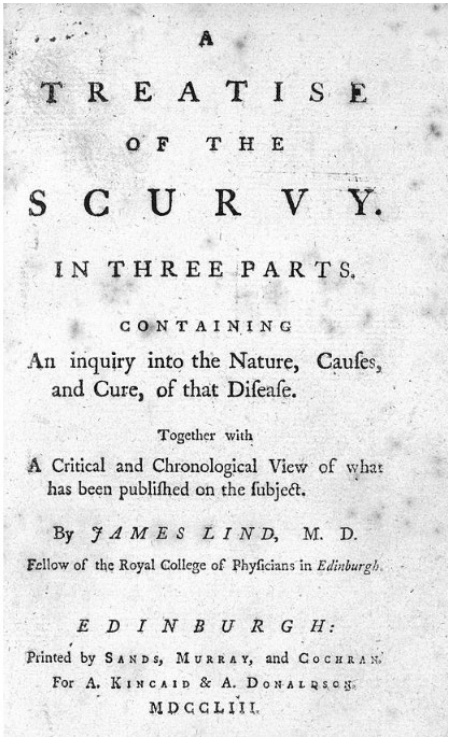
臺中榮民總醫院
Taichung Veterans General Hospital

臨床試驗中Risk of Bias的評估與統合分析技術應用

臺中榮民總醫院 醫學研究部 基礎醫學科
高級助理研究員 陳韻伃博士

百年前的「臨床試驗」

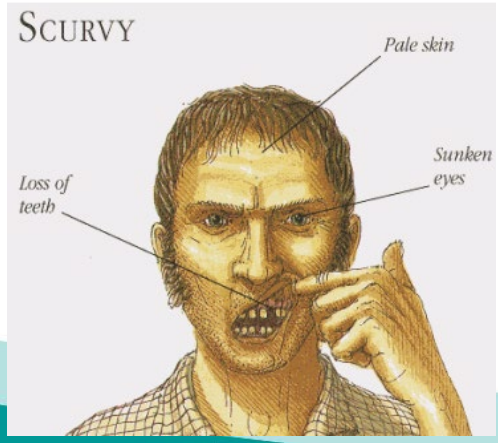
- 詹姆斯·林德（James Lind，1716年 - 1794年 6月13日）
  - 英國皇家海軍外科醫生（1739 年 - 1748年），英格蘭衛生學的創始人
  - 發起利用柑桔類水果和新鮮蔬菜治療和預防壞血病
  - A treatise of the scurvy 壞血病論


A
TREATISE
OF THE
SCURVY.
IN THREE PARTS.
CONTAINING
An inquiry into the Nature, Cauſes, and Cure, of that Diſeaſe.
Together with
A Critical and Chronological View of what has been publiſhed on the ſubject.
By JAMES LIND, M. D.
Fellow of the Royal College of Phyſicians in Edinburgh.
EDINBURGH:
Printed by SANDS, MURRAY, and COCHRAN.
For A. KINCAID & A. DONALDSON.
MDCCLIII.

TCVGH

百年前的「臨床試驗」

- 百年前的歐洲，長期在海上航行的水手經常遭受壞血病的折磨，患者常常牙齦出血，甚至皮膚淤血和滲血，最後痛苦地死去，人們一直查不出病因。奇怪的是，只 要船隻靠岸，這種疾病很快就不治而癒了。
- 問題：水手們為什麼會得壞血病呢？
- 書中提到他在1747年在船上做了一個臨床試驗：
  - 出現壞血病的船員，大家都吃完全相同的食物
  - 唯一不同的是有些病人每天吃兩個橘子和一個檸檬，其他的人喝蘋果酒、稀硫酸、醋、海水。
- 實驗的結論：吃柑橘水果的兩人好轉，其它人病情依然。



TCVGH

偏差?

- 從現代的觀點看，林德的臨床試驗不夠嚴謹：
  - 病人的分派 Allocation
  - 每一組的病人數 Sample size
  - 臨床指標 Clinical indication / Outcome
  - 統計分析 Statistical analysis

“Bias” v.s. “Risk of Bias”

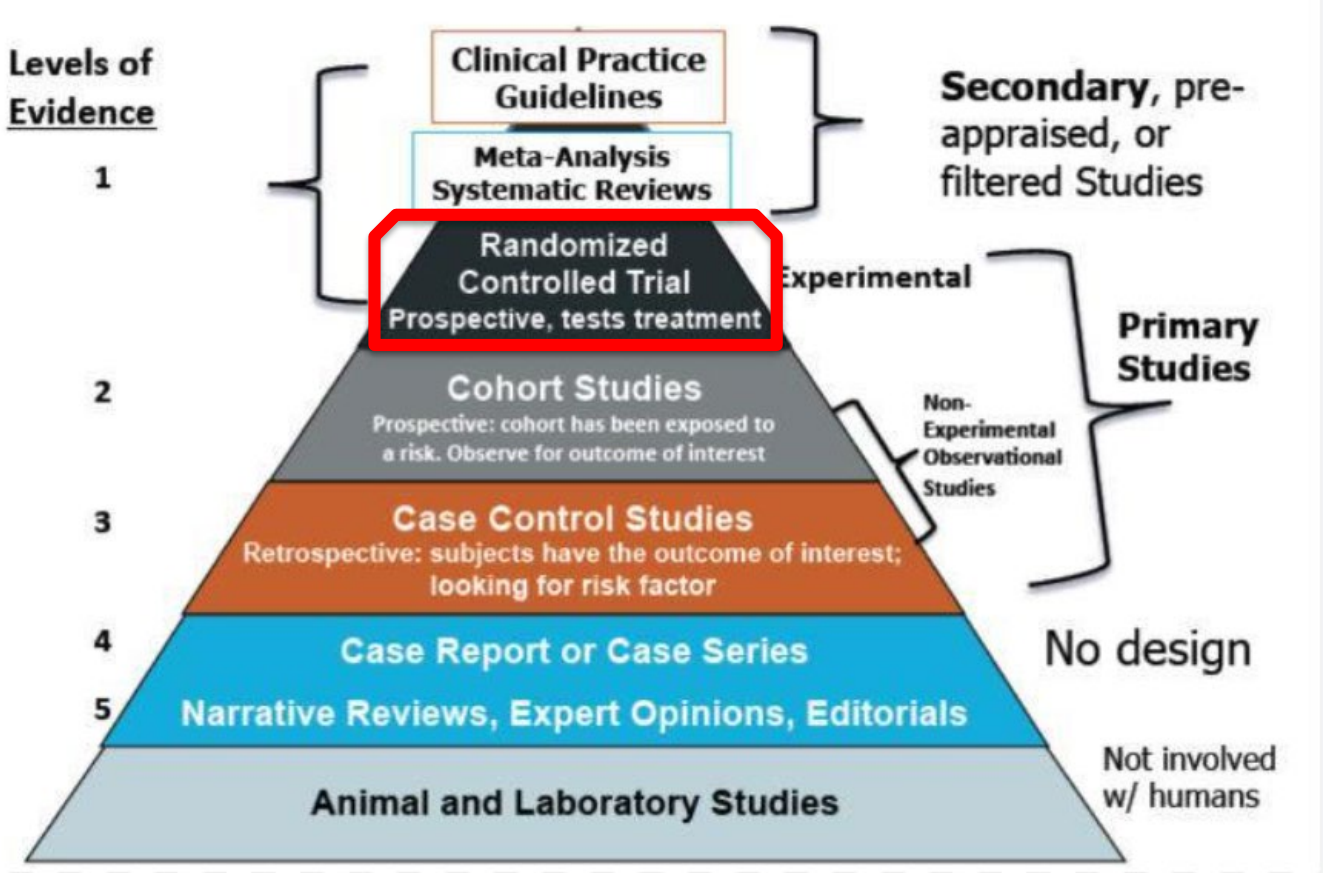
抽樣所造成測量值與真實值之間的差異，統計可處理：
通過多次測量或增加樣本數，獲得的均值儘量逼近

