High-throughput Sequencing Analysis

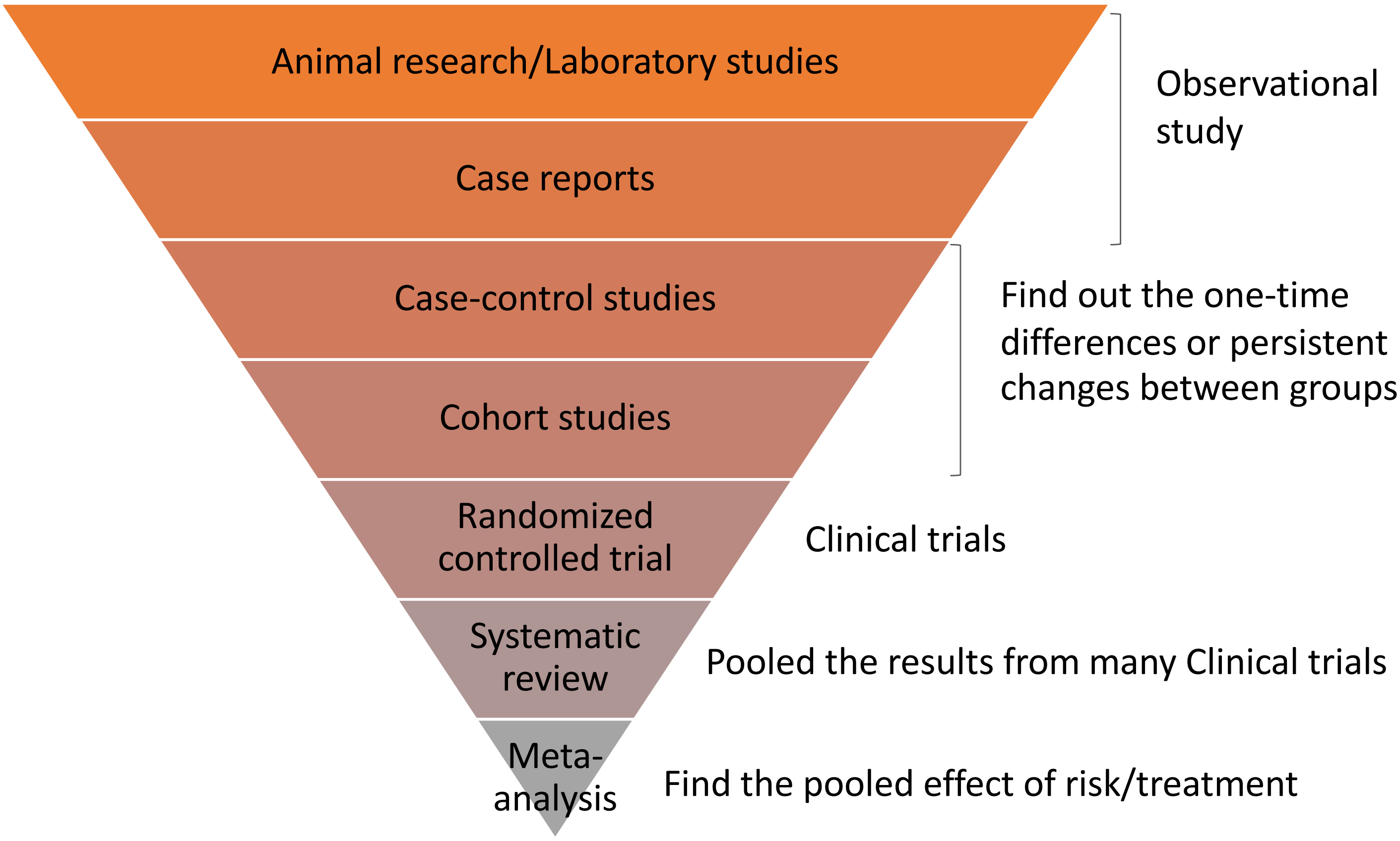
Meta-Analysis

Pei-Chien Tsai

2019-05-31

Meta-analysis

- Meta-analysis is a quantitative and epidemiological study design to conduct a systematic reviews of previous research studies.
- The outcome of meta-analysis represents the ‘pooled’ results of the studies and therefore could provide a more precise estimate of the effect of risk (or treatment) for disease.



