

# 間接比較各種隨機分派臨床試驗之治療效果 — 一種擴大利用實證資料的分析方法

## Indirect Comparison Of Treatment Effects from Randomized Controlled Trials - Utilize Evidence to Compare Effects of Different Treatments

邵文逸

財團法人醫藥品查驗中心、台灣大學臨床醫學研究所

### 摘 要

比較不同治療的效果是臨床上常見的問題，但是直接比較不同治療之效果的實證資料卻不一定存在。間接比較方法將原本針對不同治療的各別研究，做進一步的整合，透過各研究之間的共通對照組，間接比較不同治療之效果的差異，可有效利用現有的實證資料，是一項可行的替代方法。本文簡要的介紹間接比較的分析方法，以及應用該分析方法時應該注意的重要問題。

### 前 言

由於醫藥生物科技的進展，針對相同疾病的治療藥品，或治療方法都有日益增多的趨勢。如何從眾多治療中找出最適合某位病人的選擇，臨床上有許多考慮的重點，其中之一是看看有沒有臨床實證資料，比較不同治療之間效果的差異。以新藥 A 與既有藥物 B 治療效果的比較為例，臨床上關心的治療結果可包含療效、安全性、病患滿意度等具有臨床意義的指標。依照實證醫學執行方法，比較的評估從 PICO — 病患 (patient)、治療方式 (intervention)、比較對象 (comparison)、

結果指標 (outcome) 等層面開始，把臨床問題清楚界定，接著有系統的蒐集相關實證資料，再用正確有效的統計方法整合所收集到的資料。當找到足夠的實證資料時，該如何整合這些資料？

### 可用的實證資料及其證據強度

直接比較不同治療方法的隨機分派臨床試驗，以及整合這類研究的系統性回顧 (systematic review)，是目前普遍認為最好品質的實證資料。因為這些研究透過妥善的設計，將可能干擾研究結果的其他外界因素盡量排除，確保研究所觀察到，治療效果的不同，是研究所針對治療方法的

不同直接造成。

另一類常見的實證資料，來自經過適當設計，有系統收集觀察性資料的研究，以及其系統性回顧。這一類研究資料來自日常的臨床醫療，將接受不同治療病患的資料分組整理，分析治療效果是否與病患所接受的治療方法有關係。這類研究較為貼近臨床現實的治療狀況。但是在現實狀況下，病患所接受的治療方式、治療結果，經常受到外界其他因素的影響。例如，病情嚴重程度可能影響所選擇的治療，也同時影響病情的發展，此時所觀察到治療方法與治療結果的關係，很可能不是治療的真正效果。因此研究中所觀察到的結果，與真正治療效果有所差別，這樣的差別稱為研究的偏差（bias）。研究有嚴重偏差時，我們也稱該研究的效度（validity）不好。

除了研究可能的偏差外，研究也可能因為隨機誤差（random error），出現與真正效果不同的結果。這個問題需要透過統計分析來處理，常見的策略是增加較多的研究樣本，以減少研究的隨機誤差，有關研究隨機誤差的說明，讀者可參考本系列之前的報告[1]。研究者可以在一個獨立的研究中，收集較多的研究樣本；或者也可以彙整多個類似研究，透過整合各研究結果，將隨機誤差彼此相互平衡消除而減少，這就是統合分析（meta-analysis）。不過如果拿來做整合的研究存在某些固定的偏差，或是具有某些特定偏差性質的研究才被選進來，那麼這些研究整合在一起的

結果也會有偏差。因此需要一套系統性的方法，確保所收集研究的效度都是夠好，且研究收集的過程也是客觀公正，這樣的過程就是系統性回顧的基本要求。

如同系統性回顧跟統合分析把多個獨立研究整合在一起，我們也可以把原本針對不同治療的不同研究整合在一起做比較，透過一個間接的方式，採用有效的統計方法，比較多個臨床試驗的結果。在缺乏直接比較實證資料的情況下，也可提供不同治療方法的效果是否有差別的參考。這樣的作法稱為間接比較方法，是統合分析方法的一種延伸。

### 間接比較該怎樣比

比較治療效果的研究，通常作法是將病人分組，給予不同的治療（例如用 A 與 B，代表兩種不同治療），治療後觀察某項特定的效果，稱為研究終點（end point）或是結果（outcome）是否有差異。例如高血壓治療的研究，可以用血壓降低的數值；也可以用病患追蹤治療一段時期的腦中風發生率，做為研究的結果。我們用  $Y_A$  與  $Y_B$  來分別代表兩組不同治療的結果。

