

臨床醫師如何閱讀統合分析（Meta-analysis）的論文

文、圖/莊其穆

台北榮民總醫院婦產部

前言

由於研究風氣盛行，很多國家或是研究單位更是以論文作為升遷或是教職升等的重要指標，因此每年全世界發表的論文數量節節升高，舉例而言在PubMed登錄的論文數量在1987年為361,000篇，而1997年為448,000篇，而到了2007年則高達766,000篇⁽¹⁾。所以可以預期的在未來的論文數量還會更多，因此我們在閱讀論文的過程中，我們越來越難針對同一個主題閱讀完所有的相關文獻，而且對於同樣的主題，論文之間也可能有不同的結果，為了要將這些具有相同主題的論文進行統合分析，因此在1990年代開始發展了一個新興的統計方法來為相同主題的論文進行統計作出合併效果（summarized effect size）評估，這門新興統計方法稱為統合分析（以下稱統合分析）。

住院醫師們在閱讀期刊時一定有閱讀過統合分析的論文，除了很容易就可以看出的合併效果之外，很多的其他概念諸如森林圖（forest plot）、漏斗圖（funnel plot）的解讀，如何評估異質性（heterogeneity），甚麼是固定式模型（fixed-effects model）或隨機式模型（random-effects model）的統計方式一定是霧裡看花，其實這些統計概念並不難，只要你願意花些心思就可以了解上述的概念，日後甚至可以自己去針對某個題目作統合分析。

統合分析的發展歷史

統合分析的發展歷史最早可以追溯到17世紀，在當時的天文學界發現將數個觀察到的小型資料整合後會比單從資料中挑選合適資料要來的更準確⁽²⁾。而正式的運用較為正規的統計方法來整合研究資料則起源自1904年的著名統計學家Karl Pearson開始，他統合數個利用血清接種的治療方式來預防傷寒熱（enteric fever）的小型研究，來證明治療的有效性⁽³⁾。由於Pearson的演算方式仍非正規嚴謹的方式，隨後醫學界的統合分析發展便逐漸停滯。隨後的統合分析研究反而在社會學、精神醫學、和教育學逐漸成形。在1976年由精神分析師 Gene Glass首先正式提出「meta-analysis」這個名詞⁽⁴⁾。到了1979年英國的流行病學專家Archie Cochrane發現到一般民眾對於醫學議題沒有門路來收集可信賴的醫學證據，因此他認為應該成立

專門機構來統合這些研究的訊息。到了1980年代，統合分析的論文逐漸在心臟醫學，腫瘤醫學，和周產期照護的領域逐漸有相關論文發表。到了1990年代「The Foundation of The Cochrane Collaboration」正式成立，這個組織定期更新特定醫學議題的系統性文獻回顧（systemic reviews）和統合分析，從此統合分析的研究和論文發表進入了蓬勃發展的時代⁽⁵⁾。

統合分析的研究步驟

通常要做統合分析之前，一定會針對研究主題進行所謂的系統性文獻回顧。系統性文獻回顧和最簡單的敘述性回顧（narrative review）是不同的，一個嚴格的系統性文獻回顧會在正式研究前制定出protocol以明確規定列入條件（inclusion criteria）和排除條件（exclusion criteria）的標準。而統合分析也分為（1）已發表文獻的統合分析以及（2）整合式重新分析（pooled-reanalysis）。如果要做到整合式重新分析，則必須拿到相關論文的原始統計資料，然後將其整合成一個檔案再進行分析，這必須要這些論文的相關作者都願意提供原始資料給你，因此難度很高。以下表1列出systemic review和meta-analysis不同之處。本文以下的說明都是集中在已發表文獻的統合分析作說明（表1）：

由於統合分析是將數篇主要的研究論文結果作統合分析，在整個的研究過程有八項步驟要注意⁽¹⁾：

- 一、要研究的題目是否有事先定義好。在作meta-analysis之前一定要先自己是否已經嚴格定義好你的研究主題，唯有先定義好主題，你才能夠找出真正有相關性的論文。例如你要研究「同步放療化療對於局部廣泛型子宮頸癌（locally advanced cervical cancer）的治療效果」那麼對於只有接受放療，或是化療是在放射治療後才安排者，這些論文然要排除。
- 二、列入條件和排出條件是否有事先定義好。作任何研究一定要定義出你的受試者或是你欲研究的論文的列入和排除條件，這個流程可以減少差異性（heterogeneity）。
- 三、搜尋論文時要減少漏抓的疏失。要記住所有的相關資料庫（database）都要搜尋，才不會遺漏重要論文。

表1 各種review和meta-analysis的差別

	Narrative review	Systemic review	統合分析of published data	Pooled reanalysis
1. 準備詳細的study protocol及analysis plan	-	+	+	+
2. 文獻檢索 (inclusion criteria and exclusion criteria)	-	+	+	+
3. 計算pooled estimates, heterogeneity, 以及 sensitivity analysis	--	-	+	+
4. 分析各篇論文data	-	-	-	+