- Compare to a single study, meta-analysis
  - has greater study power (bigger sample size, greater diversity)
  - provides more reliable results
  - helps to find out the effects from studies that have conflicting results
  - helps to discover the effects across different populations
  - can pool the results from studies with smaller sample size that are not statistically significant

Meta-analysis

- The limitations of meta-analysis including
  - Time consuming and difficult to find appropriate studies
  - Details provided by each study might not be adequate for analysis
  - Requires advanced statistical analysis
  - There might be a publication bias – only publications with positive/significant findings were reported

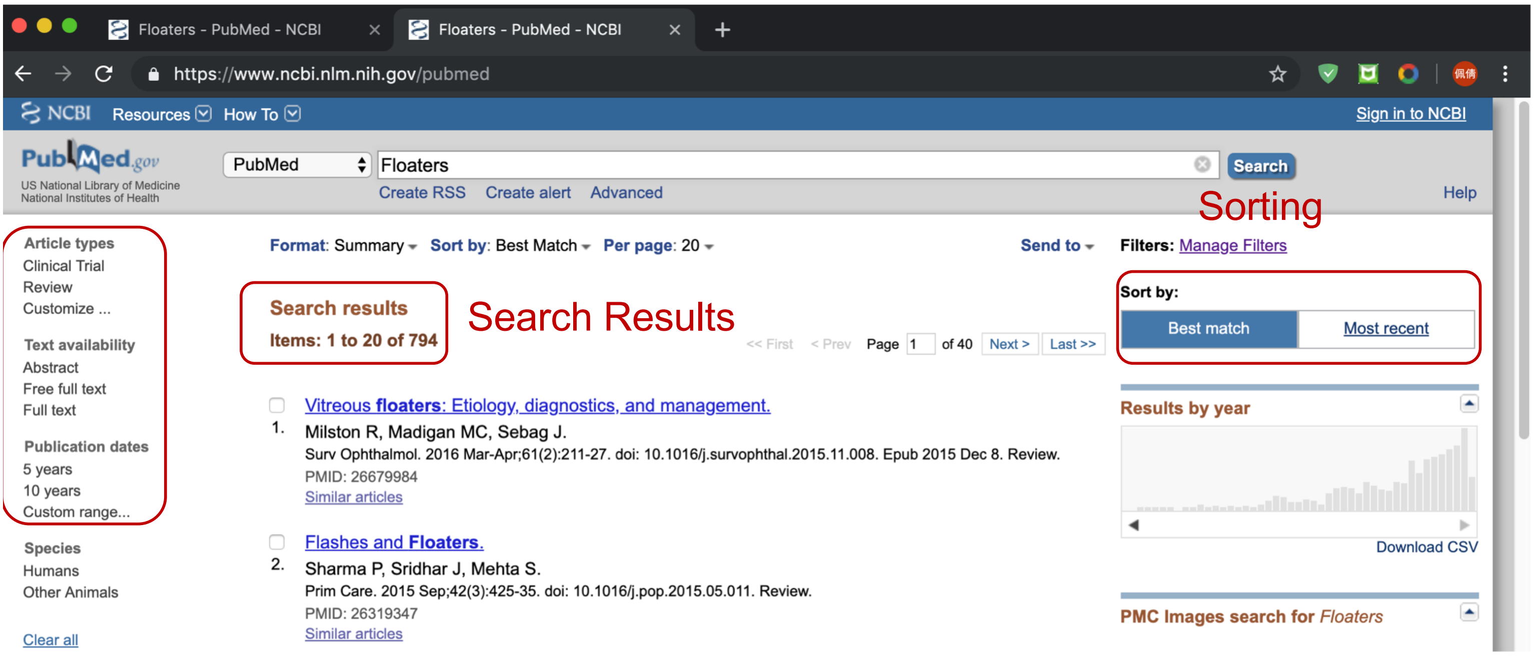
- To perform the meta-analysis, the first step is to perform a systematic review.
- Usually we aim to find as many publications as possible, and prefer studies with bigger sample size.
- Ideally, the selected studies shall be
  - Studies with large diversity
  - Study results were generated using similar statistical methods (and adjust for similar covariates)
  - Studies with same "effects".
  - Studies with both positive/negative, significant/insignificant results
  - Studies with comparable information
  - For genetic studies, sometimes we need to consider the 'molecular meanings' before analysis