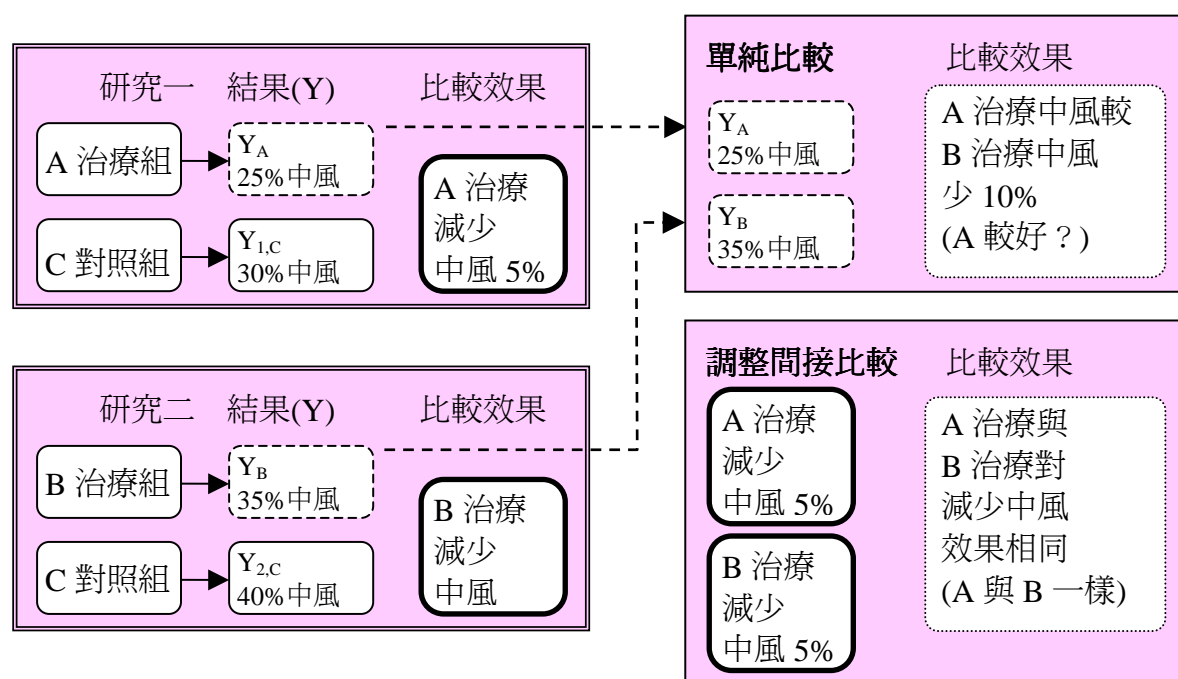
在這樣的比較當中，最重要的一點是，分派到接受 A 治療的病患與分派到 B 治療的病患，除了治療方法不同外，其他各種可能影響結果的因素必須完全相同，這個條件稱為研究分組間的可比較性（comparability）。良好的可比較性，是探討治療效果差別最重要的一項原則。透過大樣本隨機分派（randomization），可以將

具有各種不同因素的病人，相當程度的公平分配到各研究群組，以達成良好的可比性。在這樣的條件下，直接比較治療結果  $Y_A$  與  $Y_B$ ，就可以代表兩種治療效果的差異。分析時用  $Y_A - Y_B$  來表示兩者差異的大小，並檢定這個差異是否達到統計顯著。

許多時候，治療 A 與治療 B 並沒有直接比較研究，不過卻有各自的臨床研究。最常見的情況是，A 治療與安慰劑比較，且 B 也與安慰劑比較，各別皆為隨機分派臨床試驗。這是因為現今大部份藥品上市

前，多數須先完成此類臨床試驗，以符合各國的法規要求。

這些研究提供了一些關於 A 治療與 B 治療的實證資料。如此一來，是不是可以把第一個研究中，接受 A 治療病人的結果，單獨拿出來跟第二個研究中，接受 B 治療的結果直接比較呢？這樣的比較稱為單純比較 (naïve comparison)，在一些報告或討論中可能會看到這樣的分析。但是從研究方法的角度來看，這樣的分析並不恰當，有可能得到不正確的結果，請看圖一。



圖一 單純比較與調整間接比較

### 單純比較所面臨的問題

直接把不同研究的各組結果，分離出來做單純比較，面臨的問題在於，可能有些影響治療結果的因素，在不同研究之間

的分佈是不一樣的。例如，各研究選取病患的條件不同；執行研究的醫療單位，照護病人的條件不同；研究在不同地區執行時，該地區的疾病流行背景不同，……等

問題。所以，將不同研究中的各組拿出來，這樣的比較很可能不具有良好的可比較性，分析結果將發生偏差。例如，圖一的例子，研究二，兩組病人發生中風的比例都高於研究一，很有可能研究二的病人群發生中風的背景風險，原本就比研究一高。直接把第一個研究的 A 組治療結果，跟第二個研究的 B 組治療結果拿來相比，兩種治療效果的差別，並不見得是治療方法不同所造成。

### 經調整的間接比較

(adjusted indirect comparison)

上面的例子中，若兩項研究都是經過適當設計，且正確地執行，具有良好研究品質，那麼這兩項研究「內部」的比較結果是正確有效的。也就是說 — A 治療相對於對照治療 C 可減少中風 5%；B 治療相對於治療 C 也一樣可以減少中風 5%，都是正確可信的。

