財團法人醫藥品查驗中心
Center For Drug Evaluation, Taiwan
當代醫藥法規
Reg Med News

統合分析(Meta-analysis)簡介

林資荃[1]

前言

由於醫學期刊上所發表的論文如雨後春筍般一直出現,因此我們很難對某一特定藥物用於治療某特定疾病之相關文獻全部閱讀完,而且相同藥物對於相同疾病文獻之間的結果也有可能不一樣,為了要將這些相同藥物對於相同疾病的文獻進行統計分析並且計算出合併效果時,這個新興統計分析方法稱為統合分析(Meta-analysis)也就自然而然產生。

統合分析是對於相同主題系統性結合先前所有相關文獻結果給予研究者一個量化結論,因此統合分析也定義為「分析的分析(The Analysis of Analyses)」,意思就是說統合分析作的事情就是將「許多的研究結果彙整出一個總結論」。

如果要作統合分析,則必須拿到每一個研究的原始資料或是摘要性統計量(summary statistics),就統計觀點如果可以拿到每篇研究的原始資料當然是最好,因為這種情形下所有資訊最充足,然而由於時間與金錢上考量幾乎不可能得到每一個研究原始資料,大多都是利用每一個研究之摘要性統計量作統合分析。

本文章第二章節將介紹歐盟 2001 年所公佈的統合分析審查考量要點(Points to Consider on Applications with Meta-analyses),第三章節將完整介紹常用的統計分析方法,第四章節則以離散型資料為例,透過不同單位來估算 statin 治療的效應大小(effect size)並進行統合分析,最後結語則討論統合分析方法的優缺點。

統合分析審查考量要點(基於歐盟 2001 所公佈 Points to Consider on Applications with Meta-analyses)

歐盟於 2001 年所公佈的統合分析審查考量要點(Points to Consider on Applications with Meta-analyses)中提及一個好的統合分析應該在事前規劃好統合分析計畫書(包括分析時間點),而計畫書中須要清楚說明整個統合分析流程,其事前規劃流程可以分為下面四個重要步驟:

步驟一:研究目的

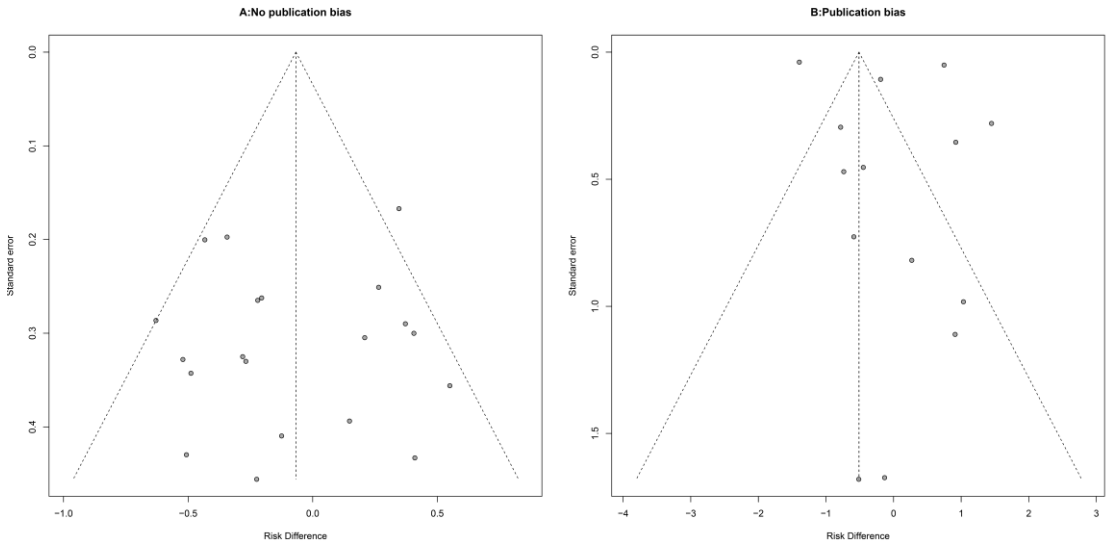
在執行統合分析前,一定要先清楚定義所要研究的主題。例如:若所要研究主題為「同時接受放射治療與化學治療對於非小細胞肺癌的療效」那麼當然只有同時接受放射治療與化學治療的文章才要納入選取。