表2 卵巢癌腹膜腔灌注化學治療相關論文之描述性分析

第一作者 (年代)	研究代號	病人數目		IV配方	IP配方
		IV	IP		
Alberts (1996)	SWOG-8501/ ECOG/ GOG-104	279	267	Cisplatin 100mg/m ² IV + Cyclophosphamide 600 mg/m ² IV every 3 weeks for six cycles	Cisplatin 100mg/m ² IP + Cyclophosphamide 600 mg/m ² IV every 3 weeks for six cycles
Armstrong (2006)	GOOG-172	210	205	Paclitaxel 135 mg/m ² over 24 h IV day 1 + cisplatin 75 mg/m ² IV day 2 every 3 weeks for six cycles	Paclitaxel 135 mg/m ² over 24 h IV day 1 + cisplatin 100 mg/m ² IP day 2 + paclitaxel 60 mg/m ² IP day 8 every 3 weeks for six cycles
Gadducci (2000)	NWOG	57	56	Cisplatin 50mg/m ² IV + epidoxorubicin 600 mg/m ² IV + cyclophosphamide 600 mg/m ² IV every 4 weeks for six cycles	Cisplatin 50mg/m ² IP + epidoxorubicin 60 mg/m ² IV+ cyclophosphamide 600 mg/m ² IV every 4 weeks for six cycles
Kirmani (1994)	UCSD	33	29	Cisplatin 100mg/m ² IV + Cyclophosphamide 600 mg/m ² IV every 3 weeks for six cycles	Cisplatin 200mg/m ² IP + etoposide 350 mg/m ² IP every 4 weeks for six cycles
Markman (2001)	SWOG/ ECOG/ GOG-114	227	235	Paclitaxel 135 mg/m ² over 24 h on day 1 + cisplatin 75 mg/m ² IV day 2 every 3 weeks for six cycles	Carboplatin (AUC=9)IV every 4 weeks for two courses, followed 4 weeks later by paclitaxel 135 mg/m ² IV over 24h on day 1 + cisplatin 100mg/m ² IP on day 2
Yen(2001) (台北榮民總 醫院顏明賢醫 師)	Veterans General Hospital, Taipei(Taipei Study)	63	55	Cyclophosphamide 500 mg/m ² IV over1 h day 1 + adriamycin or epirubicin 50 mg/m ² over 1 h IV day 1 + cisplatin 50 mg/m ² IV every 3 weeks for six courses	Cyclophosphamide 500 mg/m ² IV over1 h day 1 + adriamycin or epirubicin 50 mg/ m ² over 1 h IV day 1 + cisplatin 100 mg/m ² IP rapid infusion every 3 weeks for six courses

摘自Hess LM, Benham-Hutchins M, Herzog TJ, et al.: Int J Gynecol Cancer 2007;17:561-570.

四、是要做統合分析還是要做已發表文獻的統合分析。如果你要做到pooled-reanalysis，那麼就要和各論文的研究者聯繫，請他們提供原始資料。

五、對收集到的論文要做描述性的表格。對於升級到的相關論文要作出一個描述性的表格，好讓讀者可以清楚這些論文的特點。表2舉例台北榮民總醫院婦產部與其他國家所發表的卵巢癌腹膜腔灌

注化學治療（IP）的統合分析的描述性表格（表2）

六、是否呈現個別論文的effect分析和合併效果（effect size）。對於每篇論文的效果（效果可以是連續型變項、勝算比、相對風險、存活資料等）以及合併效果是否有呈現在統合分析中？原則上這些效果的呈現是以森林圖（forest plot）

研究代號	統計值				P值
	HR	Lower limit	Upper limit	Z value	
SWOG 8501	0.760	0.606	0.953	-2.372	0.0177
NWOG	0.670	0.389	1.155	-1.441	0.1497
GOG 114	0.810	0.627	1.047	-1.609	0.1076
Taipei	1.130	0.688	1.855	0.483	0.6290
GOG 172	0.750	0.580	0.970	-2.193	0.0283
UCSD	1.240	0.621	2.475	0.610	0.5418
Random effects	0.799	0.702	0.910	-3.383	0.0007

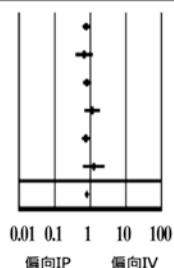


圖1 卵巢癌腹膜腔灌注化學治療的效果分析（森林圖）。摘自Hess LM, Benham-Hutchins M, Herzog TJ, et al. Int J Gynecol Cancer 2007;17:561-570.