如果以上兩項研究，都是規模夠大的隨機分派臨床試驗，此時我們可以認為，跟中風有關的其他因素，包括各種已知，或是其他未知，但是會造成中風的風險因子，透過隨機分派，在同一個研究內的比較組之間達到分配平衡，因此所觀察到「治療組中風比對照組少 5%」的結果，可有效的完全歸因於治療方法的不同，不受到其他因素的影響。在這樣的條件下，當兩項研究都使用相同的 C 治療當對照，稱為共通對照組 (common comparator)，A 與 B 的相對治療效果，可以透過這個相同的

對照來做間接比較，並排除研究內其他因素的影響。

這樣的分析，透過共通對照當作參考點，保留原本研究內的可比較性，調整了可能影響結果的其他因子，再間接的比較 A 與 B 治療效果的相對差異，稱為調整後間接比較。在這樣的分析裡面所用的共通對照，可以是臨床試驗中的安慰劑對照組 (placebo control)，也可以是具有積極療效的其他治療對照 (active control)。

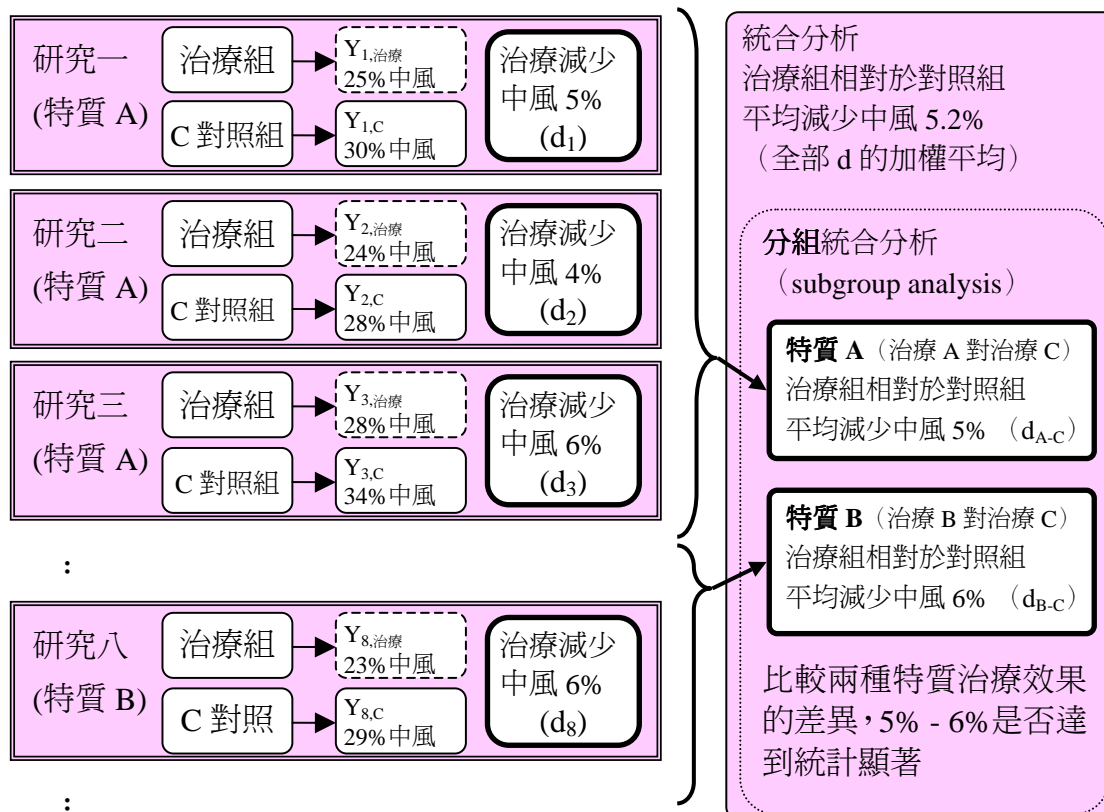
### 間接比較方法與統合分析

(meta-analysis)

統合分析是將數個獨立研究結果做整合的統計分析方法，是現今整合既有實證資料常用的分析方法。正確的統合分析，將各研究的比較結果 (用  $d$  代表，例如第一項研究的比較結果： $d_1 = Y_{1,A} - Y_{1,C}$ )，依照各研究統計上的重要性給予加權，再做平均，計算出一個整合的治療效果，請參考圖二。

這樣的分析保留各研究內設計好的可比較性，維持原本各個研究內部，比較組的有效比較結果。只有這樣分析出來的治療效果差異( $d$ )，才能有效避免分析結果的偏差。

如果將各研究打散，把所有的治療組先混合在一起，各研究的對照組另外混成一組，然後比較。此時原本經過設計，各研究內的可比較性被打散，打散後的分析可能跟原本各研究結果不同，這樣的情況稱為辛普森悖論 (Simpson's paradox) [2]。



圖二 統合分析與分組統合分析

請看表一的例子：原本兩個研究，治療組發生中風的比例都比對照組低（減少 3.0%），也就是說治療組的效果都比較好。可是，如果把兩個研究的治療組與對照組人數先分別合併，再分析，結果卻發現，治療組發生中風的比例高過對照組（增加 3.5%），有治療反而變成效果比較不好。這種改變後的結果不具有良好的可比性，成爲一個受到干擾的分析結果，是一種研究上的偏差。因此需要透過正確的統合分析，將原本兩個研究的結果作適當的整合，以排除這種偏差。