[1] 財團法人醫藥品查驗中心新藥科技組

步驟二:試驗選取及排除

一旦研究問題確定後,可以透過電子資料庫搜尋或透過相關字搜尋相關研究,選取研究可以依其試驗設計型態作為選取標準(例如:隨機、雙盲與控制組試驗應為首選);另外,不同研究之病人納入與排除條件(inclusion and exclusion criteria)、劑量(dosage)與試驗長短(study duration)盡量不要差異太大,以避免研究之間族群、劑量高低與研究長短之異質性產生。

然而,學術上所發表文章通常是正面結果(positive study)為多數,所以在蒐集相關主題的研究時,往往會出現出版性偏差(publication bias)。如果有出版性誤差問題存在時,將造成分析上不客觀且會有偏差存在,因此要評估統合分析所選取的論文有無出版性偏差的問題可以透過漏斗圖(funnel)來檢視各試驗效應的差異性。舉例來說,圖一中 A 圖所選取研究結果較對稱較無出版性偏差疑慮,而 B 圖所選取研究結果較不對稱可能有出版性偏差問題存在。



圖一:漏斗圖 A:無出版性偏差,B:有出版性偏差(X 軸:Risk Difference; Y 軸:Standard error)

步驟三:分析指標及假說

統合分析的指標必須要在統合分析計畫書中定義清楚(療效指標還是安全性指標);另外,此分析假說是優越性、不劣性或是相等性都要清楚載明。

步驟四:統計分析方法

統合分析依據不同資料型態其統計方法有所不同,茲將不同資料型態所對應統計方法摘述如下:

1. 若有每一個研究病人的原始資料，則採用混合式模型(Mixed-effects model)作分析。混合式模型指的是統計模型中將重要的解釋變數當作固定因子，而將研究間變異當作隨機因子。

2. 若是僅有每一個研究的摘要性統計量，則採用固定式模型(Fixed-effects model)或是隨機式模型(Random-effects model)作分析，其差異為前者將研究間變異當作固定因子，後者將研究間變異當作隨機因子，兩種分析方法將在第三個章節完整介紹。

由於金錢與時間等現實考量下，要得到每一個研究病人所有資料是有困難的，因此大部份情況下都是透過摘要性統計量作分析。假設我們擁有每一個研究的摘要性統計量，則執行統合分析前，首先要把每一個研究效應大小的單位作統一。一般來說對於連續型資料(continuous data)通常使用單位為平均數差異(mean difference)或標準化後的平均數差異(standardized mean difference);而二元資料(dichotomous data)，例如:有發生事件與沒有發生事件，通常使用單位為相對危險度(Relative risk)、風險差(Risk difference)與勝算比(Odds ratio);而存活資料(survival data)，通常使用單位為風險比(Hazard ratio)。

統合分析的統計方法

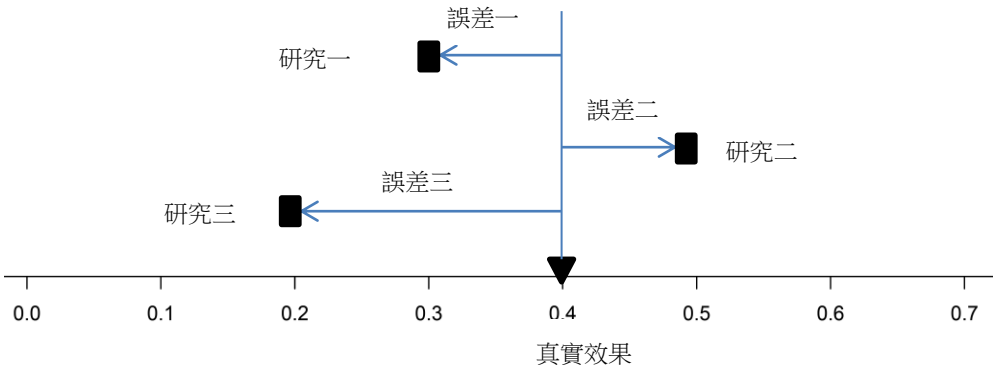
通常臨床試驗假設為:H_0:試驗藥物與安慰劑效應大小無差異 vs. H_1:試驗藥物與安慰劑效應大小有差異，若定義試驗藥物效應大小為 δ，則上面假設可以改寫成

$$H_0: \delta=0 \text{ vs. } H_1: \delta \neq 0$$

假設有 K 個研究被納入統合分析，每一個研究所觀察到試驗藥物估計效應大小為δ_i 其相對變異數為$\widehat{\sigma_i^2}$ ，$i = 1, \cdots K$。以下分別介紹固定與隨機兩種模式之統計方法來對藥物效應大小 δ 進行統合估計與檢定:

