effects assumption



Fixed effect vs. random effects models

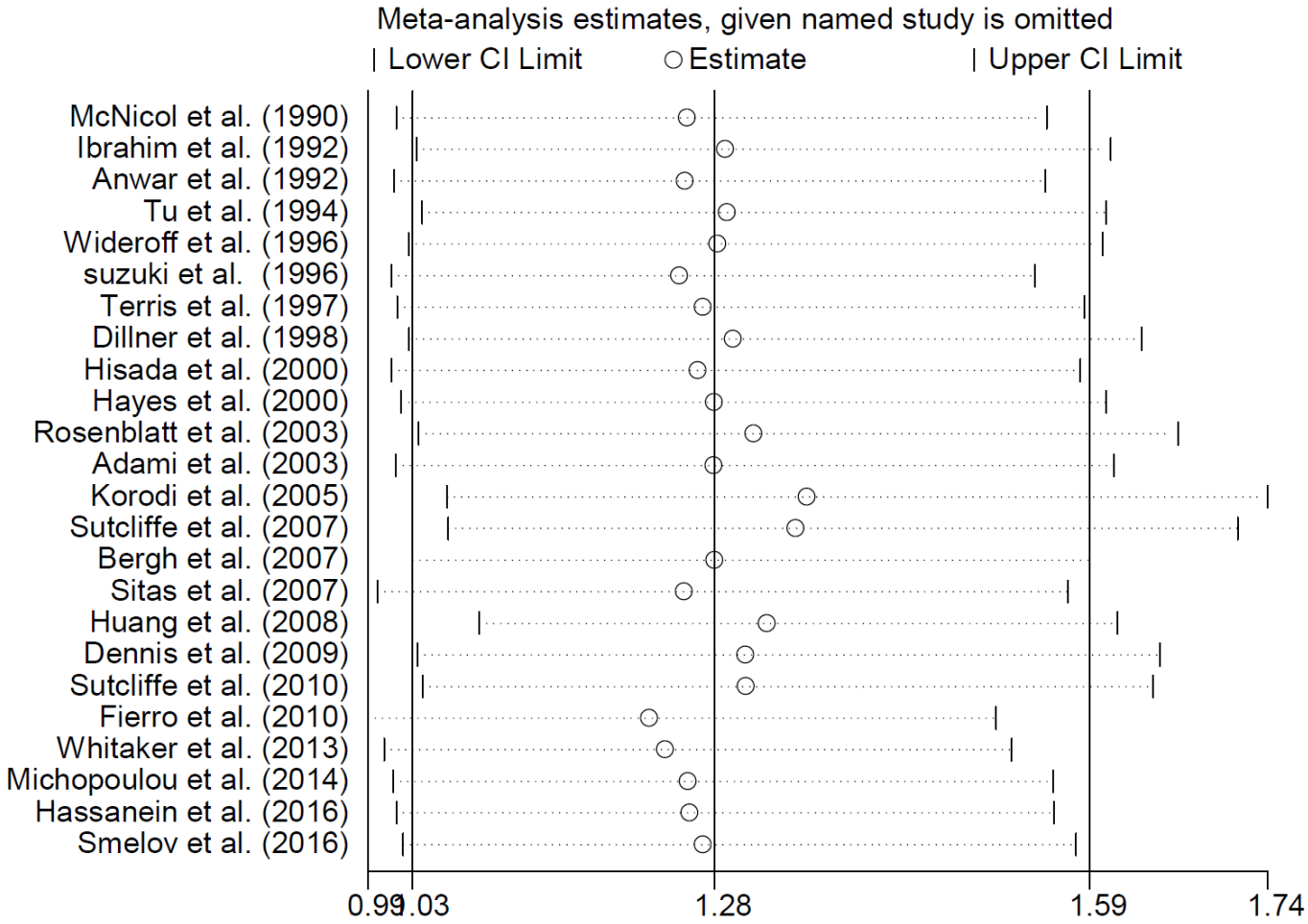
- Fixed effect model is often unrealistic when heterogeneity is considerable and unexplained. But C.I is narrow compare with random effects.

- Random effects model analysis is suitable for unexplained heterogeneity. But it is difficult to interpret.

Sensitivity Analysis

- Sensitivity analysis is perform to evaluate the consistency or robustness of our results.
- sensitivity analysis was done by successively removing a particular study or group of studies (if any) that had the highest impact on the heterogeneity of pooled effect size.
- Sensitivity analyses are sometimes confused with subgroup analysis. Although some sensitivity analyses involve restricting the analysis to a subset of the totality of studies, the two methods differ in two ways.

Sensitivity Analysis



Meta regression

- The statistical purpose of meta-regression is to see to what extent covariates can explain the between-trial component of the variance.
- Covariates in meta regression should be independent.
- In meta-regression based on the linear regression equations, linear effects of the covariates were assessed on the between studies variance component.