- Error – Systematic error v.s. Random error
- Bias – Systematic error
  - Deviation of study result from the truth，不能靠統計處理
  - 測量時即發生錯誤 (內因)：Information bias, recall bias, report bias
  - 外因：Confounding bias ~ Confounding factors
- Risk of Bias
  - In fact, we never know the truth
  - The results from a study might be unbiased despite methodological flaws
    - ✓E.g., poor randomization or lost to follow up, but unbiased results

實證醫學的證據等級

- 文獻的證據等級與研究設計相關

證據金字塔 → 隨機對照試驗 (RCT)：
Level 1 (Gold Standard)



TCVGH

實證醫學的證據等級

- 文獻的證據等級與研究設計相關
- 證據的等級：良好研究設計可以減少偏差的程度→**隨機對照試驗 (RCT)**

表一　Oxford證據等級與建議等級[6, 9]

建議等級	證據等級	證據的型態
〔A〕	1a	同質性隨機對照試驗的系統性回顧
	1b	單獨的隨機對照試驗
	1c	如果沒有給藥的全部病人會死，給藥後會有一些病人存活；或是如果沒有給藥會有一些病人死亡，而給藥後就不會有病人死亡。
〔B〕	2a	同質性世代研究的系統性文獻回顧
	2b	單獨的世代研究
	2c	結果研究或生態研究
	3a	同質性個案研究的系統性文獻回顧
	3b	單獨的個案對照研究
〔C〕	4	個案發現報告或是品質較差的世代研究和個案對照研究
〔D〕	5	未經清楚且嚴謹的專家意見

為什麼要進行Meta-analysis?

統合多個臨床研究的樣本數和結果，證據力高
花費研究經費和人力成本相對低

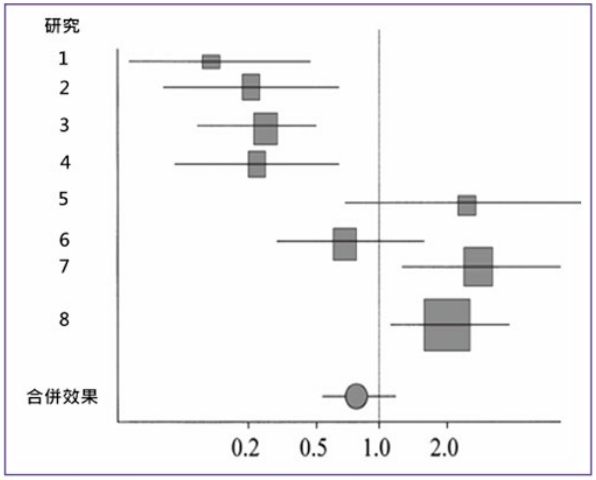
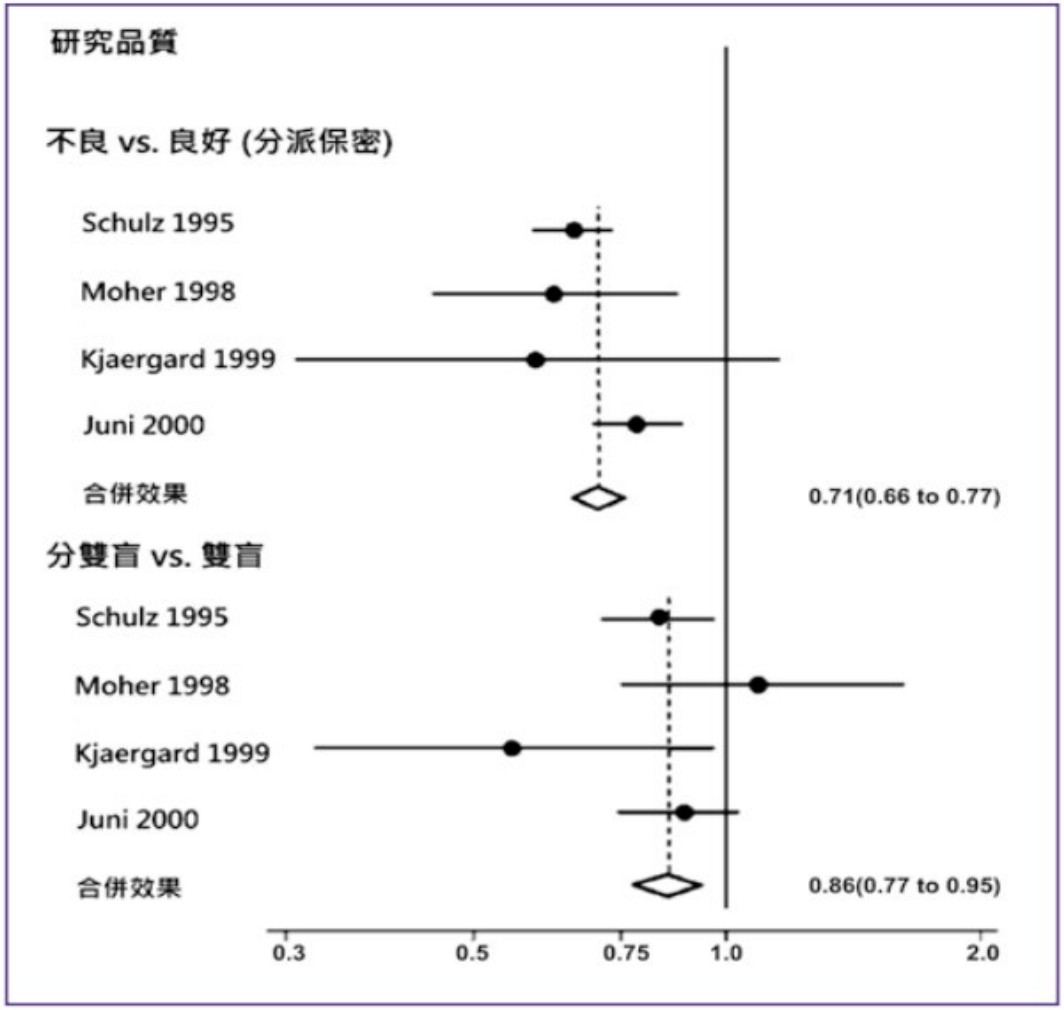