1. 固定式模型 (Fixed-effects model)

固定式模型是假設所有的研究都有一個共同的真實效果(true effect)，而每一個研究所觀察到的效果稱為觀察效果(observed effect)，每一個研究的觀察效果與真實效果不同是因為抽樣誤差(sampling error)所造成。抽樣誤差來源可能是每一個研究的劑量不同、病人疾病嚴重程度不同、年齡分布不同、或是所併用藥物不同等原因所導致。以圖二為例，假設有三個研究納入統合分析，由圖二可以看出三個研究的真實效果都是 0.4，但是所觀察到的效果在研究一是 0.3;研究二是 0.5;研究三是 0.2，因此我們可以看出研究一之抽樣誤差為-0.1;研究二為 0.1;研究三為-0.2。



圖二:固定式模式:觀察效果=真實效果+研究內誤差

我們可以把上述觀念寫成數學等式:

$$\delta_i = \delta + \varepsilon_i$$

其中δ_i為第 i 個研究觀察到效果、δ為所有研究共同的真實效果、ε_i為第 i 個研究內變異量，因為誤差是在真實效果附近上下跳動，因此可以合理假設ε_i服從平均數為 0 變異數為σ_i^2之常態分配。在固定式模型中我們所要估計的是共同的真實效果 δ，其估計方法為給予每一個研究一個權重，而權重大小為該研究內變異量之變異數的倒數$(1/\sigma_i^2)$，然後就可以計算出真實效果估計值與變異數，公式如下:

$$\hat{\delta} = \frac{\sum_{i=1}^{k} \frac{\delta_i}{\sigma_i^2}}{\sum_{i=1}^{k} \frac{1}{\sigma_i^2}} \quad \text{且} \quad \mathrm{Var}(\hat{\delta}) = \frac{1}{\sum_{i=1}^{k} \frac{1}{\sigma_i^2}}$$

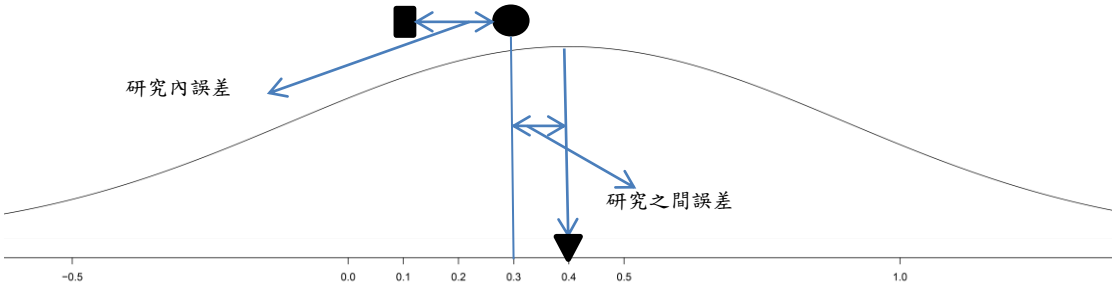
因此 95%信賴區間為$\hat{\delta} \pm 1.96\sqrt{\mathrm{Var}(\hat{\delta})}$。最後用 z test $(\hat{\delta}/\sqrt{\mathrm{Var}(\hat{\delta})})$來檢定真實效果 δ 是否為 0。單尾檢定 P-value=1-Φ(|Z|);而雙尾檢定 P-value= 2×[1-Φ(|Z|)]，其中 Φ(Z)代表標準常態分配之累積函數。

2. 隨機式模型 (Random-effects model)

隨機式模型是假設每一個研究的真實效果都不一樣，而每一個研究真實效果不同可能是每一個研究的劑量不同、病人疾病嚴重程度不同、年齡分布不同、或是所併用藥物不同等原因所導致。在隨機式模型中，我們要估計的是真實效果的整體平均值。

假設有三個研究納入統合分析，此三個研究的真實效果與上述固定式模式觀察效果一樣;也就是研究一是 0.3;研究二是 0.5;研究三是 0.2。以研究一為例(圖三)，

研究一的真實效果為 0.3 而觀察到效果是 0.1，但我們所要估計的是真實效果的整體平均值假設是 0.4。因此觀察效果與真實效果的整體平均值間差異來自於兩個變異一個是研究之間變異而另外一個是研究內變異所造成。以研究一為例:研究之間變異為−0.1(=0.3-0.4)而研究內變異為−0.2(=0.1-0.3)。