- The "effects" can be used for meta-analysis including:
  - Mean, mean difference
  - Risk ratio, odds ratio
  - Proportion, rate change
  - Coefficients (beta coefficients, correlation coefficients)

Searching the literature with Pubmed

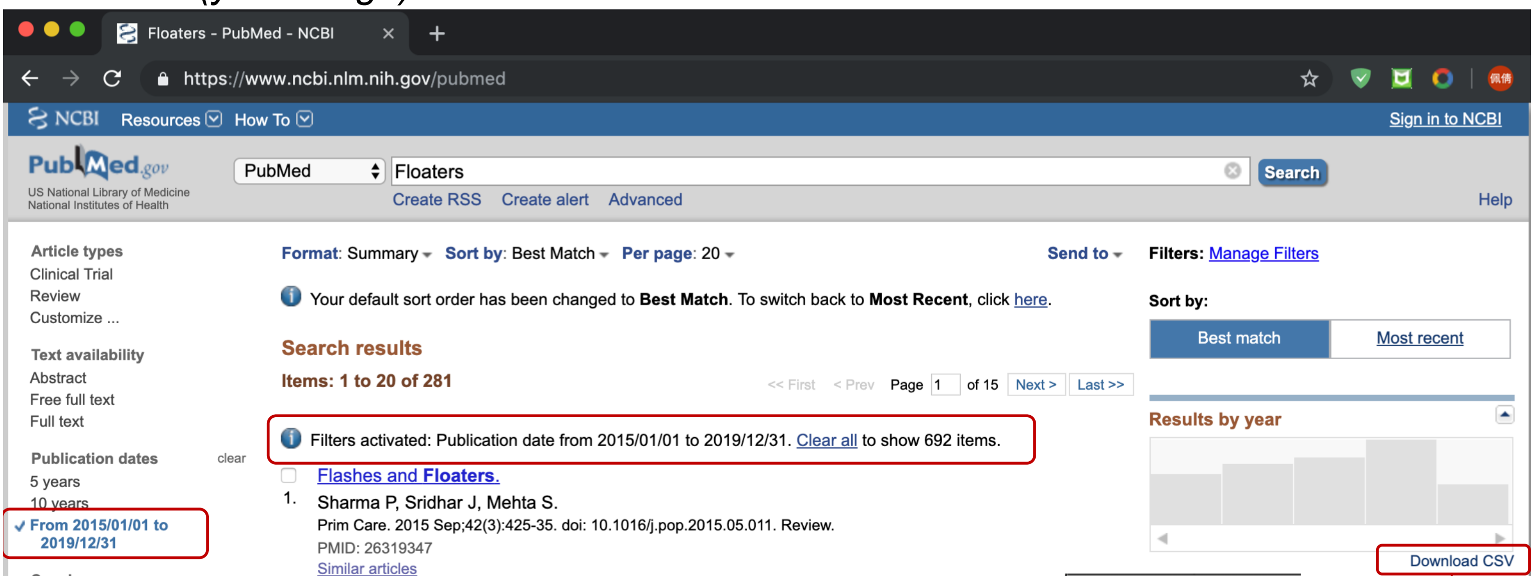
- What do you usually do for literature reviewing? Pubmed?

No filter

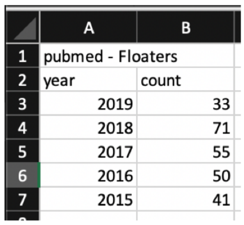


Different filters

With filter (year range)



Year filter



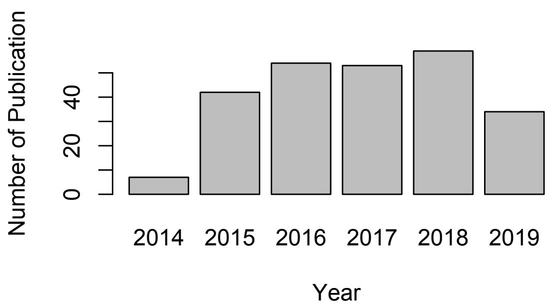
	A	B
1	pubmed - Floaters	
2	year	count
3	2019	33
4	2018	71
5	2017	55
6	2016	50
7	2015	41

- Can we do text mining easily using Pubmed?
- Probably not.