間接比較跟一般的統合分析一樣，依循相同的原理，分析時維持各研究內的有

效比較，再整合不同研究的結果，可說是統合分析方法的延伸。

進行間接比較時，每組治療對照比較（A 對 C、或 B 對 C）並不限定各來自一個獨立研究。也可以將多個同爲 A 與 C 比較的臨床試驗，先做統合分析，得到一個整合的治療效果；再與多個 B 與 C 比較的臨床試驗統合分析結果，進行間接比較。

考慮研究來源的客觀公正，做間接比較之前，最好事先擬定分析計畫，有系統，儘可能完整地收集相關主題的所有研究，並檢查所收集研究的品質，然後再做分析，是比較妥善的作法。這樣的過程跟一般的系統性回顧遵循相同的原則。

表一 辛普森悖論的例子

		病人數	治療後中風人數	中風比例 (Y)	比較效果 (d)
研究一	治療組	120	24	20.0%	<b>-3.0%</b> (治療組較好)
	對照組	200	46	23.0%	
研究二	治療組	50	26	52.0%	<b>-3.0%</b> (治療組較好)
	對照組	20	11	55.0%	
合併	治療組	170	50	29.4%	<b>3.5%</b> (治療組較差?!) 受干擾的分析結果
	對照組	220	57	25.9%	

### 間接比較與統合分析的異質性檢定

(test for heterogeneity)

在探討治療效果的系統性回顧中，常見到研究者把所彙整的研究，依照不同性質歸類，再分析不同分類之間的比較結果 (d) 是否不同，稱為次群組分析 (subgroup analysis)。常見一些可能影響臨床治療效果的因素，被用來做為分組依據，包括：

- (1) 研究收納病人群的性質。例如：有的研究針對較年輕族群，有的則是老年族群，可以分析不同年齡的病人，接受治療後的效果是否不同；分別分析不同疾病嚴重程度的病人，在相同治療下效果是否有差別；採用不同診斷條件所診斷出來的病人，接受治療後的結果是否不同；……。
- (2) 與研究設計相關的因素。例如：根據研究是短期追蹤或長期追蹤分類；整合同一類藥理機轉藥品的系統性回顧，進一步將不同成分藥品分組歸類，看看不同成分藥品的治療效果是否有差別；……。

別；……。

如果發現各類研究結果不盡相同，稱為研究的異質性 (heterogeneity)。在統合分析方法裡，有標準的分析方法來檢查研究的異質性。執行異質性檢定的理由是：即使是完美，沒有偏差的理想研究，研究觀察仍然或多或少受到一定程度隨機誤差的影響，以致於各研究結果間，必然有些許差別存在。異質性檢定的目的，就在於檢查這些研究結果之間的差異，是否可以用純粹的隨機來合理解釋。如果研究之間確實存在一些影響結果的不同條件，會造成結果的大幅度差別，此時純粹用隨機性來解釋這樣大幅度的差異，這種差異出現的機會將會很低，也就是 p 值很小，因此異質性檢定呈現統計顯著。表示這些研究之間結果的差異，可能不適合用純粹的隨機性來解釋，而是受到某些重要因素影響。

當進一步將研究依照不同性質歸類分組，如果某個性質是造成各研究結果差異的主要來源，那麼分組歸類之後，同一類

似屬性研究之間的差異將大幅減少，也就是同一組內的異質性將明顯減少。但是，如果這個分組因素不是造成研究結果差異的原因，那麼分組之後，同一組裡的各研究仍將呈現相當程度的差別。因此利用這個原理，可以檢定這個分組因素，對治療結果的影響是不是達到統計顯著。例如圖二的例子，特質 A 那群研究的治療組接受的是 A 治療，而特質 B 的研究代表 B 治療，那麼根據 A 與 B 特質所做的次群組分析，其實也就是 A 與 B 療效差異，透過共同對照組 C，所做出來的間接比較。Bucher 等學者最早在 1997 年提出，將這個原理應用在間接比較分析 [3]。

### 調整後間接比較分析方法

以下是 A 與 B 間接比較分析的一種計算方法[4]：將 A 與 C 比較的統合分析結果  $d_{A-C}$ ，與另一組 B 與 C 比較的統合分析結果  $d_{B-C}$ ，再進行比較，得到

$$(d_{A-C}) - (d_{B-C})，$$

就是 A 與 B，透過 C 當作共通對照的間接比較結果 ( $d_{\text{indirect,A-B}}$ )。當原來個別研究結果用平均值的差異來呈現時，間接比較的結果也是以平均值的差別來表示。例如：中風降低的比例；治療後血壓平均降低的毫米汞柱 (mmHg)；……等。例如：