圖三:隨機式模型:觀察效果=真實效果的整體平均值+研究之間變異+研究內變異

我們可以把上述觀念寫成數學等式:

$$\delta_i = \delta + v_i + \varepsilon_i$$

其中δ_i為第 i 個研究觀察到效果、δ為所有研究真實效果的整體平均值、v_i為第 i 個研究與其它所有研究之間變異、ε_i為第 i 個研究內變異且假設v_i與ε_i分別服從平均數為 0 變異數為τ^2與平均數為 0 變異數為σ_i^2之常態分配且互為獨立。在隨機式模型中我們所要估計的是所有研究真實效果的整體平均值 δ，其估計方法為給予每一個研究一個權重，而權重大小為該研究所有變異量之變異數的倒數$1/(\tau^2+\sigma_i^2)$，然後就可以計算出真實效果的整體平均值之估計值與變異數，公式如下:

$$\hat{\delta} = \frac{\sum_{i=1}^{k} \frac{\delta_i}{\widehat{\sigma_i^2}+\widehat{\tau^2}}}{\sum_{i=1}^{k} \frac{1}{\widehat{\sigma_i^2}+\widehat{\tau^2}}} \quad \text{且} \quad \mathrm{Var}(\hat{\delta}) = \frac{1}{\sum_{i=1}^{k} \frac{1}{\widehat{\sigma_i^2}+\widehat{\tau^2}}}$$

因此 95%信賴區間為$\hat{\delta}\pm 1.96\sqrt{\mathrm{Var}(\hat{\delta})}$。最後用 z test $(\hat{\delta}/\sqrt{\mathrm{Var}(\hat{\delta})})$來檢定真實效果的整體平均值 δ 是否為 0。單尾檢定 P-value=1-Φ(|Z|); 而雙尾檢定 P-value=2×[1-Φ(|Z|)]，其中 Φ(Z)代表標準常態分配之累積函數。

至於檢定研究之間有無異質性的問題存在，常用統計量為Q或I^2，其定義如下:

$$Q = \sum_{i=1}^{k} w_i(\delta_i - \bar{\delta})^2 \text{，} \bar{\delta} = \frac{\sum_{i=1}^{k} w_i\delta_i}{\sum_{i=1}^{k} w_i}$$

$$I^2 = \left(\frac{Q-(K-1)}{Q}\right) \times 100\%$$

其中K為納入分析試驗個數，δ_i為第 i 個研究觀察到效果、w_i為估計的權重。當

虛無假設(沒有異質性存在)成立時，統計量Q會服從卡方分配自由度為K−1，因此若在虛無假設下，$Q > \chi^2_{\text{K-1,0.95}}$即有異質性問題存在，其中$\chi^2_{\text{K-1,0.95}}$代表卡方分配自由度為K−1 之 95%分位數。而統計量I^2為 25%、50%與 75%分別代表低度、中度與高度異質性問題存在。

如果研究之間沒有異質性的問題存在時，就使用固定式模型來分析;如果存在有異質性的問題則使用隨機式模型來分析，但這是不對的觀念。正確來說，即使檢定結果沒有異質性的問題存在，如果研究者認為研究間有異質性的可能時(例如:多國多中心臨床研究)就應該使用隨機式模型分析。一般來說，建議研究者同時執行固定式模型與隨機式模型分析，再作比較會比較客觀。

案例介紹

Cannon *et al.* (2006)透過統合分析比較高劑量(high-dose)的 statin 治療(實驗組)是否較標準劑量(standard-dose)的 statin 治療(控制組)有效降低冠狀動脈性猝死(coronary death)或心肌梗塞(myocardial infarction)事件發生(較優性假設)。利用此文獻資料與依據歐盟 2001 年所公佈統合分析的審查考量要點進行分析，其分析步驟如下:

步驟一:研究目的

此分析目的是比較高劑量的 statin 治療(實驗組)是否較標準劑量的 statin 治療(控制組)有效降低冠狀動脈性猝死或心肌梗塞事件發生(較優性假設)。

步驟二: 試驗選取及排除

試驗選取準則為隨機、具控制組、試驗人數超過 1000 個病人且是用臨床上面結果當作主要評估指標才納入分析，依據此準則，作者選取了四個代表性試驗納入統合分析，此四個試驗分別為:

- PROVE IT-TIMI-22 (Pravastatin or Atorvastatin Evaluation and Infection Therapy-Thrombolysis In Myocardial Infarction-22)
- A-to-Z (Aggrastat to Zocor)
- TNT (Treating to New Targets) trials
- IDEAL (Incremental Decrease in End Points Through Aggressive Lipid Lowering)

其中每一個試驗之試驗設計與每一個試驗發生冠狀動脈性猝死或心肌梗塞事件人數與試驗總人數如表一與表二所示:

表一:每一個試驗試驗設計

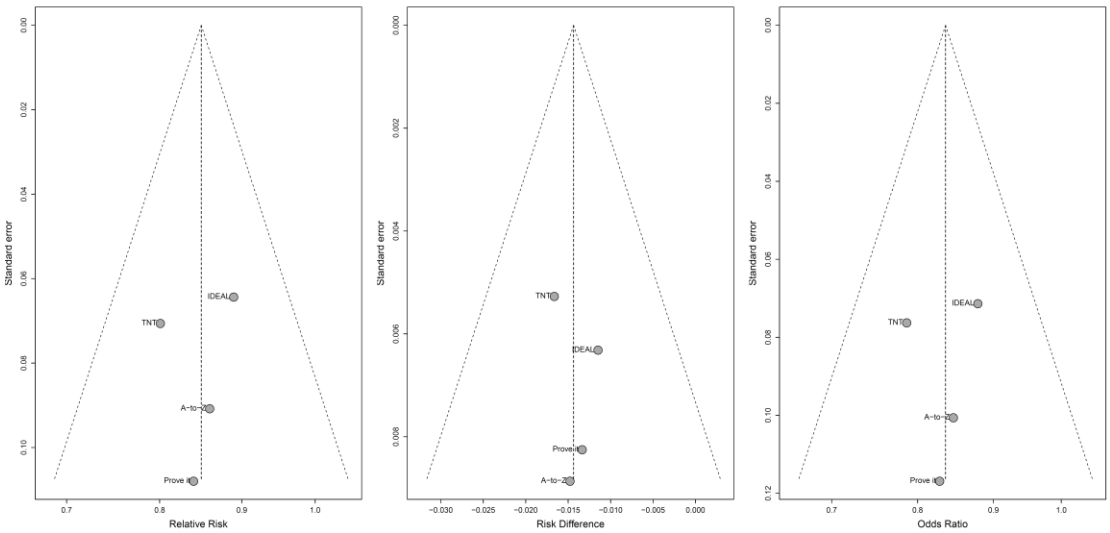
試驗名稱	PROVE IT-TIMI-22	A-to-Z	TNT	IDEAL
總人數	4162	4497	10001	8888
族群	Post-ACS	Post-ACS	Stable CAD	Stable CAD
試驗分組數	40 mg pravastatin vs. 80 mg atorvastatin	Placebo (4 months) then 20 mg simvastatin vs. 40 mg simvastatin (1 month) then 80 mg simvastatin	10 mg atorvastatin vs. 80 mg atorvastatin	20 mg simvastatin vs. 80 mg atorvastatin
試驗長短	24 months (mean)	721 days (median)	4.9 years (median)	4.9 years (median)
主要評估指標	Death, MI, UA requiring hospitalization, revascularization (>30 days), stroke	CV death, MI, readmission for ACS, stroke	CHD death, Non–procedure-related MI, resuscitation after cardiac arrest, stroke	CHD death, MI, cardiac arrest with resuscitation

ACS =acute coronary syndrome; CAD =coronary artery disease; CHD =congenital heart disease; CV =cardiovascular;
MI =myocardial infarction; UA =unstable angina.

表二:每一個試驗總人數與發生事件人數

試驗名稱	高劑量的 statin 治療 (實驗組)		標準劑量的 statin 治療 (控制組)	
	發生事件人數	總人數	發生事件人數	總人數
Prove It	147	2099	172	2063
A-to-Z	205	2265	235	2232
TNT	334	4995	418	5006
IDEAL	411	4439	463	4449

首先我們可以透過漏斗圖來檢視,所選出來這四個試驗有無出版性偏差的問題。由圖四可以看出用三種不同單位(相對危險度、風險差與勝算比)來評估所選取試驗，大致上仍對稱沒有嚴重出版性偏差問題產生。



圖四:不種單位下之漏斗圖(statin 試驗)