圖2　統合分析中呈現不同研究結果的明顯差異性

統合分析的研究品質對分析結果的影響



TCVGH

使用RCT研究進行統合分析→確認研究品質

Domain-based approach in Cochrane's "Risk of bias" (ROB) tool (for RCT)

- Sequence generation
- Allocation concealment
- Blinding of participants, personnel and outcome assessors
- Incomplete outcome data
- Selective outcome reporting
- Other sources of bias

由兩個人**double check**

TCVGH

使用RCT研究進行統合分析→確認研究品質

Domain-based approach in Cochrane's "Risk of bias" (ROB) tool (for RCT)

Domain	Signalling question	Response
Bias arising from the randomization process	1.1 Was the allocation sequence random?	Y
	1.2 Was the allocation sequence concealed until participants were enrolled and assigned to interventions?	Y
	1.3 Did baseline differences between intervention groups suggest a problem with the randomization process?	Y
	Risk of bias judgement	**Some concerns**

TCVGH

使用RCT研究進行統合分析→確認研究品質

Domain-based approach in Cochrane's "Risk of bias" (ROB) tool (for RCT)

Domain	Signalling question	Response
Bias due to deviations from intended interventions	2.1.Were participants aware of their assigned intervention during the trial?	N
	2.2.Were carers and people delivering the interventions aware of participants' assigned intervention during the trial?	N
	2.3. If Y/PY/NI to 2.1 or 2.2: Were there deviations from the intended intervention that arose because of the experimental context?	NA
	2.4 If Y/PY to 2.3: Were these deviations likely to have affected the outcome?	NA
	2.5. If Y/PY/NI to 2.4: Were these deviations from intended intervention balanced between groups?	NA
	2.6 Was an appropriate analysis used to estimate the effect of assignment to intervention?	NA
	2.7 If N/PN/NI to 2.6: Was there potential for a substantial impact (on the result) of the failure to analyse participants in the group to which they were randomized?	N
	Risk of bias judgement	**Low**

使用RCT研究進行統合分析→確認研究品質

Domain-based approach in Cochrane's "Risk of bias" (ROB) tool (for RCT)

Domain	Signalling question	Response
Bias due to missing outcome data	3.1 Were data for this outcome available for all, or nearly all, participants randomized?	Y
	3.2 If N/PN/NI to 3.1: Is there evidence that result was not biased by missing outcome data?	NA
	3.3 If N/PN to 3.2: Could missingness in the outcome depend on its true value?	NA
	3.4 If Y/PY/NI to 3.3: Is it likely that missingness in the outcome depended on its true value?	NA
	Risk of bias judgement	**Low**

TCVGH

使用RCT研究進行統合分析→確認研究品質

Domain-based approach in Cochrane's "Risk of bias" (ROB) tool (for RCT)

Domain	Signalling question	Response
Bias in measurement of the outcome	4.1 Was the method of measuring the outcome inappropriate?	N
	4.2 Could measurement or ascertainment of the outcome have differed between intervention groups?	N
	4.3 Were outcome assessors aware of the intervention received by study participants?	N
	4.4 If Y/PY/NI to 4.3: Could assessment of the outcome have been influenced by knowledge of intervention received?	NA
	4.5 If Y/PY/NI to 4.4: Is it likely that assessment of the outcome was influenced by knowledge of intervention received?	NA
	Risk of bias judgement	**Low**

使用RCT研究進行統合分析→確認研究品質

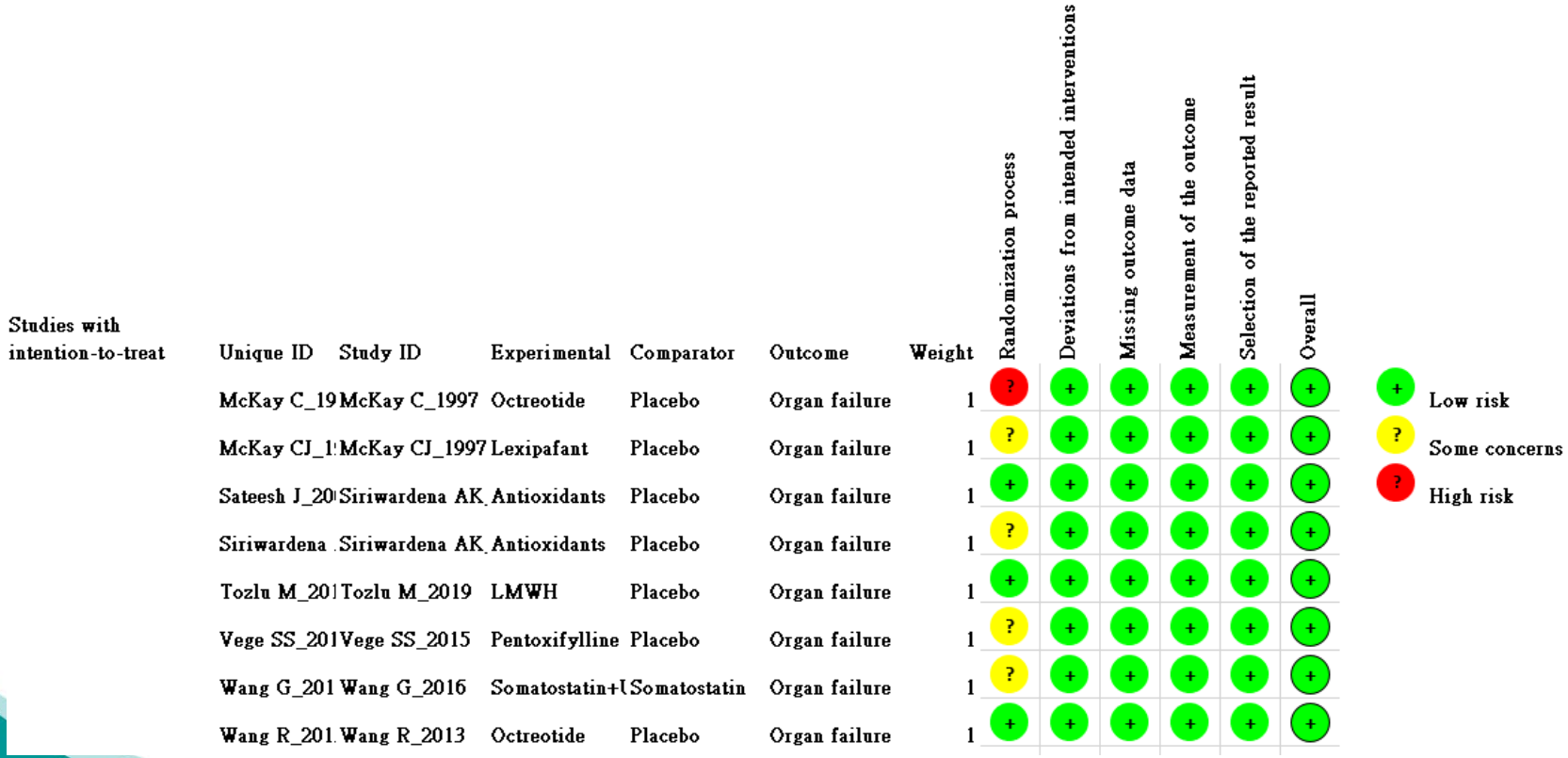
Domain-based approach in Cochrane's "Risk of bias" (ROB) tool (for RCT)