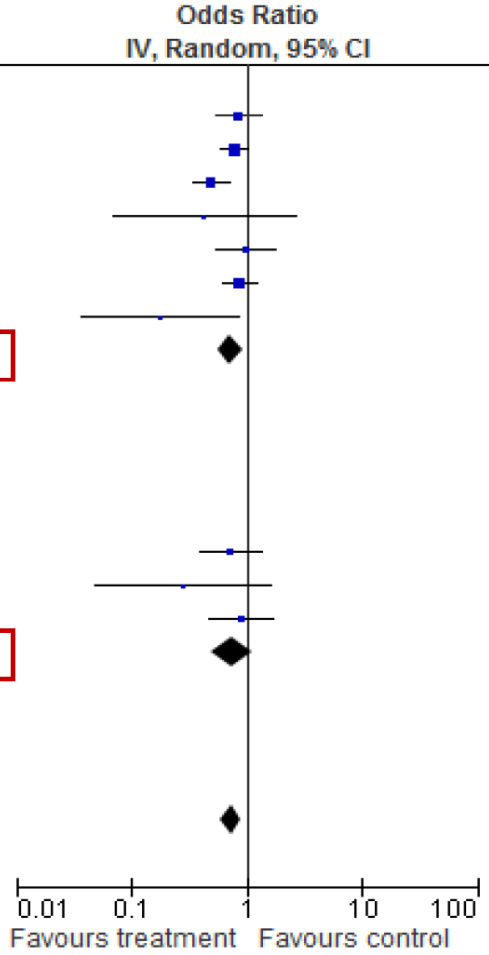
Meta regression & Subgroup analysis

- if you know which characteristics of studies may be associated with the size of effect. You can use subgroup analyses for qualitative characteristics for assessing between-trial component of the variance.

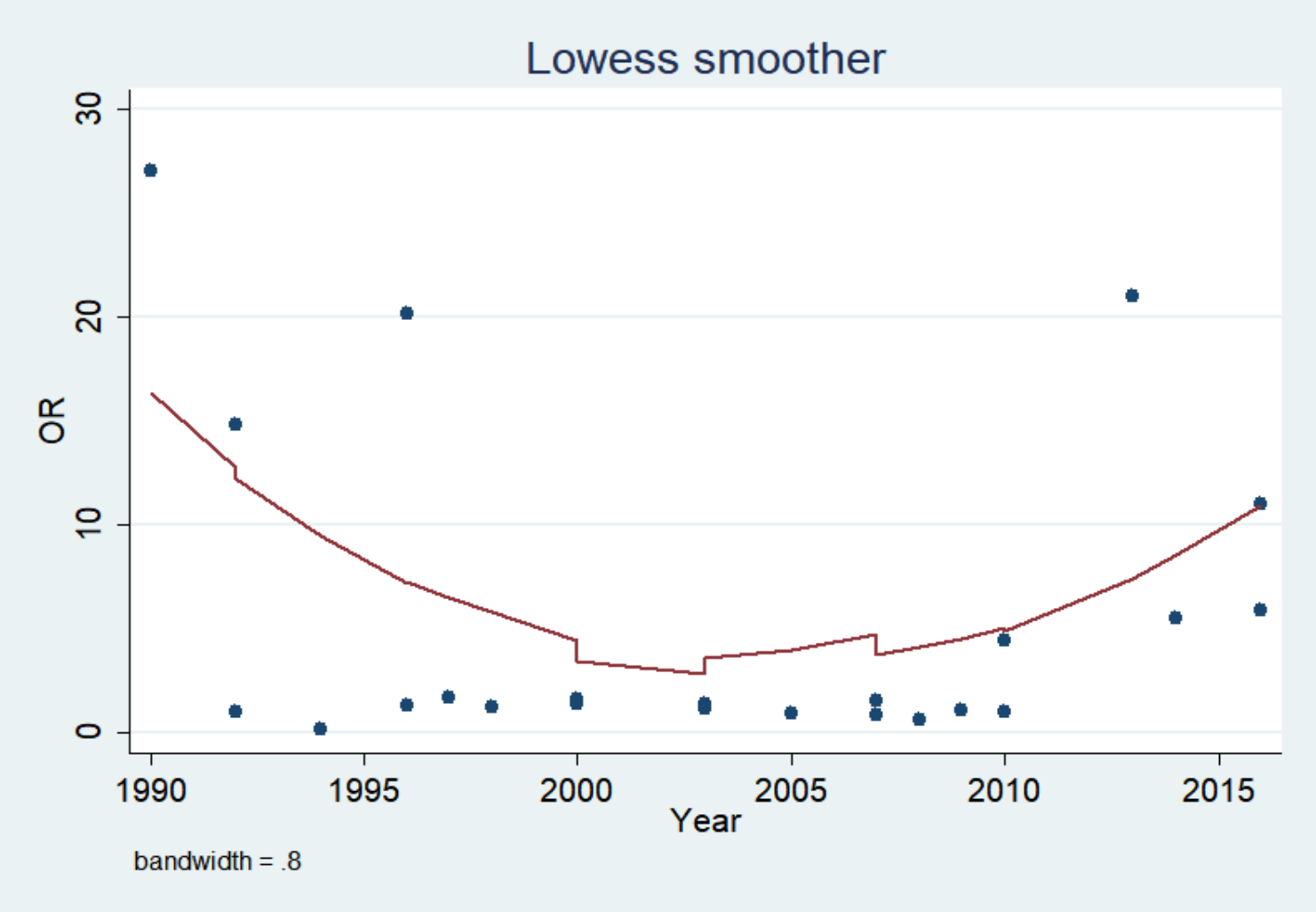
- Subgroup analysis involve splitting all of the studies into heterologous groups, often in order to make comparisons between them.