步驟三: 假說及分析指標

此統合分析主要分析指標為冠狀動脈性猝死或心肌梗塞發生事件;主要假說是高劑量的 statin 治療(實驗組)是否較標準劑量的 statin 治療(控制組)有效降低冠狀動脈性猝死或心肌梗塞發生(較優性假設)。

步驟四: 統計分析方法

由於能收集到的資料為摘要性統計量,因此採用固定式模型與隨機式模型進行統合分析。而執行統合分析前要把每一個試驗效應大小的單位作統一,此資料為二元資料(有發生事件與沒有發生事件),使用單位為相對危險度、風險差與勝算比。茲將採用上述三種不同單位,進行統合分析:

1. 相對危險度(Relative risk)

相對危險度定義為實驗組發生機率與控制組發生機率之比,公式為

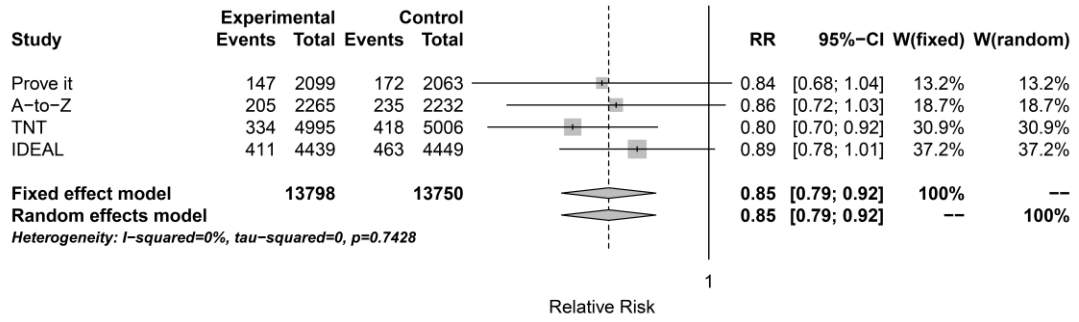
$$\mathrm{RR} = \frac{p_E}{p_C} = \frac{x_E/n_E}{x_c/n_c}$$

其中x_E為實驗組發生事件人數,n_E為實驗組總人數; x_c為控制組發生事件人數,n_c為控制組總人數。其 95%信賴區上下界分別為:

$$L_{RR} = \exp(lnRR - 1.96 \times SE_{lnRR}) \qquad U_{RR} = \exp(lnRR + 1.96 \times SE_{lnRR})$$

其中$SE_{lnRR} = \sqrt{\frac{1}{x_E} - \frac{1}{n_E} + \frac{1}{x_c} - \frac{1}{n_c}}$。

圖五顯示如果每一個試驗單獨作分析，則有三個試驗(Prove It、A-to-Z 與 IDEAL)顯示高劑量的 statin 治療與標準劑量的 statin 治療對於降低冠狀動脈性猝死或心肌梗塞事件發生是無顯著差異。但若統合這四個試驗資料作分析時，由於異質性檢定 p 值為 0.7428 顯示這四個試驗無顯著異質性存在。進一步可以發現用固定式模型與隨機式模型所得到相對危險度估計值都為 0.85 且都有達到統計上顯著(p<0.05)，因此使用相對危險度為單位，分析結果顯示高劑量的 statin 治療較標準劑量的 statin 治療有顯著降低冠狀動脈性猝死或心肌梗塞事件發生。



圖五:相對危險度之森林圖

2. 風險差(Risk difference)

風險差定義為實驗組發生機率與控制組發生機率之差，公式為

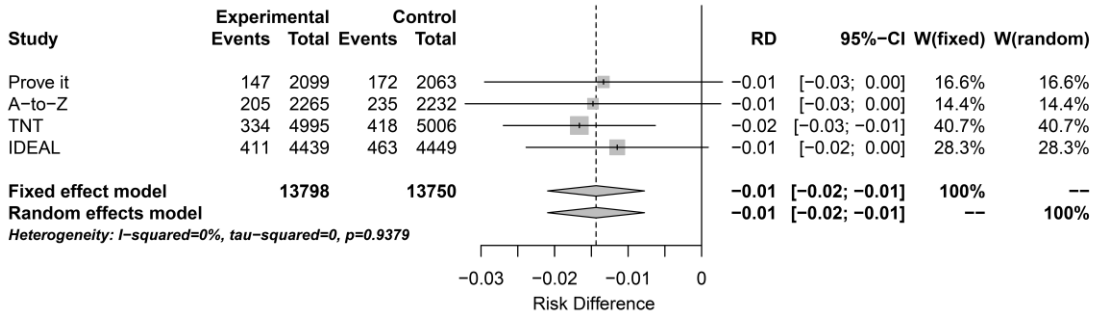
$$\text{RD} = p_E - p_c = \frac{x_E}{n_E} - \frac{x_c}{n_c}$$