呈現如圖1。森林圖在很多專門跑統合分析的軟體都可以自動幫你繪出森林圖，如comprehensive meta-analysis軟體，或RevMan等軟體。而如果是pooled-reanalysis則是要將整合後的檔案重新跑統計，常用的統計方法包括羅吉斯回歸分析（logistic regression）或線性迴歸分析（linear regression），關於回歸統計方法分析，日後會有專文另行介紹。

至於如何計算合併效果，目前有兩種重要的統計分析方法應用於統合分析中，第一種是固定式模型（fixed-effects model），第二種是隨機式模型（random-effects model）^(6,7)。這兩種分析方法的差異會在下一個段落詳述。

七、要考量論文間是否有存在異質性、是否有作敏感性（sensitivity analysis）、以及是否有出版性偏差（publication bias）的問題。

1.異質性的問題：要分析各論文間是否有很大的異質性，可以用卡方異質性檢定（chi-square test for heterogeneity, Cochran Q test）統計法來分析。分析異質性的目的是要看個別論文的效果和綜合性效果是否有極大的差異性，通常 p 值 < 0.1 我們就可以判定這些論文間存在有明顯的異質性。（8）異質性的原因來自於：（1）不同的病人群研究（patient population studies），（2）治療方法（interventions used），（3）附加治療（co-interventions），

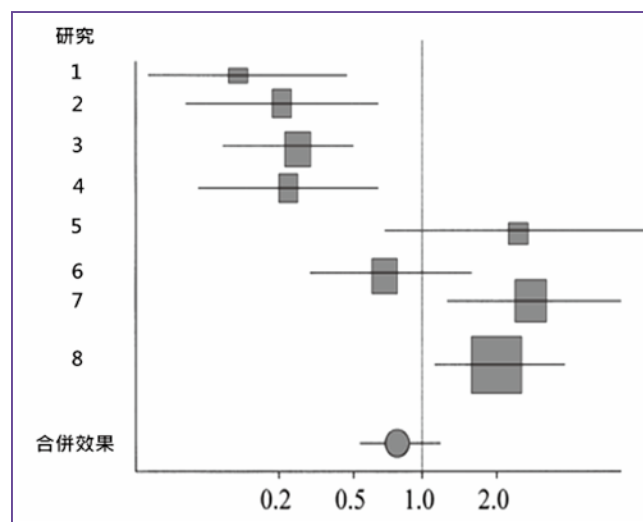


圖2 統合分析中呈現不同研究結果的明顯差異性

（4）結果評估方式（outcomes measured），（5）研究設計不同（different study design features），（6）研究品質（study quality），（7）隨機誤差（random error）。圖2就顯示出很明顯的異質性：

原則上如果出現了異質性，則不要先急著作統合分析，而是應該找出可能的原因，或是看看有那幾篇論文造成了異質性的產生，必要時可以再重新評估，將某些論文刪除後，再重新計算異質性的統計值。當我們遇到欲選取的論文存在有明顯的異質性時，目前有兩種方法解決：

A. 次群組分析（subgroup-analysis）：也就是將欲選取的論文，找出具有明顯的category 差別的變項，按照此變項的level分別作統合分析。因此就可能產生兩個或三個的森林圖。
B. 統合性迴歸分析（meta-regression）：原則上如果總論文數小於10篇以下，則不要作統合性迴歸分析。統合性迴歸分析的目的是在將某些變數當作共變數（covariates），去探索（explore）有哪幾個變項會造成異質性。