Subgroup analysis

Study or Subgroup	Vitamine K Events	Vitamine K Total	Control Events	Control Total	Weight	Odds Ratio IV, Random, 95% CI
1.1.1 Over 50s						
Bayes	48	183	54	183	12.7%	0.85 [0.54, 1.34]
Cochrane	125	624	152	631	23.3%	0.79 [0.60, 1.03]
Fisher	132	259	172	253	17.3%	0.49 [0.34, 0.70]
Gosset	3	10	5	10	1.1%	0.43 [0.07, 2.68]
Jeffreys	47	91	48	92	9.0%	0.98 [0.55, 1.75]
Markov	86	311	93	302	17.9%	0.86 [0.61, 1.22]
Pearson	3	18	9	17	1.5%	0.18 [0.04, 0.85]
Subtotal (95% CI)		**1496**		**1488**	**82.9%**	**0.72 [0.57, 0.92]**
Total events	444		533			
Heterogeneity: Tau² = 0.04; Chi² = 10.78, df = 6 (P = 0.10); I² = 44%						
Test for overall effect: Z = 2.58 (P = 0.010)						
1.1.2 Under 50s						
Hill	41	83	49	85	8.3%	0.72 [0.39, 1.32]
Wilks	5	11	9	12	1.2%	0.28 [0.05, 1.62]
Yates	24	94	27	97	7.6%	0.89 [0.47, 1.69]
Subtotal (95% CI)		**188**		**194**	**17.1%**	**0.75 [0.49, 1.15]**
Total events	70		85			
Heterogeneity: Tau² = 0.00; Chi² = 1.51, df = 2 (P = 0.47); I² = 0%						
Test for overall effect: Z = 1.34 (P = 0.18)						
Total (95% CI)		**1684**		**1682**	**100.0%**	**0.73 [0.60, 0.89]**
Total events	514		618			
Heterogeneity: Tau² = 0.02; Chi² = 12.29, df = 9 (P = 0.20); I² = 27%						
Test for overall effect: Z = 3.16 (P = 0.002)						
Test for subgroup differences: Chi² = 0.02, df = 1 (P = 0.90), I² = 0%						



Caution in Meta-regression



Cluster analysis & Subgroup analysis

- if you don't know which characteristics of studies may be associated with the size of effect. You can use cluster analysis based on the all of related and independent characteristics of studies in order to identify heterogeneous clusters.
- Finally, you can use subgroup analysis based one clustering running. Cluster analysis make comparisons between clusters automatically.
- Comparisons between clusters are expected to significantly differ in subgroup analysis.

Partitioning clustering algorithm

- K-means is simplest partitioning algorithm (MacQueen, 1967).

- This algorithm is the most commonly used unsupervised machine learning algorithm for partitioning a given data set into a set of k groups (i.e. k clusters), where k represents the number of groups pre-specified by the analyst.

Partitioning vs. Fuzzy clustering

- Unlike partitioning clustering methods (e.g. k-means), in fuzzy clustering methods (e.g. c-means) each observation has a set of membership coefficients or membership probabilities.
- The fuzzy clustering is considered as soft clustering and partitioning clustering is considered as hard or non-fuzzy clustering.
- Fuzzy clustering has many advantages compare with hard clustering such as flexibility and clustering noisy data samples. Very sensitive to good initialization can be considered as a main disadvantage of fuzzy clustering.