Searching the literature with R

- Alternatively, we can use a package “RISmed” to make things easier

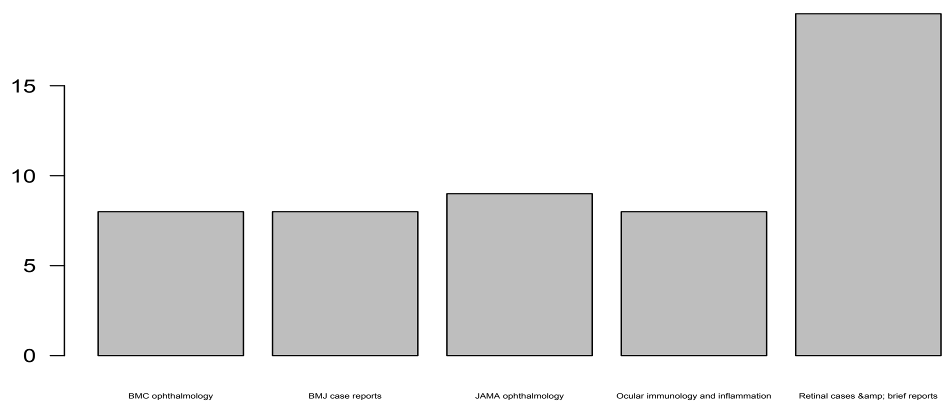
```
# Install and use the package
> install.packages("RISmed")
> library(RISmed)

# Let’s see how it works
> fit <- EUtilsSummary("Floaters", type="esearch", db="pubmed",
datetype='pdat', mindate=2015, maxdate=2019)
# "Floaters": this is the keyword you are interested in. Try to
replace it with other character
# db: database, we search publications on pubmed
# mindate: publication year (min)
# maxdate: publication year (max)

# Check how many references you found
> fetch <- EUtilsGet(fit)
> QueryCount(fit)
[1] 250

# Check how it distributed during 2015 and 2019
> y <- YearPubmed(fetch)
> table(y)
2014 2015 2016 2017 2018 2019
   7   42   54   53   59   34
```




```
# The results are ‘slightly’ different
  from what we’ve got using Pubmed…

> barplot(table(y),ylab="Number of Publication",xlab="Year")
```

Searching the literature with R

```
# We can get all kind of information from these publications
> alllist <- EUtilsGet(fit)
> slotNames(alllist)
[1] "Query"                "PMID"                    "YearReceived"
[4] "MonthReceived"        "DayReceived"             "HourReceived”
…
[58] "CollectiveName"       "Mesh”

# Alternatively, just select the information you are interested in
> data <- data.frame(pmid=PMID(fetch),year=YearPubmed(fetch),
title=ArticleTitle(fetch),journal=Title(fetch),country=Country(fetch))

# Check the dimension now, we got 5 columns listing the pmid, year,
title, journal, and country information
> dim(data)
[1] 249   5

# You can get some interesting information from the fetch results
# For example, check which journal potentially wants to take a related
articles in the future
> journal1 <- table(data$journal)
> barplot(journal1[journal1>7],horiz=F,las=1.5,cex.names=0.4)
```




```
# Or check who are interested in this topic
> sort(table(data$country),decreasing=T)

     United States              England          Switzerland
               112                   70                   21
```