2. 敏感度分析（sensitivity analysis）：敏感度分析主要的目的是將某些不合適的論文（例如壁報或品質差的論文）刪除後，看看剩餘論文的

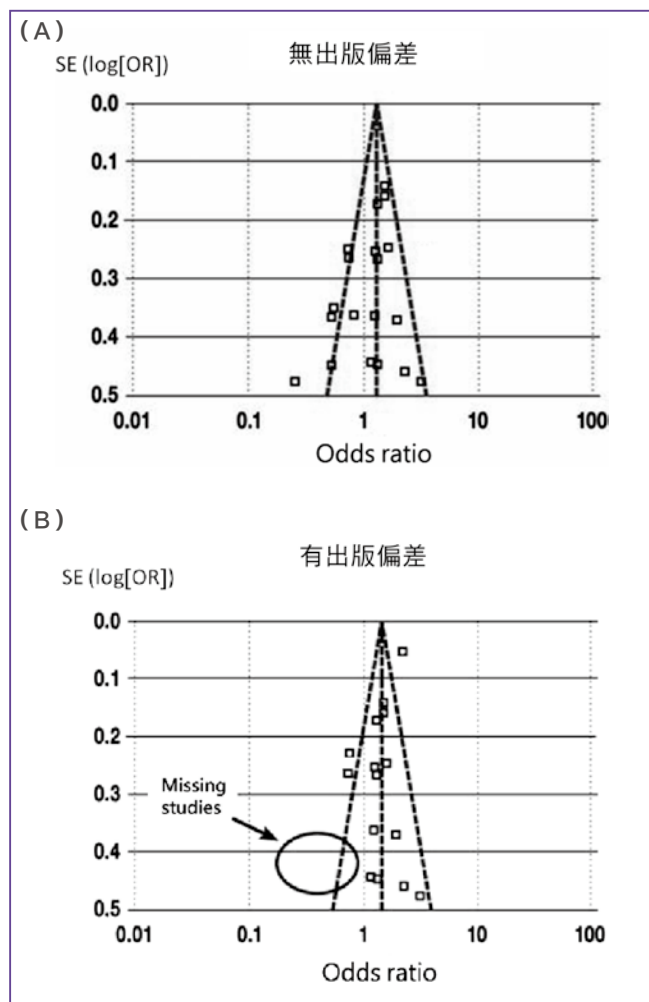


圖3 (A) 無出版性偏差，(B) 有出版性偏差

合併效果是否會因此更改，藉以測試綜合性效果的穩定度。如果某篇論文被刪除後，造成剩餘論文的綜合性效果明顯改變，那麼就應該在論文的討論部分說明此篇論文對於整體綜合性效果的重要性。

3. 出版性偏差 (publication bias) 的問題：要診斷統合分析所選取的論文有無出版性偏差的問題可以運用漏斗圖 (funnel) 來檢視⁽⁹⁾，圖3是一個漏斗圖的案例，(A) 代表沒有出版性偏差，而 (B) 則有出版性偏差。

八、你如何詮釋研究結果。在詮釋統合分析結果時對於所選取的每篇論文的品質，病人年齡分布，治

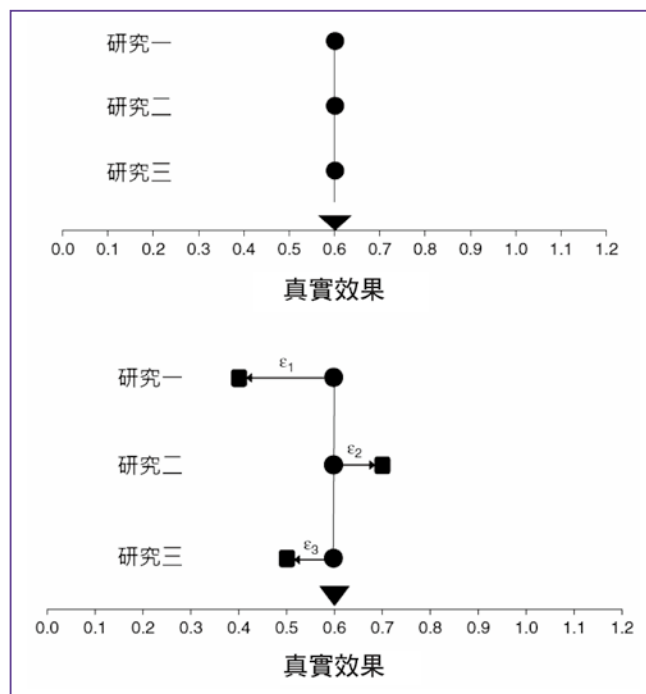


圖4 固定式模型，觀察效果 = 真實效果 + 誤差

療方式，給藥劑量，或追蹤長短，都要和其他論文比較，找出這些論文的相同或相異之處。

統合分析的統計方法

統合分析的統計方法有兩種，第一是固定式模型 (fixed-effects model)，而另一種則是隨機式模型 (random-effects model)，此兩種方法各有其假設 (assumption)。

一、固定式模型

固定式模型是假設所有的研究都有一個共通的真實效果 (true effect)，而每篇論文所觀察到的效果稱為觀察效果 (observed effect)，之所以每篇論文的觀察效果不同乃是導因於取樣誤差 (sampling error) (因為每篇論文的病人群不同、年齡分布不同、藥劑使用量不同、或是追蹤時間不同等)。由於我們假設這些論文都有相同的真實效果，因此當每篇論文的病人數目如果能增加到無限大時，則觀察效果會等於真實效果⁽¹⁰⁾。在此舉圖四為例，總共有三個研究列入統合分析的研究 (●：真實效果，■：觀察效果)，從圖中可以看出三個研究的true effect都是0.6，但是我們所

觀察到的effect卻是研究一：0.4；研究二：0.7；研究三：0.5，因此我們可以看出對於每個研究的觀察效果 $Y_i = \text{真實效果} + \text{誤差}$ ($Y_i = \text{true effect } (\theta) + \text{error } (\varepsilon_i)$)。(圖4)

由於在固定式模型中我們會為每篇研究給予一個權重 (weight)，公式如下：

$W_i = 1/V_{Y_i}$ ($V_{Y_i} = \text{within-study variance}$)，然後就可以算出加權綜合效果平均數 (weighted mean of summary effect size, M) 以及綜合效果變異量 (variance of the summary effect, V_M)

$$M = \frac{\sum_{i=1}^k W_i Y_i}{\sum_{i=1}^k W_i} \quad V_M = \frac{1}{\sum_{i=1}^k W_i}$$