第一組研究 A 藥品治療病人血壓平均降低 12mmHg，其對照 C 治療病人血壓平均降低 2mmHg， $d_{A-C} = 10\text{mmHg}$ ；

第二組研究 B 藥品治療病人血壓平均

降低 18mmHg，其對照 C 治療病人血壓平均降低 5mmHg， $d_{B-C} = 13\text{mmHg}$ 。則 A 治療與 B 治療透過 C 當作共通對照，間接比較的結果如下

$$\begin{aligned} d_{\text{indirect,A-B}} &= 10 - 13 \\ &= -3， \end{aligned}$$

B 治療相對於 A 治療，病人平均血壓多降低 3mmHg。

### 比值 (ratio) 研究結果的間接比較

另一類臨床研究常用的分析，採用相對比值作為結果。例如勝算比 (odds ratio, OR)，可以將各組比較的整合分析結果， $OR_{A/C}$  以及  $OR_{B/C}$ ，再比一次

$$OR_{A/C} \div OR_{B/C}$$

就可以得到 A 與 B 間接比較的勝算比分析結果 ( $OR_{\text{indirect,A-B}}$ )。同樣原理適用在風險比 (risk ratio, RR)、危害率比 (hazard ratio, HR) 的分析。這類採用比值的分析，統計上習慣先取自然對數後相減

$$\begin{aligned} \ln(OR_{A/C} \div OR_{B/C}) \\ = \ln(OR_{A/C}) - \ln(OR_{B/C})， \end{aligned}$$

此時各研究整合後的結果分別為

$$\begin{aligned} d_{A-C} &= \ln(OR_{A/C})；d_{B-C} = \ln(OR_{B/C})，而 \\ d_{\text{indirect,A-B}} &= \ln(OR_{A/C}) - \ln(OR_{B/C})。 \end{aligned}$$

因此基本上，調整後間接比較分析，就是把兩組分別比較的研究結果，先整合，再相減。將這個對數結果取其指數，就可以算出勝算比

$$OR_{\text{indirect,A/B}} = \exp(d_{\text{indirect,A-B}})。$$

### 間接比較結果的方向性

整合各研究比較結果時，要注意各分

組比較的正負號方向。共通對照 C，在兩組比較裡面都是當作對照，應放在減號後面，或是分母的部份。這樣 C 所代表的治療效果，才會在間接比較中消去。當 C 不是安慰劑對照，而是其他的積極治療時，在原本的研究報告裡，可能會放在分組比較的減號前方，或分子的部份，如果不小心拿出來作間接比較，會變成這樣的結果：

$$\begin{aligned} & d_{A-C} - d_{C-B} \\ &= (Y_A - Y_{1,C}) - (Y_{2,C} - Y_B) \\ &= Y_A + Y_B - Y_{1,C} - Y_{2,C} \end{aligned}$$

這樣就沒有辦法正確計算出 A 與 B 效果的差異。

除了用相減（或相除）的方法之外，正確方向的分析也可安排如下：

$$\begin{aligned} & d_{A-C} - d_{B-C} \\ &= (Y_A - Y_{1,C}) - (Y_B - Y_{2,C}) \\ &= (Y_A - Y_{1,C}) + (Y_{2,C} - Y_B) \\ &= d_{A-C} + d_{C-B} \end{aligned}$$

這樣的安排，把共通對照放在兩個主要比較治療的中間，可以讀成 A 對 B 的相對效果等於「A 對 C 的效果“加上”C 對 B 的效果」。

間接比較的共通對照組，並不一定只限定用一種治療。當比較透過多個共通對照時，以上使用加法的安排閱讀起來較方便。例如：A 與 B 透過 X、Y、Z 三種不同的治療做間接比較，可以這樣分析

$$d_{\text{indirect,A-B}} = d_{A-X} + d_{X-Y} + d_{Y-Z} + d_{Z-B}$$

其中  $d_{A-X}$ 、 $d_{X-Y}$ 、 $d_{Y-Z}$ 、 $d_{Z-B}$  為各組治療，兩兩直接比較的研究結果。

做間接分析的時候，需注意各研究分

組的比較方向，與間接比較所用的加減號是否正確的配合。

## 間接比較分析的抽樣誤差

(standard error, SE)

除了 A 與 B 治療效果差異的大小跟方向，我們還要考慮研究的隨機誤差，才能分析間接比較結果是不是具有統計顯著意義。由於用來分析間接比較的各組結果，分別來自獨立的研究，根據統計原理， $(d_{A-C}) - (d_{B-C})$  的變異數 (Variance，用  $V^2$  表示)，等於兩組各別變異數的和。而抽樣誤差又等於變異數除以分析樣本數 (研究個數，以  $k$  表示) 之後的平方根，根據這樣的原理，間接比較結果的抽樣誤差可以這樣分析

$$\begin{aligned} & SE_{\text{indirect,A-B}} \\ &= \sqrt{V^2_{A-C}/k_{A-C} + V^2_{B-C}/k_{B-C}} \end{aligned}$$

如果兩組比較都是經過統合分析的結果，那麼研究個數已經在個別的統合分析裡考慮過，此時間接比較抽樣誤差可以簡化成

$$SE_{\text{indirect,A-B}} = \sqrt{(SE_{A-C})^2 + (SE_{B-C})^2}$$

$SE_{A-C}$  與  $SE_{B-C}$  為各組比較，統合分析結果的標準誤差。

## 間接比較分析的統計顯著檢定

(statistical test)