Domain	Signalling question	Response
Bias in selection of the reported result	5.1 Were the data that produced this result analysed in accordance with a pre-specified analysis plan that was finalized before unblinded outcome data were available for analysis?	Y
	5.2 ... multiple eligible outcome measurements (e.g. scales, definitions, time points) within the outcome domain?	N
	5.3 ... multiple eligible analyses of the data?	N
	Risk of bias judgement	**Low**

TCVGH

使用RCT研究進行統合分析→確認研究品質

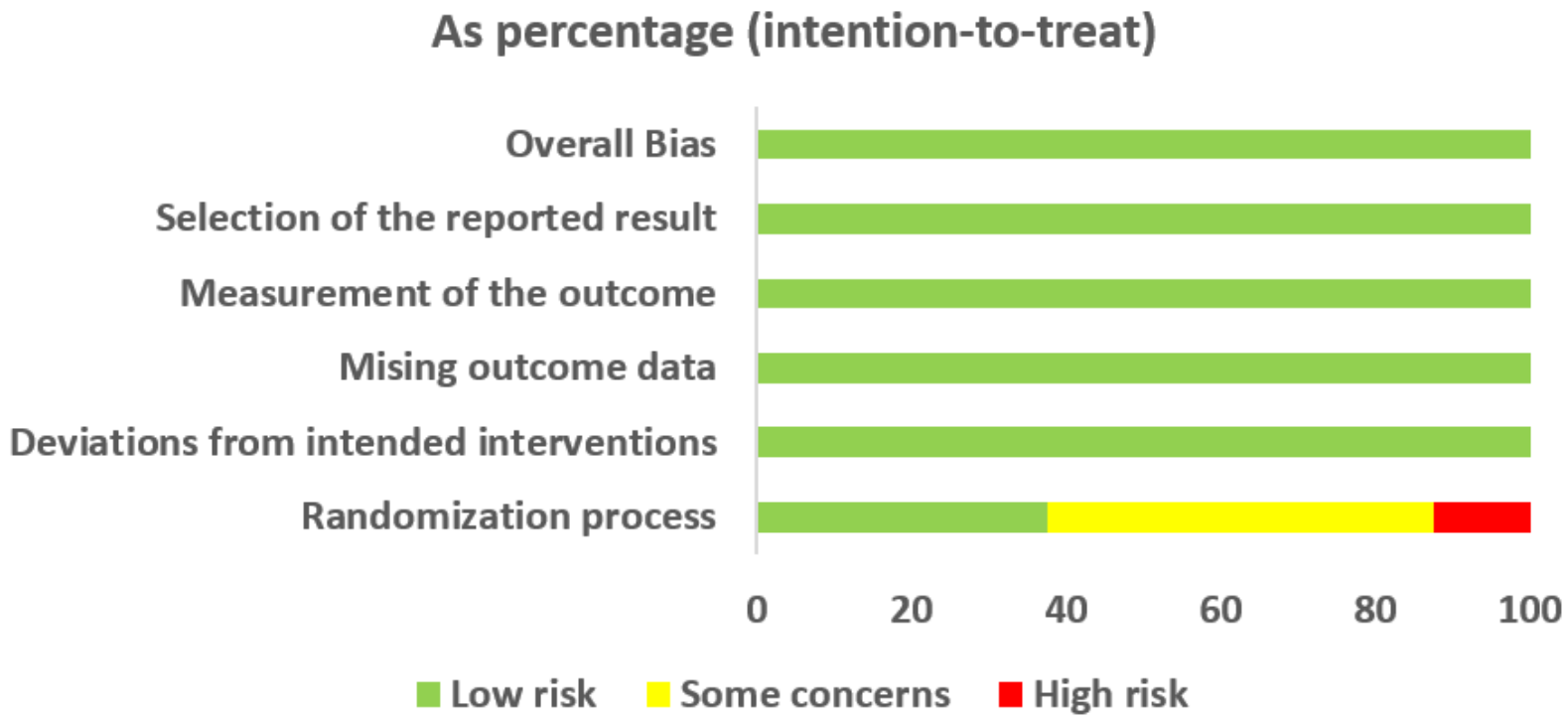
ROB Summary



TCVGH

使用RCT研究進行統合分析→確認研究品質

ROB graph across studies: example



TCVGH

使用Cohort研究進行統合分析→確認研究品質

Newcastle-Ottawa Quality Assessment Scale: Cohort Studies

- Selection (4)
- Comparability (1)
- Outcome (3)

由兩個人**double check**

- Most observational studies do not have a protocol: Risk of "cherry-picking" outcomes **(reporting bias)**
- Particular care should be given to **confounding**
- Focus on the **design** **"features"** rather than design "labels" **(cohort vs. case-control, prospective vs. retrospective)**

TCVGH

使用Cohort研究進行統合分析→確認研究品質

Newcastle-Ottawa Quality Assessment Scale: Cohort Studies

Selection

1. Representativeness of the exposed cohort
   a) truly representative of the average _____ (describe) in the community ♦
   b) somewhat representative of the average _____ in the community ♦
   c) selected group of users eg nurses, volunteers
   d) no description of the derivation of the cohort

2. Selection of the non exposed cohort
   a) drawn from the same community as the exposed cohort ♦
   b) drawn from a different source
   c) no description of the derivation of the non exposed cohort