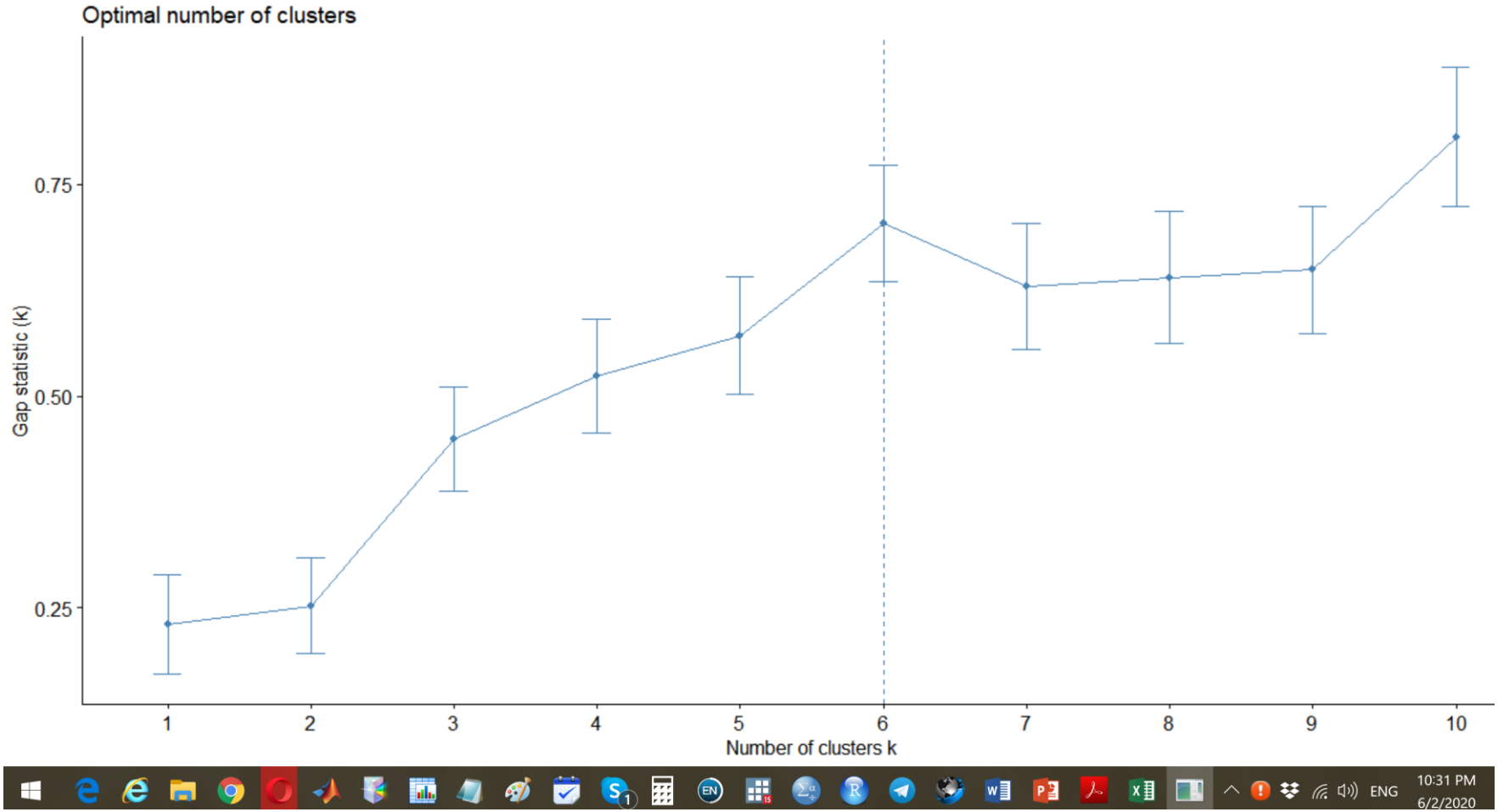
Example for k-means algorithm

ID	Study	Location	Year	Number of case	Number of control	Type of control	Case positive	Case negative	Control positive	Control negative	Sample	HPV detection method
1	McNicol et al.	Canada	1990	4	5	Healthy	4	0	1	4	Tissue	PCR
2	Ibrahim et al.	USA	1992	48	16	Healthy	6	42	2	14	Tissue	PCR
3	Anwar et al.	Japan	1992	68	10	Healthy	28	40	0	10	Tissue	PCR
4	Tu et al.	USA	1994	60	1	Healthy	3	57	0	1	Tissue	PCR
5	Wideroff et al.	USA	1996	56	42	Healthy	7	49	4	38	Tissue	PCR
6	suzuki et al.	Japan	1996	51	51	Healthy	8	43	0	51	Tissue	PCR
7	Terris et al.	USA	1997	73	37	Healthy	18	55	6	31	Tissue	PCR
8	Dillner et al.	Finland	1998	165	290	Healthy	40	125	60	230	Serum	PCR
9	Hisada et al.	USA	2000	48	63	Healthy	20	28	19	44	Serum	Elisa
10	Hayes et al.	USA	2000	276	295	Healthy	19	257	15	280	Serum	Elisa