有了抽樣誤差，就可以做間接比較效果的統計顯著性檢定，計算方法一般原則如下，



$$z = d_{\text{indirect,A-B}} / SE_{\text{indirect,A-B}},$$

$z$  為標準常態分數，與標準常態分配比較 (standard normal distribution)，當  $z$  小於 -1.96 或大於 1.96 時， $d_{\text{indirect,A-B}}$  與 0 的差異達統計顯著。檢定結果  $d_{\text{indirect,A-B}}$  不等於 0，表示接受比較的 A 與 B 兩組治療效果不相同。

### 間接比較分析的信賴區間

(confident interval)

A 與 B 治療效果差異間接比較的 95% 信賴區間，計算方法如下

$$d_{\text{indirect,A-B}} \pm 1.96 \times SE_{\text{indirect,A-B}}。$$

### 文獻上間接比較分析實例

Lim 等學者於 2003 年發表，比較低劑量與中劑量阿斯匹靈 (aspirin) 治療，對接受過冠狀動脈血管繞道手術後病人，預防其血管再阻塞的間接比較分析[5]。經系統性回顧，收集到 2 個中劑量以及 3 個低劑量阿斯匹靈治療比上安慰劑對照組，隨機分派臨床試驗研究報告。統合分析整合結果，中劑量治療比上對照組，發生再阻塞的相對風險為 0.55 倍；低劑量比上對照組，再阻塞風險則為 0.74 倍。兩種治療都能夠降低再阻塞的發生，效果均達統計顯著。利用上面的結果，間接分析中劑量與低劑量治療的相對效果如下，

$$\begin{aligned} RR_{\text{indirect}} &= \exp \{ \ln(0.55) - \ln(0.74) \} \\ &= 0.74, \end{aligned}$$

95% 信賴區間為 (0.52, 1.06)，統計檢定  $p$  值為 0.10。中劑量組發生再阻塞風險，為

低劑量組的 0.74 倍，較為降低，但是沒有達統計顯著。

這樣的分析，探討低劑量與中劑量阿斯匹靈，預防手術後血管再阻塞效果是否不同，分析上也稱為，劑量與治療效果的交互作用 (treatment by dosage interaction)。如果把以上分析的不同劑量，換成不同藥品，就可以用來比較不同藥品的治療效果。

例如，Wells 等學者 2006 年發表，系統性回顧三種不同成分雙磷酸鹽類藥物 (bisphosphonate)，預防停經後婦女，因骨質疏鬆所引起骨折的研究[6]。利用這個系統性回顧及統合分析結果，進一步做間接比較，評估 alendronate 與 etidronate，對於非脊椎骨折的預防效果[7]。該回顧收集了五個 alendronate 與安慰劑對照的臨床試驗，統合分析結果 alendronate 減少非脊椎骨折風險，相對危險比 (relative risk, RR) 為 0.84，95% 信賴區間為 (0.74, 0.94)；七個 etidronate 與安慰劑對照臨床試驗，統合分析結果，etidronate 比上安慰劑組，發生非脊椎骨折風險的相對危險為 0.95，95% 信賴區間為 (0.66, 1.36)。系統性回顧結果，雖然沒有發現 alendronate 與 etidronate 的直接比較研究，利用間接比較分析，比較兩種治療對非脊椎骨折發生風險比為 0.88，95% 信賴區間為 (0.60, 1.29)，統計檢定  $p$  值為 0.79，沒有達到統計顯著差異。

間接比較分析與直接比較研究，統計分析上的差異

Glenny 等學者就「直接比較研究」與「間接比較分析」的抽樣誤差作了以下說明 [4]：當各項研究數目相同、設計與結果也類似，例如，總共收集到 10 個 A 與 B 的直接比較研究，另外各有 10 個 A 與 C 的比較，以及 10 個 B 與 C 的比較研究，並且各種比較結果的變異數都相同 ( $V^2 = V^2_{A-C} = V^2_{B-C}$ )，整合 10 個直接比較研究結果，抽樣誤差為

$$SE_{\text{direct,A-B}} = \sqrt{\frac{V^2}{10}} ;$$

而間接比較的誤差大小為

$$SE_{\text{indirect,A-B}} = \sqrt{\frac{V^2}{10} + \frac{V^2}{10}} = \sqrt{2 \frac{V^2}{10}} ,$$

後者是直接比較抽樣誤差的  $\sqrt{2}$  倍。

根據這樣的原理，需要各有 20 個 A 與 C 以及 B 與 C 的研究，間接比較的抽樣誤差，才會跟 10 個直接比較研究整合的誤差相同，

$$\begin{aligned} SE_{\text{indirect,A-B}} &= \sqrt{\frac{V^2}{20} + \frac{V^2}{20}} = \sqrt{2 \frac{V^2}{20}} \\ &= \sqrt{\frac{V^2}{10}} . \end{aligned}$$