，其 95%信賴區上下界分別為:

$L_{RD} = \text{RD} - 1.96 \times SE_{RD}$　　$U_{RD} = \text{RD} + 1.96 \times SE_{RD}$

其中$SE_{RD}=\sqrt{\frac{p_E(1-p_E)}{n_E}+\frac{p_c(1-p_c)}{n_c}}$。

圖六顯示如果每一個試驗單獨作分析，則有三個試驗(Prove It、A-to-Z 與 IDEAL) 顯示高劑量的 statin 治療與標準劑量的 statin 治療對於降低冠狀動脈性猝死或心肌梗塞事件發生是無顯著差異。但若統合這四個試驗資料作分析時，由於異質性檢定 p 值為 0.9379 顯示這四個試驗無顯著異質性存在。進一步可以發現用固定式模型與隨機式模型所得到風險差估計值都為-0.01 且都有達到統計上顯著(p<0.05)，因此使用風險差為單位結果顯示高劑量的 statin 治療較標準劑量的 statin 治療有顯著降低冠狀動脈性猝死或心肌梗塞事件發生。



圖六:風險差之森林圖

3. 勝算比(Odds ratio)

勝算定義為發生某事件的人數與未發生該事件人數的比值,而勝算比即為實驗組中發生事件的勝算與控制組中發生事件的勝算比值。以 TNT 試驗(表三)為例,依勝算比定義,其計算值為:

$$OR_{TNT} = \frac{334/4661}{418/4588} = 0.787$$

表三:TNT 試驗發生事件人數與沒有發生事件人數

	發生事件人數	沒有發生事件人數
High-dose	334	4661
Standard-dose	418	4588

若寫成公式為:

$$\text{OR} = \frac{\frac{p_E}{1-p_E}}{\frac{p_c}{1-p_c}} = \frac{A/B}{C/D} = \frac{AD}{BC}$$

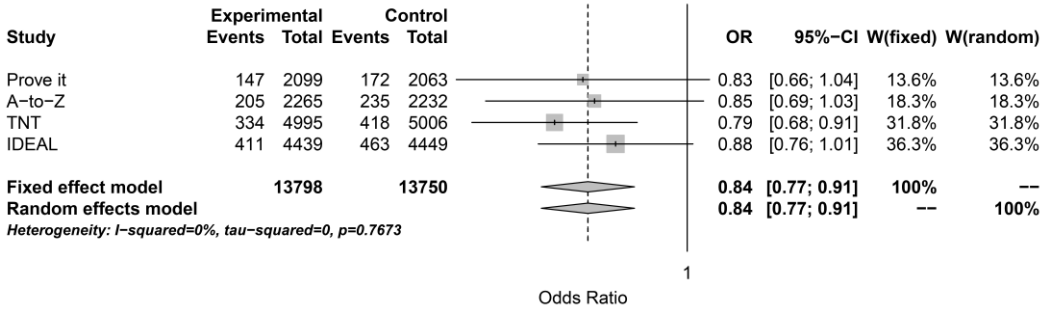
其中 A 為實驗組發生事件數目、C 為控制組發生事件數目、B 為實驗組未發生事件數目、D 為控制組未發生事件數目。其 95%信賴區上下界分別為:

$L_{OR} = \exp(lnOR - 1.96 \times SE_{lnOR})$ $U_{OR} = \exp(lnOR + 1.96 \times SE_{lnOR})$

其中$SE_{lnOR}=\sqrt{\frac{1}{A}+\frac{1}{B}+\frac{1}{C}+\frac{1}{D}}$。

圖七顯示如果每一個試驗單獨作分析,則有三個試驗(Prove It、A-to-Z 與 IDEAL)顯示高劑量的 statin 治療與標準劑量的 statin 治療對於降低冠狀動脈性猝死或心肌梗塞事件發生是無顯著差異。但若統合這四個試驗資料作分析時,由於異質性檢定 p 值為 0.7673 顯示這四個試驗無顯著異質性存在。進一步可以發現