For example: R code for k-means algorithm

```
install.packages("cluster")
install.packages("factoextra")
library(cluster)
library(factoextra)
data<-read.table("clipboard",h=TRUE,sep="\t")
data<-na.omit(data)
data<-scale(data)
fviz_nbclust(data, kmeans, method ="gap_stat")
km.res <- kmeans(data, 6, nstart = 25)
print(km.res)
```

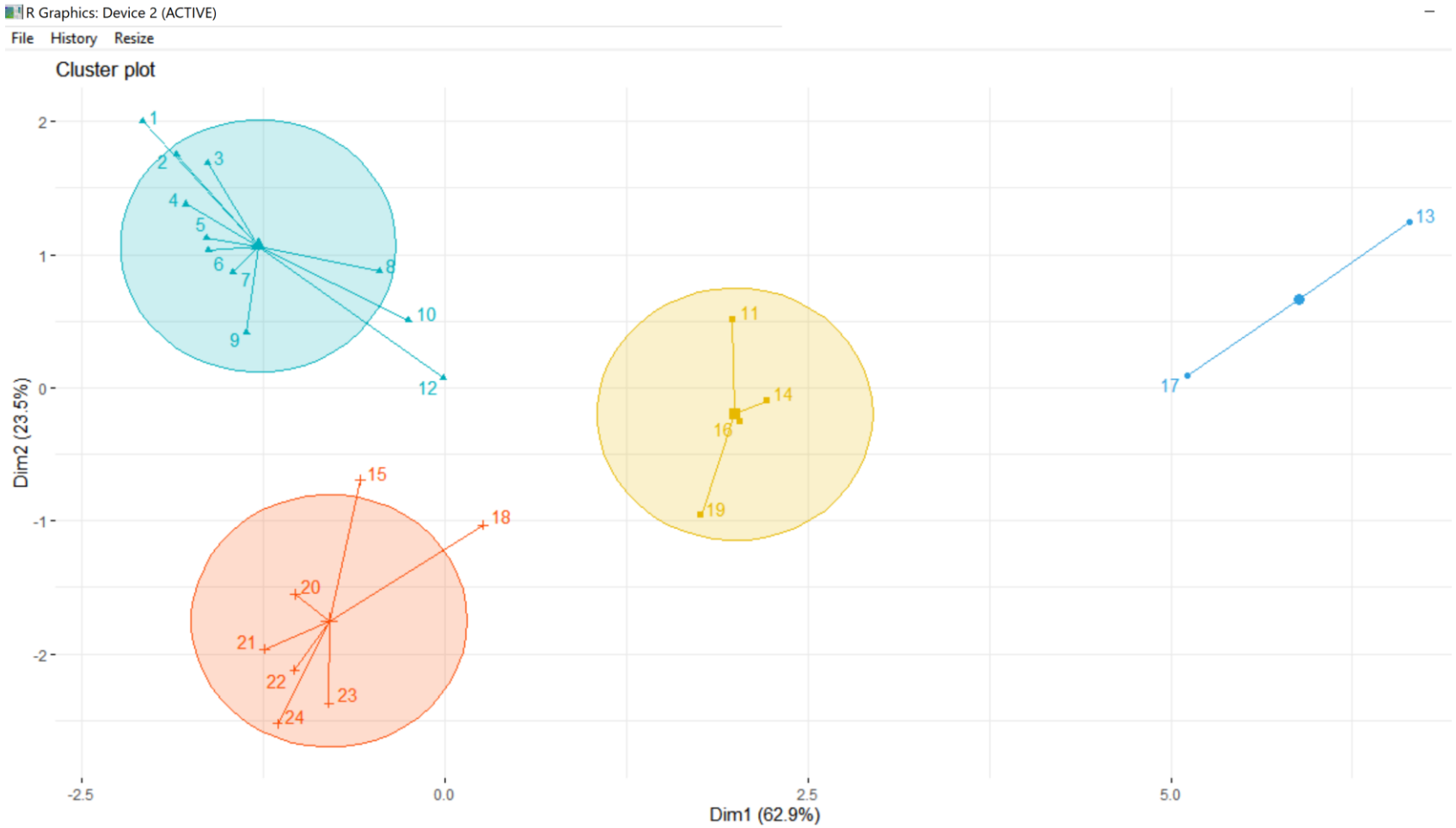
Example for k-means algorithm



Example for k-means algorithm

- ✓ K-means clustering with 6 clusters of sizes 5, 2, 1, 3, 5, 8
- ✓ Clustering vector:
- ✓ 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
- ✓ 6 6 6 6 6 6 6 1 6 1 4 1 2 4 1 3 2 1 4 5 5 5 5 5
- ✓ Within cluster sum of squares by cluster:
- ✓ [1] 4.5460675 6.9232220 0.0000000 3.3145160 0.8259981 2.5880018
- ✓ (between_SS / total_SS = 90.1 %) ????? I-square statistics