Searching the literature with R

```
# And next, let’s check if there is a popular word (or words) in the
publication title
# Let’s check with the first publication, which list on the first row of your
data
> a<- data$title[1]
> a
[1] "Ocular and orbital side effects of ALK inhibitors: a review article”

# We can remove all the punctuations at the beginning
> a <- gsub("[:]","",a)
> a <- gsub("[;]","",a)
> a <- gsub("[.]","",a)
# Now, split the string into single characters
> a1 <- strsplit(as.character(a)," ")
> a1
[[1]]
 [1] "Ocular"     "and"         "orbital"    "side"        "effects"
 [6] "of"         "ALK"         "inhibitors" "a"           "review"
[11] "article”

# Let’s try to do it altogether
> a <- data$title
> a <- gsub("[:]","",a)
> a <- gsub("[;]","",a)
> a <- gsub("[.]","",a)
> a1 <- strsplit(as.character(a)," ")

# a1 is a “list”
# You need to ”unlist” it before counting it

> a2 <- unlist(a1)
> sort(table(a2),decreasing=T)[1:20]
a2
        of        and         in        the          a        for       with
       159         96         92         59         55         49         40
         A   Floaters   floaters   Vitreous   vitreous         to         an
        36         28         27         27         24         23         22
   Retinal Vitrectomy      after       case    retinal         as
        20         14         13         13         13         12
```

Example of Meta-analysis: OR

```
# After fetching the publications of interests, you can select the
publications for the meta-analysis
# Let's use an simple example here, please download R package "meta"
> install.packages("meta")
> library(meta)

# We will use a dataset 'Fleiss93' in meta package
# The use of aspirin in preventing death after myocardial infarction
> data(Fleiss93)
> Fleiss93
   study year event.e  n.e event.c  n.c
1  MRC-1 1974      49  615      67  624
2    CDP 1976      44  758      64  771
3  MRC-2 1979     102  832     126  850
4   GASP 1979      32  317      38  309
5  PARIS 1980      85  810      52  406
6   AMIS 1980     246 2267     219 2257
7 ISIS-2 1988    1570 8587    1720 8600
```

	Aspirin	Placebo
Death #	49	67
Alive #	566	557

```
# Study: study label
# year: year of publication
# event.e: number of deaths in aspirin group
# n.e: number of observations in aspirin group
# event.c: number of deaths in placebo group
# n.c: number of observations in placebo group

# Basically we will compare the odds ratio for each study
```

	Aspirin	Placebo
Death number		
Alive number		
Total number		

Example of Meta-analysis: outcome

```
# metabin function helps you to calculate odds ratio first
> result <- metabin(event.e, n.e,event.c,n.c,data=Fleiss93,sm="OR")
> result
```

1.

```
      OR            95%-CI %W(fixed) %W(random)
1 0.7197 [0.4890; 1.0593]       3.2        8.2
2 0.6808 [0.4574; 1.0132]       3.1        7.8
3 0.8029 [0.6065; 1.0629]       5.7       13.2
4 0.8007 [0.4863; 1.3186]       1.8        5.4
5 0.7981 [0.5526; 1.1529]       3.2        8.9
6 1.1327 [0.9347; 1.3728]      10.2       20.7
7 0.8950 [0.8294; 0.9657]      72.9       35.8
```

Weight of individual studies

- Fixed effect
  - Same mean, zero bt study variance
- Random effect
  - Different mean, with bt study variance

```
Number of studies combined: k = 7
```

4.

```
                         OR           95%-CI     z p-value
Fixed effect model   0.8969 [0.8405; 0.9570] -3.29  0.0010
Random effects model 0.8763 [0.7743; 0.9917] -2.09  0.0365
```

```
Quantifying heterogeneity:
tau^2 = 0.0096; H = 1.29 [1.00; 1.99]; I^2 = 39.7% [0.0%; 74.6%]
```

variance of the true effect sizes

```
Test of heterogeneity:
    Q d.f. p-value
 9.95    6  0.1269
```

2.

3.