用固定式模型與隨機式模型所得到勝算比估計值都為 0.84 且都有達到統計上顯著($p<0.05$)，因此使用勝算比為單位結果顯示高劑量的 statin 治療較標準劑量的 statin 治療有顯著降低冠狀動脈性猝死或心肌梗塞事件發生。



圖七:勝算比之森林圖

總結，此 statin 臨床試驗例子選用三種不同單位(相對危險度、風險差與勝算比)來計算 statin 治療的效應大小，透過統合分析(固定式模型與隨機式模型）其結果都是一致的，均顯示高劑量的 statin 治療較標準劑量的 statin 治療有顯著降低冠狀動脈性猝死或心肌梗塞事件發生。

結語

近年來統計軟體蓬勃發展，統合分析更可以透過軟體來協助完成，比如在 R 軟體(Chen, 2013)只要下載 meta 這個程式套件就可以幫助我們執行固定式模型與隨機式模型分析，而執行統合迴歸分析也可透過 metafor 這個程式套件來執行。近年來更發展出精確且有效率(exact and efficient)方法對罕見事件(rare events)分析，而這些方法都在 gmeta 這個程式套件裡面。

統合分析為醫學研究帶來莫大的幫助，例如:許多醫院醫師都會執行小型研究來探討試驗藥物有效性與安全性，然而小型研究常被人質疑的問題就是病人數目太少，造成沒有足夠資訊來了解試驗藥物的效應大小，這時如果透過統合分析來統合這些相同試驗的小型研究，可以得到比較公正客觀結果。

臨床試驗中，統合分析常用於估計不劣性試驗之臨界值(non-inferiority margin)與安全性方面評估。例如:糖尿病患者使用 Rosiglitazone 可能會增加心肌梗塞的風險之發現(Singh *et al.*, 2007)，便是利用統合分析所得到結論。至於統合分析結果是否能作為適應症之樞紐性證據，EMA 在審查要點中提到，如果依賴統合分析結果作為適應症之樞紐性證據，那麼通常會要求其 p 值要遠小於一般之顯著水準 0.05。另外，計畫書也應提供一致性(consistency)及穩健性(robustness)之評估計畫，對於不同指標、不同次族群與不同試驗子集(不同區域、不同試驗

長短或不同設計之試驗或品質佳之試驗等等)之各種分析皆有相同之效果。最後，指引也說明如果依賴回顧性(retrospective)統合分析提供充分的證據，其必要條件為: (1) 有些試驗很明顯的是正面的結果，(2)無結論的試驗在主要指標上趨勢是正面的，(3)沒有統計顯著的異質性存在，(4)整合的 95%信賴區間遠離 0（或風險比遠離 1、或是遠離不劣性試驗所預先定義的邊界值），(5)有合理的理由說明沒有選擇性試驗及/或指標的偏差發生，以及(6)敏感性分析驗證其結果之穩健性。然而，目前法規單位仍傾向不能以統合分析結果作為樞紐性證據唯一依據。

最後，值得注意的是統合分析結果的可信度與所選取的文獻有相當大關係，如果所選取文獻都是品質不好的，那統合分析出來品質當然也是不好，而透過漏斗圖來檢視也未必能檢查出所有出版性偏差問題。所以為了減少這些缺點產生，應在蒐集與統合相同主題文獻時，結合醫師、統計學家與相關專家的參與。

參考文獻

1. Cannon, C.P., Steinberg, B.A., Murphy, S.A., Mega, J.L., and Braunwald, E. Meta-Analysis of Cardiovascular Outcomes Trials Comparing Intensive Versus Moderate Statin Therapy. *Journal of the American College of Cardiology*, 2006; 438-445.
2. Chen, D.G. Applied Meta-Analysis with R. Deming Conference, Atlantic City, New Jersey December 9, 2013.
3. EMA, Points to Consider on Applications with Meta-analyses 2001.
4. Singh, S., Loke, Y.K., and Furberg, C.D. Long-term Risk of Cardiovascular Events With Rosiglitazone: A Meta-analysis. *Journal of the American Medical Association*, 2007;298:1189-95.
5. 莊其穆，臨床醫師如何閱讀統合分析(Meta-analysis)的論文。臺灣醫界，Vol.54，No.2，2011。
6. 李宛柔、林怡君、于耀華與賴玉玲，後設分析之介紹。牙醫學雜誌，29-2:63-68，2009。