3. Ascertainment of exposure to implants
   a) secure record (eg surgical records) ♦
   b) structured interview ♦
   c) written self report
   d) no description

4. Demonstration that outcome of interest was not present at start of study
   a) yes ♦
   b) no

TCVGH

使用Cohort研究進行統合分析→確認研究品質

Newcastle-Ottawa Quality Assessment Scale: Cohort Studies

Comparability

1. Comparability of cohorts on the basis of the design or analysis

a) study controls for _____ (select the most important factor) ♦

b) study controls for any additional factor (This criteria could be modified to indicate specific control for a second important factor.) ♦

TCVGH

使用Cohort研究進行統合分析→確認研究品質

Newcastle-Ottawa Quality Assessment Scale: Cohort Studies

Outcome

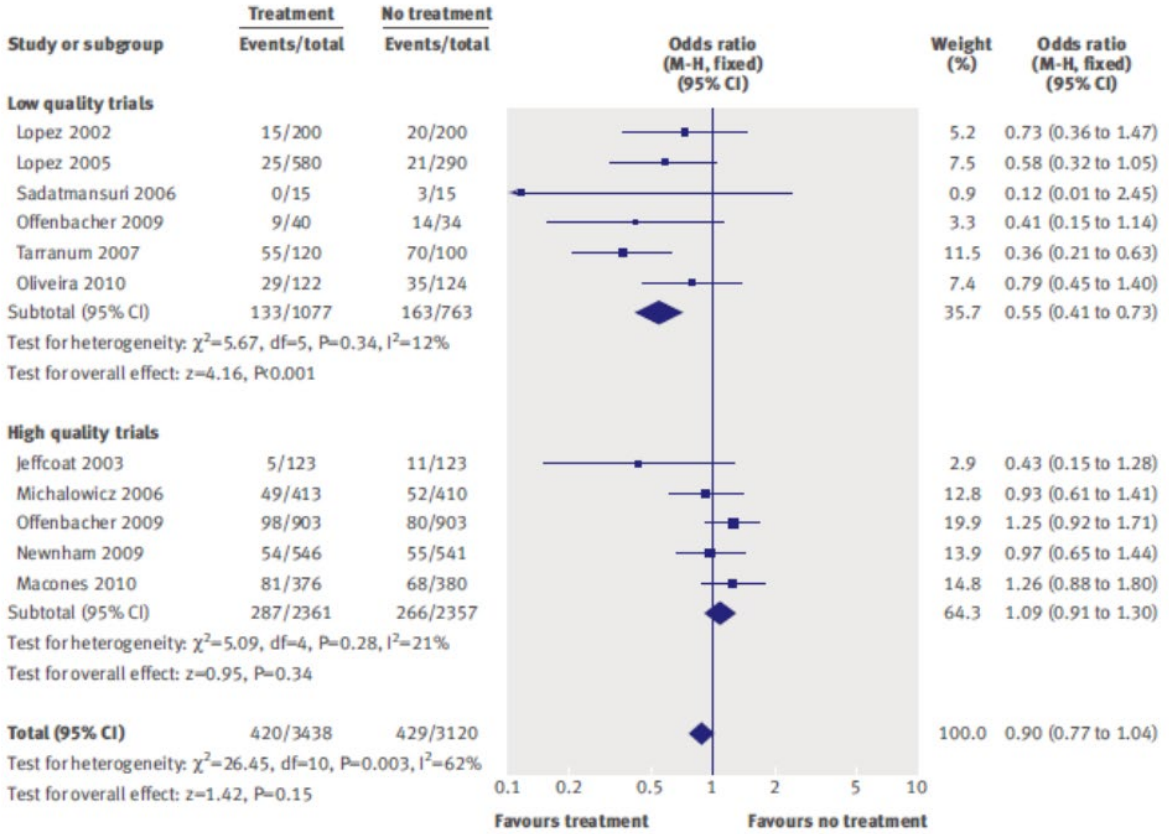
1. Assessment of outcome
   a) independent blind assessment ♦
   b) record linkage ♦
   c) self report
   d) no description

2. Was follow up long enough for outcomes to occur
   a) yes (select an adequate follow up period for outcome of interest) ♦
   b) no

3. Adequacy of follow up of cohorts
   a) complete follow up - all subjects accounted for ♦
   b) subjects lost to follow up unlikely to introduce bias - small number lost - > ___ % (select an adequate %) follow up, or description of those lost) ♦
   c) follow up rate < ___% (select an adequate %) and no description of those lost
   d) no statement

TCVGH

使用RCT研究進行統合分析→確認研究品質



Study or subgroup	Treatment Events/total	No treatment Events/total	Odds ratio (M-H, fixed) (95% CI)	Weight (%)	Odds ratio (M-H, fixed) (95% CI)
Low quality trials					
Lopez 2002	15/200	20/200		5.2	0.73 (0.36 to 1.47)
Lopez 2005	25/580	21/290		7.5	0.58 (0.32 to 1.05)
Sadatmansuri 2006	0/15	3/15		0.9	0.12 (0.01 to 2.45)
Offenbacher 2009	9/40	14/34		3.3	0.41 (0.15 to 1.14)
Tarranum 2007	55/120	70/100		11.5	0.36 (0.21 to 0.63)
Oliveira 2010	29/122	35/124		7.4	0.79 (0.45 to 1.40)
Subtotal (95% CI)	133/1077	163/763		35.7	0.55 (0.41 to 0.73)
Test for heterogeneity: χ^2=5.67, df=5, P=0.34, I^2=12%					
Test for overall effect: z=4.16, P<0.001					
High quality trials					
Jeffcoat 2003	5/123	11/123		2.9	0.43 (0.15 to 1.28)
Michalowicz 2006	49/413	52/410		12.8	0.93 (0.61 to 1.41)
Offenbacher 2009	98/903	80/903		19.9	1.25 (0.92 to 1.71)
Newnham 2009	54/546	55/541		13.9	0.97 (0.65 to 1.44)
Macones 2010	81/376	68/380		14.8	1.26 (0.88 to 1.80)
Subtotal (95% CI)	287/2361	266/2357		64.3	1.09 (0.91 to 1.30)
Test for heterogeneity: χ^2=5.09, df=4, P=0.28, I^2=21%					
Test for overall effect: z=0.95, P=0.34					
Total (95% CI)	420/3438	429/3120		100.0	0.90 (0.77 to 1.04)
Test for heterogeneity: χ^2=26.45, df=10, P=0.003, I^2=62%					
Test for overall effect: z=1.42, P=0.15					