Example for k-means algorithm



Subgroup analysis after clustering

Test(s) of heterogeneity:

	Heterogeneity statistic	degrees of freedom	P	I-squared**	Tau-squared
6	10.23	7	0.176	31.6%	0.2915
1	0.45	3	0.929	0.0%	0.0000
4	0.96	2	0.619	0.0%	0.0000
2	4.42	1	0.035	77.4%	0.0448
3	0.00	0	.	.%	0.0448
5	1.41	4	0.842	0.0%	0.0000
Overall	56.74	22	0.000	61.2%	0.1104

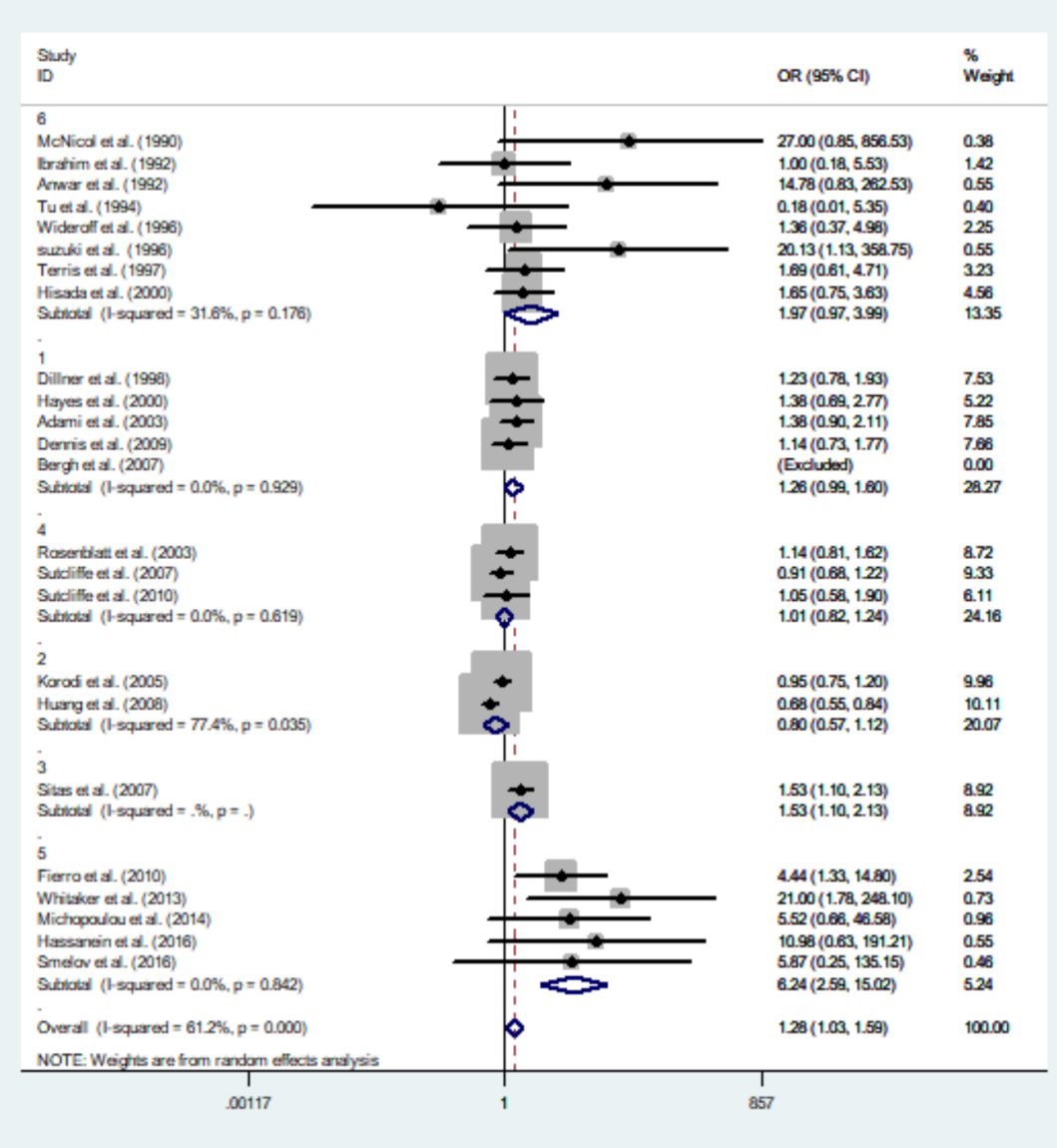
** I-squared: the variation in OR attributable to heterogeneity)

Note: between group heterogeneity not calculated;
only valid with inverse variance method

Significance test(s) of OR=1

6	z= 1.88	p = 0.060
1	z= 1.90	p = 0.057
4	z= 0.05	p = 0.957
2	z= 1.31	p = 0.190
3	z= 2.51	p = 0.012
5	z= 4.09	p = 0.000
Overall	z= 2.23	p = 0.026

Subgroup analysis after clustering



Applications of Bayesian Modeling in meta analysis

- ✓ For violation of the symmetry assumption by a skewed dis.
- ✓ For small number of individual studies by a prior dis.
- ✓ **For dealing with considerable heterogeneity by a random dis.**
- ✓ For using prior knowledge about true effect size
- ✓ For dealing with skewed data a skewed dis.
- ✓ For dealing with outlier data by a Laplace dis.
- ✓ For controlling false positive results by a wider 95% (Cr.I)

Why Bayesian Modeling?