而綜合效果標準誤 (standard error of the summary effect) 則是

$$SE_M = \sqrt{V_M}$$

而95%信賴區間 $M \pm 1.96 \times SEM$ 。最後用 Z-value 來測試虛無假設： $= \text{真實效果是否為0}$ (Z value = M/SEM)。單尾檢定統計量 $P = 1 - \Phi(|Z|)$ ；而雙尾檢定量為 $P = 2 \times [1 - \Phi(|Z|)]$ ， $\Phi(Z)$ 代表標準累積分布 (standard normal cumulative distribution)。

二、隨機式模型

隨機式模型的假設是每篇研究的真實效果都不同 (因為每篇論文的病人群不同、年齡分布不同、藥劑使用量不同、或是追蹤時間不同等)，這些因子的不同而造成了每篇論文的真實效果不同⁽¹¹⁾，我們可以從圖5A來看隨機式模型的概念：假設有三個研究，此三個研究的觀察效果和上述的固定式效果都一樣：研究一：0.4；研究二：0.7；研究三：0.5。(圖5)

我們舉研究三為例：研究三的真实效果 = 0.5，而 $\mu = \text{實際效果平均數}$ (grand mean of true effects)，而 $\zeta = \text{實際效果變異量}$ (true variation in effect size)，所以就隨機式模型而言 $Y_i = \mu + \zeta_i + \varepsilon_i$ 。務必要記住，在隨機式模型中，我們要測量的是所有真實效果的整體平均值，而95%信賴區間是代表我們對這個平均值的不確定性；相對的我們在固定式模型中所預測量的

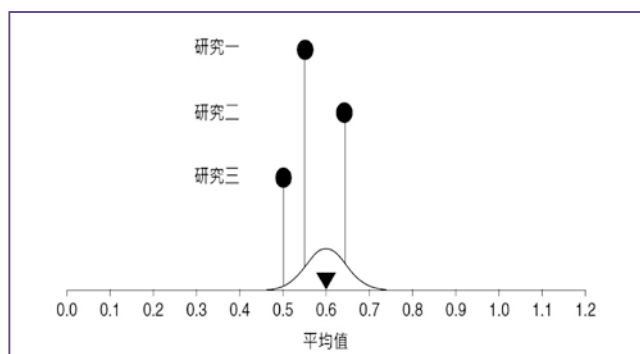


圖5A 隨機式模型 (每篇研究的真實效果都不同)

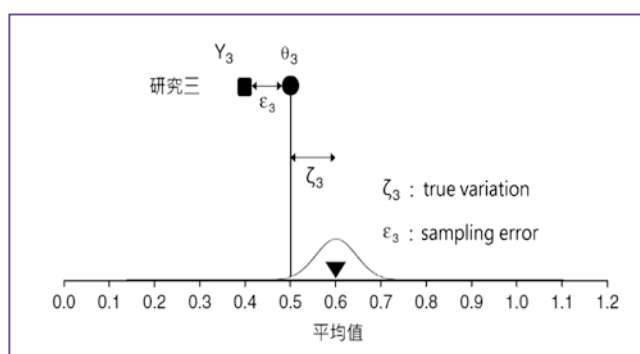


圖5B 隨機式模型的變異量 = 研究間變異量 + 研究內變異量

是真正的真實效果值。在固定式模型中只有研究內變異量 (within-study variance)，而在隨機式模型中則同時有研究內變異量以及研究間變異量。在隨機式模型一樣也會計算每篇論文權值 (W_i)，其中 T^2 就是指研究間變異量。

$$W_i^* = \frac{1}{V_{Y_i}^*}$$

$$V_{Y_i}^* = V_{Y_i} + T^2.$$

接下來就可以算全部論文的加權平均綜合效果為

$$M^* = \frac{\sum_{i=1}^k W_i^* Y_i}{\sum_{i=1}^k W_i^*}$$

而綜合效果變異量為

$$V_{M^*} = \frac{1}{\sum_{i=1}^k w_i}$$

而綜合效果標準誤則是

$$SE_{M^*} = \sqrt{V_{M^*}}$$

而95% 信賴區間 $M \pm 1.96 \times SE_{M^*}$ 。最後用Z-值來測試需無假設： $=$ 真實效果為0 ($Z = M/SE_{M^*}$)。單尾檢定統計量 $P = 1 - \Phi(\pm|Z|)$ ；而雙尾檢定量為 $P = 2 \times [1 - \Phi(|Z|)]$ ， $\Phi(Z)$ 代表標準累積分布。要查此函數的統計值可以查統計教科書，或是利用Excel的函數公式 = NORMSDIST (Z*)。

三、固定式模型和隨機式模型的區別

在前面已經解釋過，固定式模型是將所有論文的真實效果視為相同，而隨機式模型則將所有論文的真實效果視為不同。固定式模型會根據每篇論文的病人數目給予不同的權值，一篇論文的病人數目越大則給予的權值就越大，而病人數目少的研究則因它提供的訊息相對較差，所以就會給予較小的權值。