Fig 6| Meta-analysis plot for overall adverse pregnancy outcome (preterm birth <37 weeks of gestation and spontaneous abortions/stillbirths). M-H=Mantel-Haenszel model

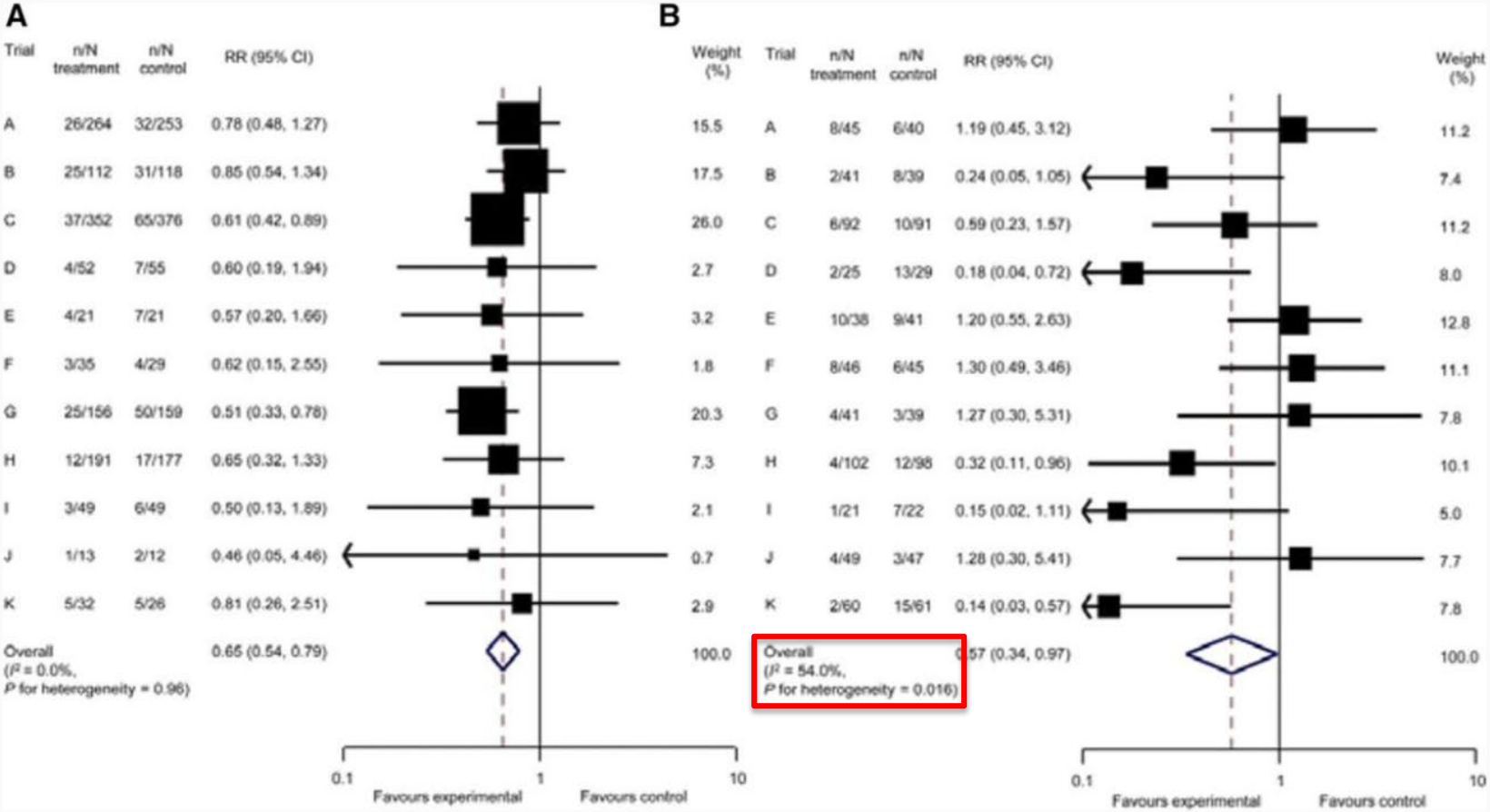
Polyzos et al, BMJ 2010

TCVGH

研究出現高異質性怎麼辦?

$I^2 \leq 50\%$: Homogeneous (fixed effect)

$I^2 > 50\%$: Heterogeneity (random effect mode)



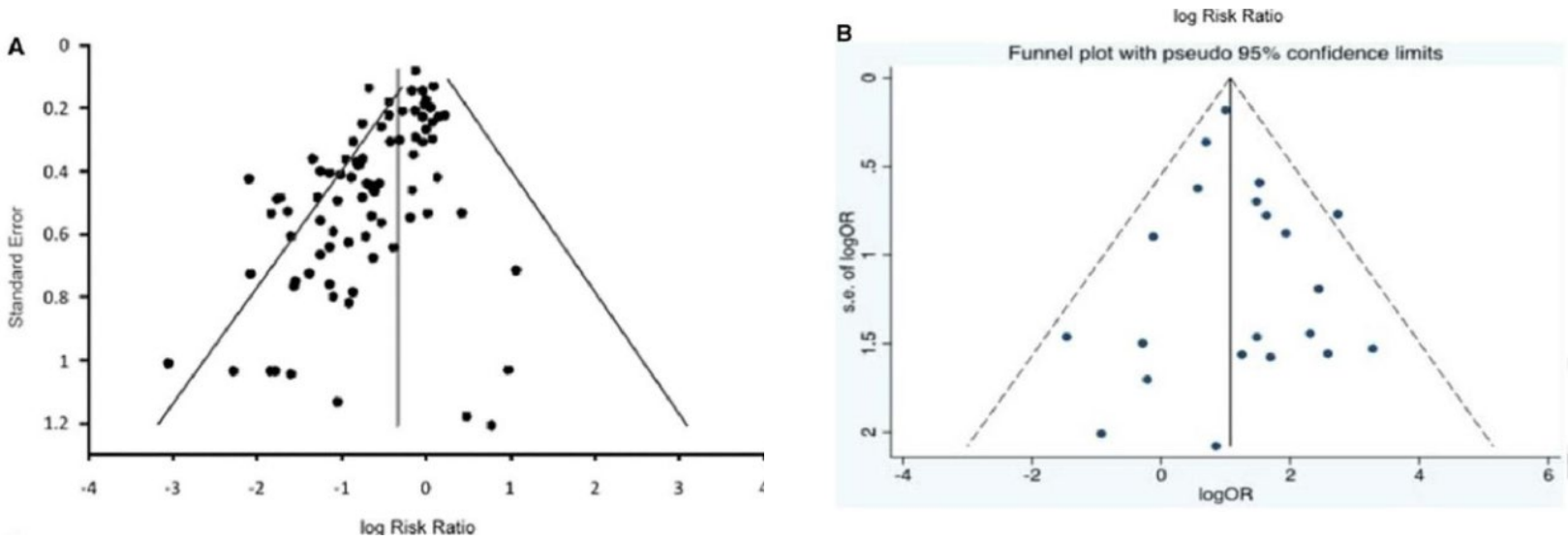
TCVGH

研究出現高異質性怎麼辦?