I^2 ranged between 0% and 100%

- <25% : low
- 25%-75%: middle
- >75%: high

- If P<0.1, there is a significant heterogeneity problem (that means the studies have great variability)
- You need to find out the reasons for heterogeneity, and use random-effect result instead.

```
Details on meta-analytical method:
- Mantel-Haenszel method
- DerSimonian-Laird estimator for tau^2
```

Example of Meta-analysis: forest plot

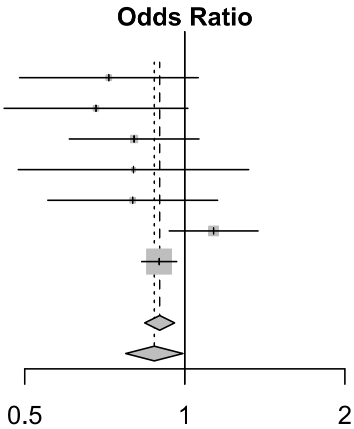
```
# Now we can generate the famous forest plot to visual our result
> forest(result)
```

We can see Aspirin seems to help preventing death after myocardial infarction in most of the studies

Therefore, the overall effect is significant (but most likely, the results were highly influenced by the study result of Study 7.

Study	Experimental Events	Experimental Total	Control Events	Control Total	OR	95%-CI	Weight (fixed)	Weight (random)
1	49	615	67	624	0.72	[0.49; 1.06]	3.2%	8.2%
2	44	758	64	771	0.68	[0.46; 1.01]	3.1%	7.8%
3	102	832	126	850	0.80	[0.61; 1.06]	5.7%	13.2%
4	32	317	38	309	0.80	[0.49; 1.32]	1.8%	5.4%
5	85	810	52	406	0.80	[0.55; 1.15]	3.2%	8.9%
6	246	2267	219	2257	1.13	[0.93; 1.37]	10.2%	20.7%
7	1570	8587	1720	8600	0.89	[0.83; 0.97]	72.9%	35.8%
Fixed effect model		**14186**		**13817**	**0.90**	**[0.84; 0.96]**	**100.0%**	**--**
Random effects model					**0.88**	**[0.77; 0.99]**	**--**	**100.0%**

Heterogeneity: $I^2 = 40\%$, $\tau^2 = 0.0096$, $p = 0.13$



For the next step, let's consider the potential publication bias in this meta-analysis

We will use 'Funnel Plot' to visual the bias

A funnel plot is a scatter plot of intervention effect estimates against a measure of study precision.

Asymmetry in the funnel plot suggest there might be a publication bias for the study

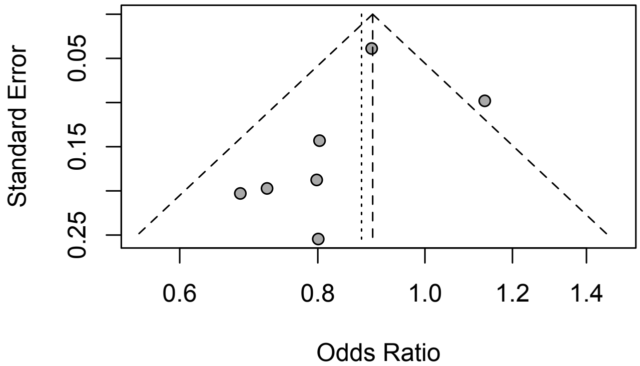
The bias could be due to

- Reporting bias or selective reporting
- Poor methodological quality (e.g. sample size too small)
- Heterogeneity among studies
- Simply by chance

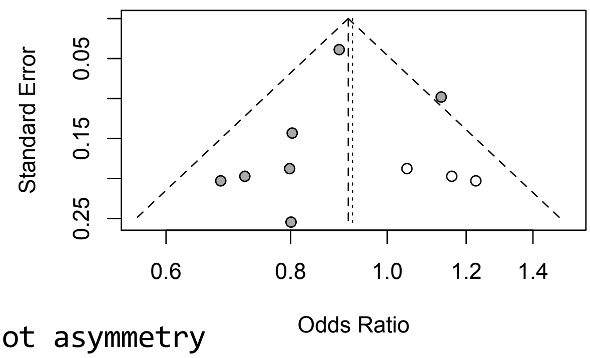
When you have studies < 10, funnel plot is not suggested for judging the publication bias (lack of power)

Example of Meta-analysis: asymmetry analysis

```
# Asymmetry analysis
> funnel(result)
```