然而在隨機式模型中，由於每篇研究都有其獨特性，因此不能單以病人數目多寡大小來衡量權重，以下舉相同的一篇統合分析研究就可以看出固定式模型和隨機式模型的不同之處：從圖6的上圖是用固定式模型作的分析，可以看出每篇研究的黑色方格面積不盡相同，但是同樣的這些論文如果是用隨機式模型分析，可以看出黑色方格大小的差異性減少了（研究4的相對權重由39%降到23%），因此可以看出隨機式模型會將小型研究的權值調升。

四、如何選擇固定式模型或是隨機式模型來分析

目前認為如果卡方檢定沒有異質性的問題，就使用固定式模型來計算；而如果存在有異質性的問題則使用隨機式模型來計算合併效果，這其實是個錯誤的觀念。正確的觀念是：即使是沒有存在異質性，如果研究者認為研究彼此之間仍存在有異質性時，則仍可以使用隨機式模式；而相對的如果統計發現有明顯異質性時，則當然考慮使用隨機式模型，或者是作敏感

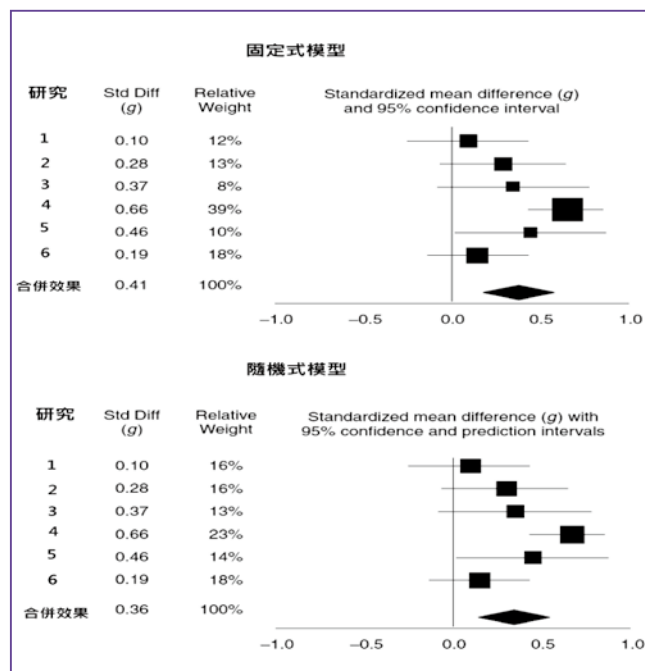


圖6 同一篇研究的固定式模型和隨機模型的相對權重差別（本圖為模擬之統合分析）

度分析，將不合適的論文剔除之後，再重新分析。另外研究者也要問自己，一旦出現異質性時，這些研究是否可合併？如果不能合併，那麼就應該試著排除一些論文後，再試著分析，另外一個思考就是如果這小型的研究資料無法合併時，此時就應該要執行大規模的研究以釐清這個問題。

統合分析的優缺點

統合分析的成就在於它為醫學研究帶來了許多前所未有的優點（當然它也有許多潛在的缺點），其優缺點如下：

- 一、統合分析了許多小型的研究：小型研究為人所詬病的就是病人數目不足，沒有足夠的檢定力（power）來檢測兩種治療方法的真實差異。早期幾例有名的統合分析研究，如beta-blockers對於心肌梗塞的二度預防效果研究⁽¹²⁾、steroid對於胎兒肺成熟度的影響⁽¹³⁾、和adjuvant tamoxifen對乳癌的影響⁽¹⁴⁾，都是統合了幾篇相關的小型研究，而使相關醫學議題獲得統合性的評估。統合