- 不要先急著作統合分析
  - 統合性迴歸分析（meta-regression）
  - 次群組分析（subgroup-analysis）：找出具有明顯的 category 差別的變項
    - 總論文數小於10篇以下，盡量不要作統合性迴歸分析 → Egger’s test
  - 敏感度分析（sensitivity analysis）：
    - 將某些不合適的論文（例如壁報或品質差的論文） 刪除
  - 使用Random effect model

圖像化評估 Publication Bias: Funnel Plot

「出版性偏差」（publication bias）：研究的質素相若，但報告較大效應值的大型研究，相比於報告較小、或沒有效應的小型研究更常被發表出版的情況。
「出版性偏差」的風險：會令綜合性的研究並不能準確地代表某主題的所有研究，而只偏重於較極端的結果。



漏斗圖：(A) 有出版性偏差、(B) 無出版性偏差

TCVGH

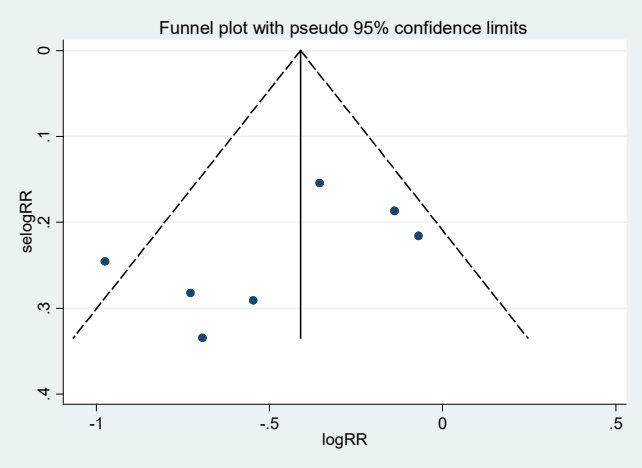
圖像化評估 Publication Bias: Funnel Plot

Introduction to the "metafunnel" Module

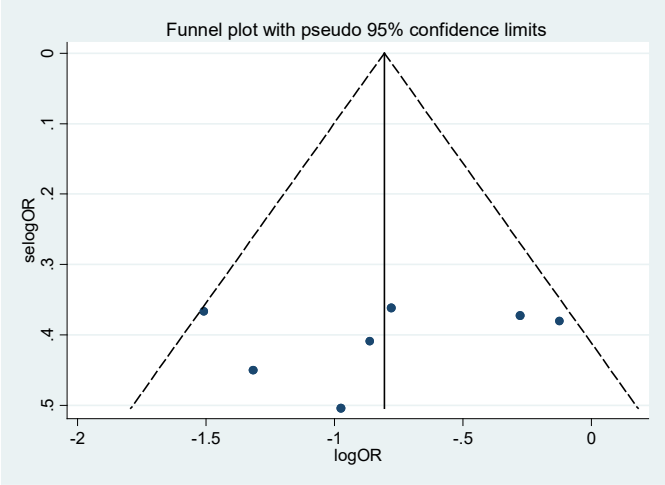
search(metafunnel)
search(metabias)

insheet using "D:\助理研究員\中榮醫研部-生統小組\全院教育課程規劃-2022oct\111年第4季\20221228-初探Meta-analysis\afreg.csv"

metafunnel logrr selogrr



metafunnel logor selogor



TCVGH

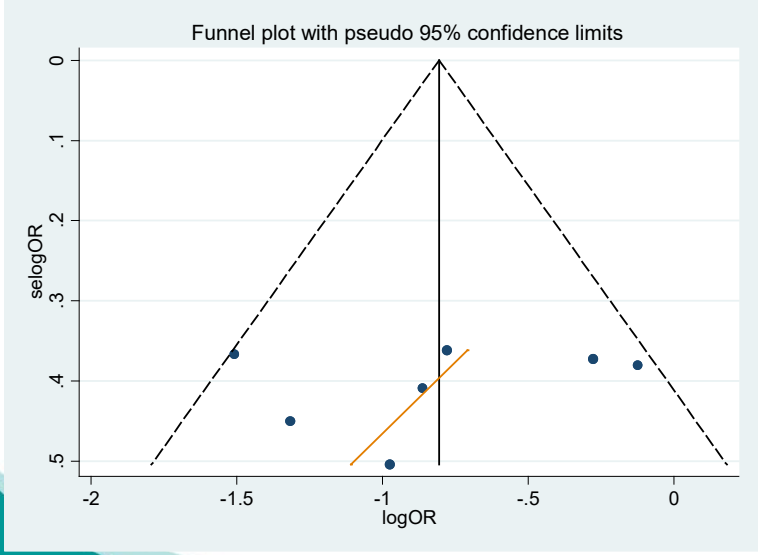
圖像化評估 Publication Bias: Funnel Plot

樣本數太少?
Small size effect: Egger's test

search(metafunnel)
search(metabias)

insheet using "D:\助理研究員\中榮醫研部-生統小組\全院教育課程規劃-2022oct\111年第4季\20221228-初探Meta-analysis\afreg.csv"

metafunnel logor selogor, egger



metabias logor selogor, egger

```
. metabias logor selogor, egger graph

Note: default data input format (theta, se_theta) assumed.

Tests for Publication Bias

Begg's Test

  adj. Kendall's Score (P-Q) =      -12
        Std. Dev. of Score =    18.27 (corrected for ties)
      Number of Studies    =       14
                         z  =    -0.66
                   Pr > |z| =    0.511
                         z  =     0.60 (continuity corrected)
                   Pr > |z| =    0.547 (continuity corrected)

Egger's test
```

Std_Eff	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
slope	.3068297	1.247459	0.25	0.810	-2.41115	3.024809
bias	-2.8082	3.130834	-0.90	0.387	-9.629702	4.013302

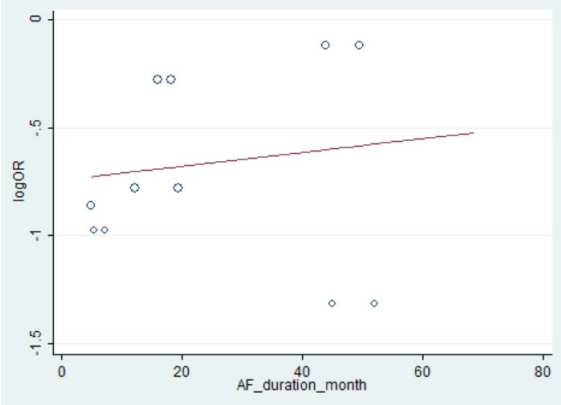
若有 Publication Bias，則做 Meta-regression去看是否有差異？

search (metareg)