當間接比較跟直接比較的真正效果一致時 ( $d_{\text{direct,A-B}} = d_{\text{indirect,A-B}}$ )，採用間接比較分析，需要比相同規模的直接比較，多四倍的研究個數，才能得到相同的統計結果。因此相對來說，間接比較分析較不容易達到統計顯著，統計檢定力比較低，信賴區間的範圍也較寬。也就是說，間接比較分析的不確定性較高，這是間接比較分

析的一個弱點。並且，兩個主要比較治療之間，所經過的共通對照越多，分析的誤差也會隨之增加。

以上所介紹，是間接比較分析中較簡單的一種方法，採用頻率學派固定效果分析的觀點 (frequentist, fixed effect model)，是目前實務上應用較為普遍的方法。加拿大藥品與科技評估機構 (Canadian Agency for Drugs and Technologies in Health, CADTH)，提供了一套簡易的軟體 ITC (Indirect Treatment Comparison)，供上述間接比較分析。可由該機構網站下載 (<http://www.cadth.ca>)。有興趣的讀者可進一步參考該單位於 2009 年所發表，各種間接比較統合分析方法的整理[8]。

## 使用間接比較分析時 需注意的問題

### 收納研究的效度 (validity)

效度也就是研究的可比較性，是首先需注意的問題。可分為兩個層面來討論：1、各研究本身的內部效度；2、研究之間的外部效度。

1. 研究的內部效度 (internal validity)，各研究內比較群組之間的可比較性：

內部效度指的是各研究本身的品質。要確保統合分析結果的可靠，需先確認所收納的每項研究，都具有良好的研究品質，沒有嚴重的偏差。因此執行系統性回顧時，所收納每項研究結果的內部品質都

須詳細評讀 (critical appraisal)，以評估可能造成研究結果偏差的各項因素 (risk of bias)，對各研究結果的可能影響，評估的重點包括：

- (1) 研究內的分組是否客觀公平，包括隨機分派的過程跟分派結果，比較分組之間，各種可能影響治療結果因素的分佈是否相當；
- (2) 除了研究要比較的治療之外，其他可能影響結果的醫療照護條件，包括研究治療的調整、以及研究所容許或不容許的其他共同治療，在比較分組之間是否相當；
- (3) 研究過程病患接受治療的配合度與追蹤的完整性是否良好且各組相當；
- (4) 各比較分組研究結果的測量，需採用相同客觀公正的有效方法，所獲得資料的品質也需一致；……等等

2. 研究間的外部效度 (external validity)，不同研究結果的可比較性：

彙整各獨立研究結果時，還需要注意各研究的相似程度，整合在一起的結果需要有足夠的臨床相似性，針對不同病人特質的研究，結果必須類似才能整合在一起。將不同特質，且結果不一致的研究，用統合分析勉強地算出一個共同的結果，可能會掩蓋了真正的治療效果，也可能誤導正確的治療方向。

例如，一個治療用在病情較嚴重的病患正面的療效 (治療反應率增加 12%)，但是用在病情較輕微的病患治療效果不明

顯 (治療反應率只增加 2%)，且病患有可能受到治療的不良副作用影響。從治療的風險與效益角度來看，嚴重度高的病人應積極治療，但病情輕微的病人就不適合該項治療了。如果忽略疾病嚴重程度這個特質，把全部研究結果合併在一起分析，得到一個整合後的反應率，結果是稍微增加 7%，將可能誤認為該治療只有一些不甚明顯的效果，而錯失治療嚴重病患的機會；也可能不恰當的把治療用在病情輕微的病患，而增加治療副作用造成傷害的機會。

類似這種影響治療效果的特質，稱為療效的調控因子 (effect modifier)，研究應盡可能將結果依重要的調控因子分開分析。可增加積極療效的因子，臨床上會被考慮為治療的適應症 (treatment indication)；相反地，可能增加治療所造成危害的因子，將被考慮為治療的禁忌症 (contraindication)。

統計上可透過檢定某個研究因子與療效是否發生交互作用 (test for interaction)，做為判斷該因子對治療是否有調控效果的參考。當療效因某個因子的不同而明顯改變時，統合分析的異質性檢定也會顯示顯著的結果。因此整合研究結果時，需要注意是否有明顯的異質性存在。

從另一個角度來看，當各種不同特質的研究，都一致地顯示類似治療效果時，我們認為該項治療具有良好的外部效度。這樣情況下，來自不同研究特質的各項研究結果可以進行整合。例如，一個藥物在不同人種的研究，均顯示一致的治療效果

時，該類研究具有良好的外部效度，也認為該治療具有良好的一般性（generalizability）。

間接比較分析裡，各研究的外部效度可以從三個層面來考量：

- (1) 同一群研究裡面(A對C或是B對C)，各研究的異質性是否顯著；
- (2) 兩組比較所包含研究的相似性(similarity)；
- (3) 與直接比較試驗結果相比，分析結果是不是一致(consistency)。

第(1)與(3)項，可以參考一般統合分析所用的異質性檢定方法來檢查。但是對於(2)的問題，檢定兩組比較之間的異質性，也同時解釋為該兩種治療經過間接比較是有差異的，因此不適合用這個方法來檢查兩組比較的相似性。