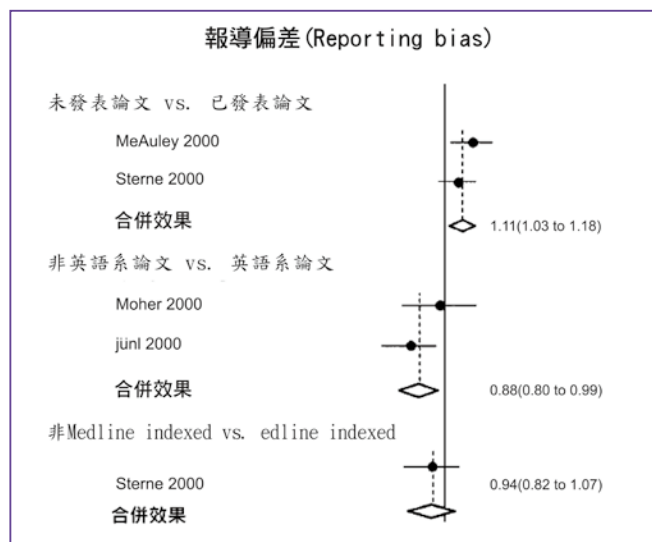


圖7 統合分析的報導偏差的影響

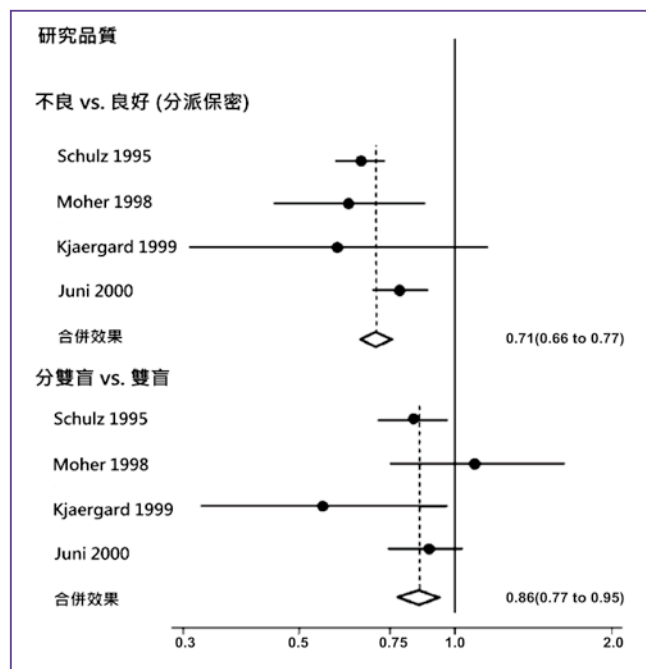


圖8 統合分析的研究品質對分析結果的影響

分析的發展也使西方醫學由教授的威權式個人的經驗傳承，進步到由較客觀公正的統計方法來針對特定醫學議題進行統合分析，由於分析的結果較客觀也較公正，因此也較經得起考驗，也較有機會在不同的醫院或國家傳播研究的結果，年輕醫師也可以藉由公正客觀的分析獲得某特定醫學議題的統合分析結果的知識。

- 二、系統性文獻回顧 (systemic reviews) 和統合分析可以點出具有爭議性的醫學題目：例如傳統在腸道手術後一般認為應該禁食，但是隨後的統合分析卻發現禁食有可能帶來較多缺點，因此這篇統合分析的結論建議針對此議題進行一個大規模 (large-scale) 研究，才能解決其爭議性⁽¹⁵⁾。
- 三、統合分析可以提供較有證據性的次群組分析 (subgroup analysis)：例如某一針對tamoxifen對於可手術性乳癌的55個研究的統合分析發現tamoxifen對於estrogen接受體陰性的病人而言是不具有效益的，因此這篇統合分析的研究結果影響到日後的乳癌病理檢驗必須提供estrogen接受體和progesterone 接受體的免疫染色報告⁽¹⁶⁾。
- 四、統合分析的研究結果絕非是完全值得信賴的：有很多的統合分析忽略了偏差 (bias) 的存在。由於統合分析的分析絕非盡善盡美，因此也有許多

研究指出統合分析的缺失。如圖7的統合分析研究就指出尚未發表研究結果會低估約10%的治療效果；而非英語系的論文會高估約12%的治療效果；而非medline index的論文則會高估約5%的治療效果。而圖8則指出如果所列入的論文含有不足夠或不明的治療分派保密 (concealment of allocation) 時，則會高估30%的治療效果；而沒有雙盲的研究會比有雙盲的研究高估了15%的治療效果。(圖7、8)

- 五、統合分析結果有可能和隨機分派研究 (randomized controlled trials, RCT) 的結果相抵觸：統合分析和隨機分派的結果相抵觸時有所聞，最有名的例子就是「鎂離子對於心肌梗塞」的研究，統合分析指出了鎂離子的有效性⁽¹⁸⁾，然而隨後的大型Fourth International Study of Infarct Survival (ISIS-4) 卻發現鎂離子是無效的⁽¹⁹⁾，當遇到統合分析和隨機分派結果相抵觸時，我們必須花時間去找出兩者間的研究差異性。隨後的探討發現ISIS-4的鎂離子灌注時機和統合分析的鎂離子灌注方法是不同的。另外一個重要的概念