#單看特定變數會不會影響結局 (只能放一個變數)

AF duration會不會影響復發

metareg logor af_duration_month , wsse(selogor) graph




```
. metareg logor af_duration_month , wsse(selogor) graph

Meta-regression                                 Number of obs   =        12
REML estimate of between-study variance         tau2            =    .02237
% residual variation due to heterogeneity       I-squared_res   =    13.84%
Proportion of between-study variance explained  Adj R-squared   =   -82.67%
With Knapp-Hartung modification
```

logor	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
af_duratio~h	.0032216	.0072379	0.45	0.666	-.0129054	.0193486
_cons	-.7446471	.2116059	-3.52	0.006	-1.216134	-.2731598

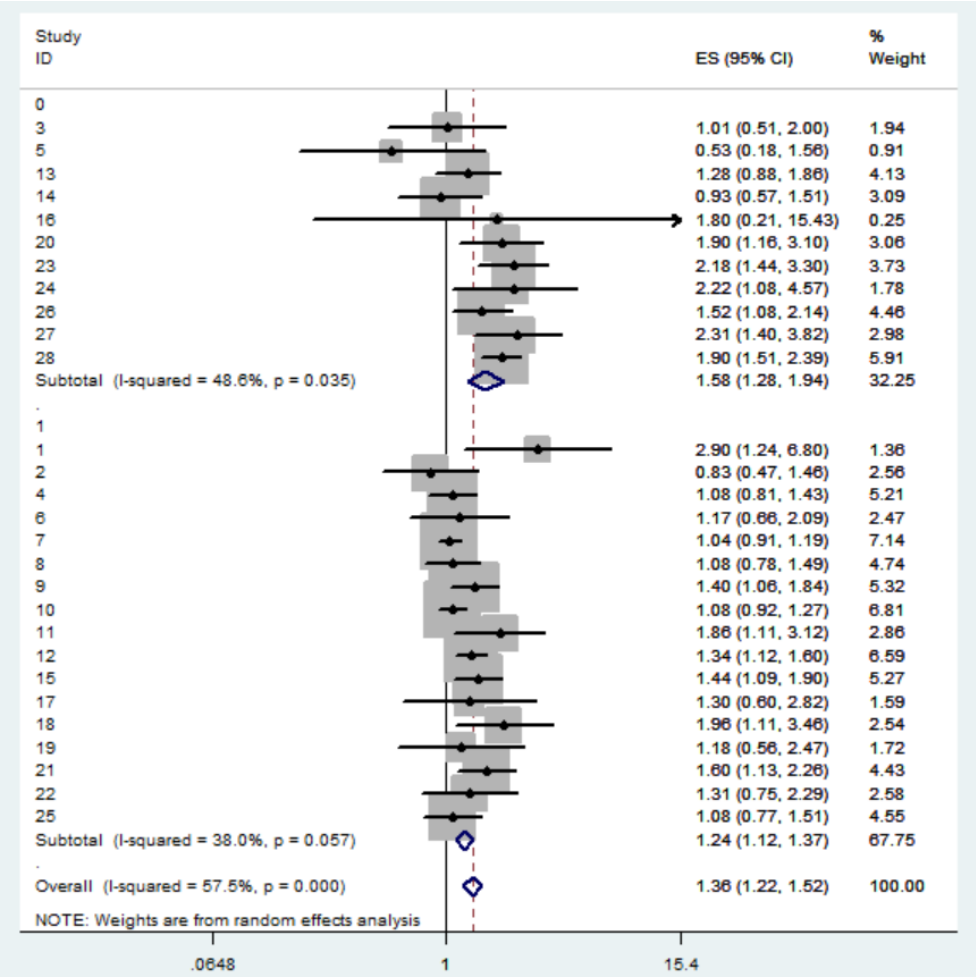
afreg.csv

TCVGH

若Meta-regression有差異 →
把有差異的那群分層去比較 (Sub-group analysis)

insheet using "D:\助理研究員\中榮醫研部-生統小組\全院教育課程規劃-2022oct\111年第4季\20221228-初探Meta-analysis\nodm.csv"

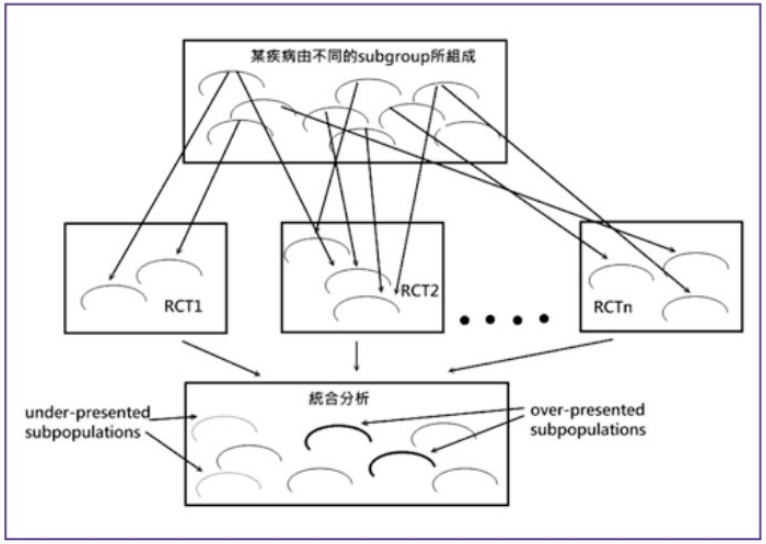
#Sub-group analysis
(依adjust與否分層): Random-effects model
metan logrr selogrr, random eform by(sex)



總結

- 僅由單一個隨機分派研究的結果來下結論是一種比較危險的行為，萬一這個結果有隨機錯誤時（error by chance），我們就有可能對某個醫學議題造成誤判。
- 統合分析可以提供較客觀的整合分析結果，對於不合適的研究我們也可藉由敏感性分析將其剔除，而使分析結果更正確。
- 隨機分派研究與觀察性研究的證據強度（level of evidence）是不同的，我們在看一篇統合分析的論文時一定要注意所選取論文的研究種類、品質、和訊息強度。

統合分析和隨機分派研究論文結果牴觸的可能原因:
某些特定族群被過度呈現



記得填寫滿意度問卷喔！

生統小組：統計方法教育訓練

臺中榮民總醫院 Taichung Veterans General Hospital

TCVGH