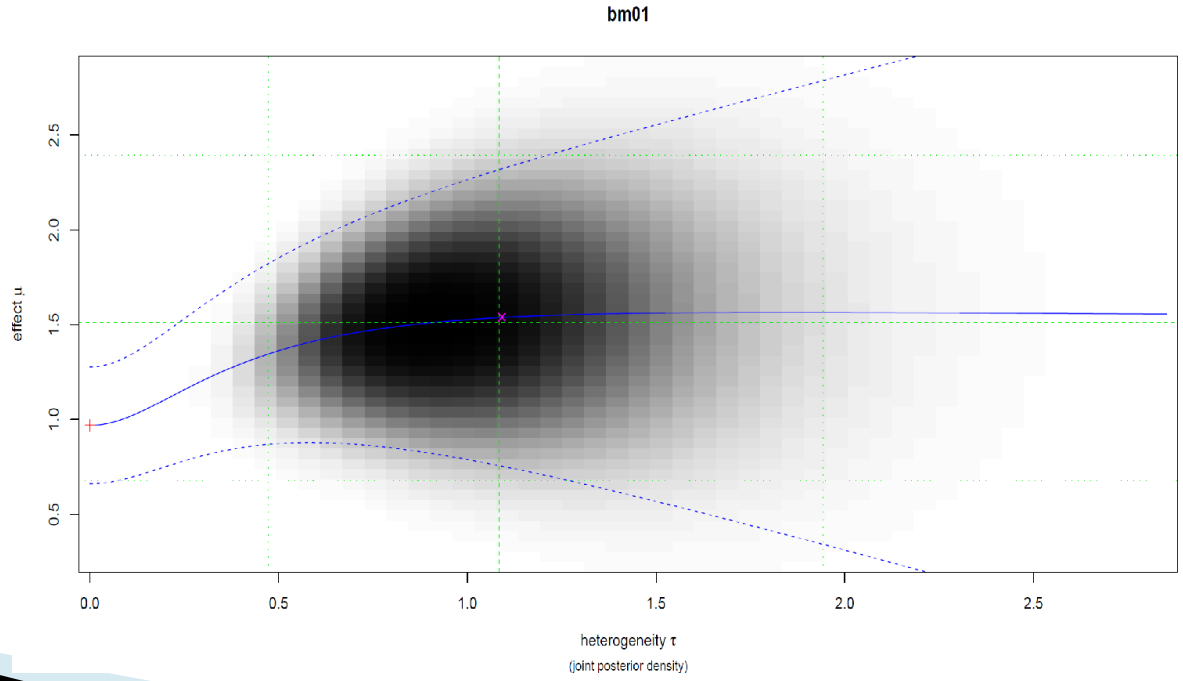
- Because Bayesian approach takes into account all sources of variation and reflects these variations in the pooled effect size.

For example: R code for Bayesian Modeling

- install.packages("bayesmeta")
- library(bayesmeta)
- OR <- escalc(measure="OR",ai=tpos, bi=tneg, ci=cpos, di=cneg, data=data)
- bm01 <- bayesmeta(y=OR,sigma=SE,labels =`First author` ,mu.prior.mean=mu, mu.prior.sd=10.0,tau.prior=function(x){ dinvgamma(x, 2+(1/10000),1.38*(1+(1/10000)))})
- print(bm01)
- > plot(bm01,which=1)
- > plot(bm01,which=2)
- > plot(bm01,which=3)
- > plot(bm01,which=4)
- > forestplot(bm01)

Example for Bayesian Modeling

- marginal posterior summary:
- tau mu
- mode 0.9757218 1.4906700
- median 1.0843731 1.5117491
- mean 1.1457243 1.5213724
- sd 0.3986335 0.4304754
- 95% lower 0.4723850 0.6793157
- 95% upper 1.9420972 2.3922495



Multivariate meta analysis

- ✓ In a meta-analysis clinical interest does not always concern only one specific outcome measure.

- ✓ Sometimes the focus is on the combination of several related outcome measures that are presented in the individual studies, for instance when there are more outcome variables. Related multiple outcome in the studies is one of the reasons of methodological heterogeneity. For dealing with related multiple related outcome, multivariate meta analysis can useful rather than univariate meta analysis.
- ✓ Unlike of multivariate meta analysis, network meta-analysis combines direct and indirect estimates across a network of interventions in a single analysis in order to comparison between interventions/or outcomes.