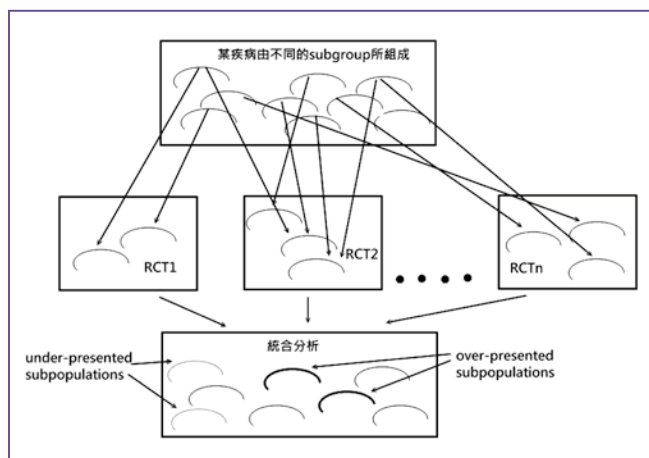


圖9 統合分析和隨機分派研究論文結果牴觸的可能原因: 某些特定族群被過度呈現 (over-presented)。

是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；相對的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。圖9指出了統合分析和隨機分派研究為何有時會有不同的結果，原因就是在於隨機分派研究之間會有異質性的存在，在作統合分析有可能會讓特定族群過度呈現 (over-presented)。

結語

統合分析研究方法的問世帶給了醫學界很大的衝擊，雖然此研究方法可以為某特定醫學主題帶來系統性的研究，而且也應用了客觀的統計法來計算合併效果，然而統合分析結果的可信度是取決於所選取論文的研究品質。所謂的垃圾進，垃圾出 (garbage in, garbage out) 就是指如果所選取讀論文品質是差的，那麼統合分析的品質當然也就差。我們在看任何一篇論文都應該要有能力來分析出該篇論文的品質好壞，而非只是無條件的全然接受論文的研究結果。

統合分析的研究方法仍持續的進展，多數統合分析是分析多篇隨機分派研究的結果，但是目前已有多篇統合分析朝向分析觀察性研究 (observational

studies)。由於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。

參考文獻

1. Resning M, Blettner M, Klug SJ: Systemic literature reviews and meta-analysis. Dtsch Arztebl Int 2009;106:456-463.
2. Plackett RL: Studies in the history of probability and statistics: VII. The principle of the arithmetic mean. Biometrika 1958;45:130-135.
3. Pearson K: Report on certain enteric fever inoculation statistics. Br Med J 1904;3:1243-1246.
4. Glass GV: Primary, secondary and meta-analysis of research. Educ Res 1976;5:3-8.
5. Bero L, Rennie D: The Cochrane Collaboration. Preparing, maintaining, and disseminating systemic reviews of the effects of health care. JAMA 1995;274:1935-1938.
6. DerSimonian R, Laird N: Meta-analysis in clinical trials. Control Clin Trials 1986;7:177-188.
7. Normand SL: Meta-analysis: formulating, evaluating, combining, and reporting. Stat Med 1999;18:321-359.
8. Higgins JP, Thompson SG: Quantifying heterogeneity in a meta-analysis. Stat Med 2002;21:1539-1558.
9. Dwan K, Altman DG, Amaiz JA, et al.: Systemic review of the empirical evidence of study publication bias and outcome reporting bias. PLoS ONE 2008;3:e3081-3085.
10. Thompson SG, Sharp SJ: Explaining heterogeneity in meta-analysis. A comparison of methods. Stat Med 1999;18:2693-2708.
11. Viechtbauer, W. 2006 MIMA: an S-Plus/R function to fit meta-analytic mixed-, random-, and fixed effects models [Computer software and manual]. See <http://www.wvbauer.com/>
12. Yusuf S, Peto R, Lewis J, et al.: Beta blockade during and after myocardial infarction: an overview of the randomized trials. Prog Cardiovasc Dis 1985;17:335-371.
13. Crowley P: Corticosteroids prior to preterm delivery. In: The Cochrane Library, Issue 1, 1999, Oxford: update software.

14. Early Breast Cancer Trialists' Collaboration Group: Effects of adjuvant tamoxifen and of cytotoxic therapy on mortality in early breast cancer. An overview of 61 randomized trials among 28,896 women. *N Engl J Med* 1988;319:1681-1692.
15. Lewis SJ, Egger M, Sylvester PA, et al.: Early enteral feeding versus 'nil by mouth' after gastrointestinal surgery: systemic review and meta-analysis of controlled trials. *Br Med J* 2001;323:773-776.
16. Early Breast Cancer Trialists' Collaboration Group: Tamoxifen for early breast cancer: an overview of the randomized trials. *Lancet* 1991;338:1127-1130.
17. Juni P, Holenstein F, Sterne J, et al.: Direction and impact of language bias in meta-analyses of controlled trials: empirical study. *Int J Epidemiol* 2002;31:115-123.
18. Teo KK, Yusuf S, Collins R, et al.: Effects of intravenous magnesium in suspected acute myocardial infarction: overview of randomized trials. *Br Med J* 1991;303:101-105.
19. Collaborative Group: ISIS-4. A randomized trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58,050 patients with suspected acute myocardial infarction. *Lancet* 1995;345:669-687.
20. O' Rourke K, Detsky AS: Meta-analysis in medical research; strong encouragement for higher quality in individual research efforts. *J Clin Epidemiol* 1989;42:1021-1024.