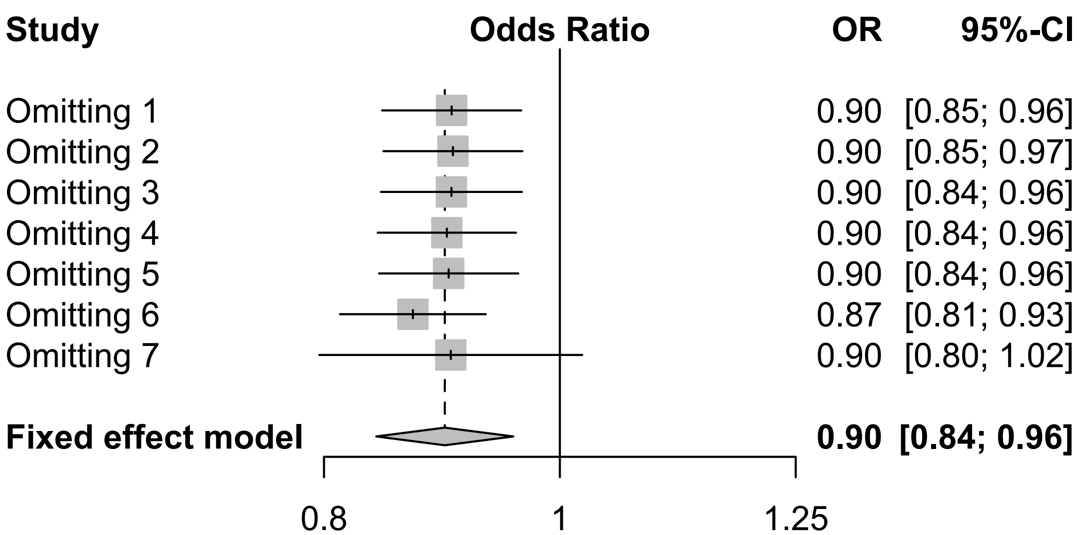



```
# If you have studies > 10, you can use this function to test your
publication bias
> metabias(result,method.bias="peters")
Warning message:
In print.metabias(x) :
  Number of studies (k=7) too small to test for small study effects
(k.min=10). Change argument 'k.min' if appropriate.

# We do see Asymmetry in the funnel plot, we can adjust our results using
either trim and filled or copas model.

> tf1 <- trimfill(result, comb.fixed=TRUE)
> summary(tf1)
Number of studies combined: k = 10 (with 3 added studies)

                        OR           95%-CI     z p-value
Fixed effect model   0.9140 [0.8587; 0.9727] -2.83  0.0047
Random effects model 0.9231 [0.8252; 1.0327] -1.40  0.1622

Quantifying heterogeneity:
tau^2 = 0.0102; H = 1.26 [1.00; 1.83]; I^2 = 37.4% [0.0%; 70.1%]

Test of heterogeneity:
     Q d.f. p-value
 14.37    9  0.1099

Details on meta-analytical method:
- Inverse variance method
- DerSimonian-Laird estimator for tau^2


- Trim-and-fill method to adjust for funnel plot asymmetry
> funnel(tf1)
```



Example of Meta-analysis: sensitivity test

```
# We can exclude 1 study a time to check the meta-analysis results each time
> metainf(result, pooled="fixed")

Influential analysis (Fixed effect model)

                        OR            95%-CI  p-value   tau^2    I^2
Omitting 1          0.9027 [0.8452; 0.9641]    0.0023  0.0099  42.3%
Omitting 2          0.9038 [0.8462; 0.9652]    0.0026  0.0082  37.9%
Omitting 3          0.9025 [0.8443; 0.9648]    0.0026  0.0129  46.3%
Omitting 4          0.8986 [0.8417; 0.9594]    0.0014  0.0123  48.7%
Omitting 5          0.9001 [0.8427; 0.9615]    0.0018  0.0124  47.6%
Omitting 6          0.8702 [0.8122; 0.9324]  < 0.0001  0.0000   0.0%
Omitting 7          0.9020 [0.7965; 1.0214]    0.1040  0.0268  49.7%

Pooled estimate     0.8969 [0.8405; 0.9570]    0.0010  0.0096  39.7%

Details on meta-analytical method:
- Mantel-Haenszel method
> forest(metainf(result), comb.fixed=TRUE)
```



- The result again suggest that the overall results were highly influenced by the study result of Study 7.