Multivariate meta analysis

- ✓ In a multivariate analysis all outcome measures are analyzed jointly, therefore also revealing information about the correlations between the multiple outcome variables.
- ✓ In multivariate meta analysis I^2_R and multivariate H^2 statistic are as the heterogeneity indices.
- ✓ "mvmeta" r package is available package to perform fixed and random-effects multivariate meta analysis (https://cran.r-project.org/web/packages/mvmeta/mvmeta.pdf) .

For example: R code for multivariate meta analysis

- install.packages("mvmeta")
- library(mvmeta)
- model<-
  mvmeta(cbind(Y1,Y2)~year,S=S,data=data,method="fixed")
- print(summary(model))
- summary(model)$coef
- coef(model,format="matrix")
- > forestplot(model)

Application of structural equation modeling in meta-analysis

- ✓ On one hand, related multiple outcome in the studies is one of reasons of methodological heterogeneity. In other hand, related outcomes can play a variety of roles in multiple regression equations. For this situation, structural equation modeling can useful.

- ✓ Please refer to "metaSEM" as a r package for meta-analysis using Structural Equation Modeling (https://cran.r-project.org/web/packages/metaSEM/vignettes/metaSEM.pdf)

meta-analysis modeling for skewed data

- ✓ As you know that all of traditional models in meta analysis are only correct asymptotically, while assume that the true effects are normally distributed.
- ✓ In practice, meta analysis models are frequently applied when study numbers are small and the normality of the effect distribution unknown or unlikely.
- ✓ If meta-analyses involve outcomes with skewed distributions, you can use skewed normal random effect model. However, log-transformed distribution can be useful.

meta-analysis modeling for skewed data

- ✓ "**altmeta**" R package provides alternative statistical methods for meta-analysis, including new heterogeneity tests and measures that are robust to outliers or skewed data; measures, tests, and visualization tools for publication bias; meta-analysis methods for synthesizing proportions; models for multivariate meta analysis, etc
- ✓ (https://cran.r-project.org/web/packages/altmeta/altmeta.pdf).

Some of the important functions in altmeta r package

- ✓ metaoutliers(y, s2, model) ## Calculates the standardized residual for each study
- ✓ mvma.Bayesian () ## Bayesian Random-Effects Multivariate Meta-Analysis
- ✓ mvma.hybrid () ## Performs a multivariate meta-analysis using the hybrid random-effects model when the within-study correlations are unknown
- ✓ mvma.hybrid.Bayesian() ## Bayesian Hybrid Model for Random-Effects Multivariate Meta Analysis

References

- Patsopoulos NA, Evangelou E, Ioannidis JP. Sensitivity of between-study heterogeneity in meta-analysis: proposed metrics and empirical evaluation. International journal of epidemiology. 2008 Oct 1;37(5):1148-57.
- Schroll JB, Moustgaard R, Gøtzsche PC. Dealing with substantial heterogeneity in Cochrane reviews. Cross-sectional study. BMC medical research methodology. 2011 Dec;11(1):22.
- Higgins JP, White IR, Anzures-Cabrera J. Meta-analysis of skewed data: combining results reported on log-transformed or raw scales. Statistics in medicine. 2008 Dec 20;27(29):6072-92.
- Kriston L. Dealing with clinical heterogeneity in meta-analysis. Assumptions, methods, interpretation. International journal of methods in psychiatric research. 2013 Mar;22(1):1-5.
- Arends L. Multivariate meta-analysis: modelling the heterogeneity mixing apples and oranges; dangerous or delicious?. 2006 Jun 28.

References

- Cheung MW. metaSEM: An R package for meta-analysis using structural equation modeling. Frontiers in Psychology. 2015 Jan 5;5:1521.
- Gasparrini A, Gasparrini MA. Package 'mvmeta'.
- Roever C, Friede T, Roever MC. Package 'bayesmeta'.
- Mohammadi FS, Aslani S, Mostafaei S, Jamshidi A, Riahi P, Mahmoudi M. Are genetic variations in IL-21-IL-23R-IL-17A cytokine axis involved in a pathogenic pathway of rheumatoid arthritis? Bayesian hierarchical meta-analysis. Journal of cellular physiology. 2019 Oct;234(10):17159-71.
- Mostafaei S, Keshavarz M, Nahand JS, Hassankiadeh RF, Moradinazar M, Nouri M, Babaei F, Ahadi M, Payandeh M, Esker AS, Hajighadimi S. Viral infections and risk of thyroid cancer: A systematic review and empirical bayesian meta-analysis. Pathology-Research and Practice. 2020 Feb 13:152855.