研究群組之間相似性的分析，是一個研究方法上仍在進展中的課題。這個問題的基本思考是，如果將原本接受A與C比較的病人群，換做接受B與C的比較，結果是否跟現今所看到的B與C比較相同。因此著眼點是怎樣的因素會造成這些研究比較結果的不同。

目前對於間接比較研究群之間的相似性，多以質性敘述的方式進行討論。從各研究所收納的病人群特質、各研究採用的治療方式、各研究結果的測量方法來比較，並加入研究者的專業判斷。舉例來說，臨床治療經常考慮人種的差異，如果兩組治療比較的研究對象在人種分布上有顯著的不同，且缺乏證據顯示不同人種接受相

同治療的效果一樣時，人種這個因素的差異將是一個需要注意的問題。相反的，如果有充分資料顯示，同一組治療比較的多個研究分別包含各種廣泛特質的研究對象，例如不同年齡層、性別、地區、種族、疾病嚴重程度、與各種形式的醫療照護搭配、…等，且這些研究結果之間沒有發現明顯的異質性，這樣的狀況可以視為該治療效果具有良好的外部效度。當兩組比較的外部效度都屬良好的時候，我們對於間接比較的結果也會比較有信心。

### 間接比較與直接比較的實證應用

另外一個值得思考的問題是，如果已經有直接比較的研究，是否該執行間接比較分析？

從研究的內部效度，以及統計上資料分析的效率來看，隨機分派直接比較對照研究，是比較理想的研究方法。但是現實上能夠找到的直接比較研究結果並不常見，或者數量有限，也可能只侷限在某些特定的病人群，這些都將影響到實證資料的可用性。這樣情況下，間接比較分析可補充一些不足。尤其是在研究的外部效度方面，間接比較分析有機會收集到特質較為廣泛的各種研究，以評估治療效果的外部效度。

建議先有系統地搜尋相關题目的直接比較研究結果，再擴充到間接比較分析，會是一個較完整的做法。如果找到直接比較分析結果，將這些直接比較先做系統性回顧與統合分析，再看看與間接比較分析

所得到的治療效果是否一致。如有不一致的地方，可進一步探索可能造成治療效果差異的因素。如果兩種方法所得到的效果估計一致，則可考慮用統計方法把兩種結果進一步整合，以增加分析的信度。

在間接與直接比較研究結果一致性的探討上，Song 等學者於 2003 年收集了國際上發表，44 個同時包含兩種研究的統合分析結果[9]。44 個分析中，3 組間接比較與直接比較分析結果呈現統計顯著差異；有 9 組分析，兩種方法呈現方向一致的統計顯著結果；23 組分析，兩種方法的效果比較均沒有達到統計顯著；並沒有發現兩種分析呈現統計結果顯著矛盾（例如間接比較 A 顯著優於 B，但是直接比較 B 顯著優於 A 治療）的情況。整體而言，兩種分析結果呈現中等程度的一致性。

### 統計檢力

一般來說，可納入統合分析的研究數量並不會很多，可能只有數個到十數個。間接比較又把可納入的研究，依照不同的治療對照分組再比較，導致分析誤差的增加，因此經常會遇到統計檢力不足的困境。也就是說當我們想要比較的治療方法之間真的有差異時，統計檢定可能得到不顯著的分析結果。再加上間接比較分析較低的統計效率，使得檢力成爲此類分析更需要注意的問題。

常見的狀況是，分析的結果顯示不同治療之間效果的差異並不小，例如兩種治療有效比例，A 治療可降低中風 5%，而 B

治療可降低中風 15%，治療效果差異達 3 倍，但是檢定 p 值卻大於 0.05，無法宣稱兩種治療的差異達統計顯著。因此這個比較結果，有可能是因爲資料選樣的隨機性，碰巧得到；也有可能臨床上這兩種治療效果真的有差別，但是因爲研究樣本數不足，現有資訊無法提供統計上令人肯定的分辨。

### 間接比較的應用及臨床與實證價值

Song 等學者整理 2000 到 2007 年之間，國際上所發表的 88 個使用間接比較分析的回顧性研究[10]，指出此類研究近年來有逐漸增加的趨勢。所整理的報告中，有 67 個回顧研究（佔 76%）使用正式の間接比較統計方法。回顧的主題中，72 個爲藥物治療，其他包括探討醫療器材、醫療諮詢、手術、以及診斷的效果。回顧內容除了探討臨床效果外，也用在成本效果分析等醫療科技評估方面。

從分析方法的角度，作者歸納發現：當前大家對於間接比較分析的基本假設認識程度不一致；有些回顧使用的分析方法不恰當；執行相關資料搜尋的完整度不盡相同；在相同題目直接比較研究上的搜尋，以及直接與間接比較結果的進一步整合方面較容易被忽略。方法學上，用來做間接比較的研究群之間的相似性，該如何客觀評估仍有待進一步發展。

國際上使用間接比較方法評估不同治

療的效果，已經有一段不算短的時間。但是相對於數量龐大的系統性回顧報告，間接比較分析報告仍不算普遍。執行間接比較分析，比起直接比較研究結果的系統性整合，有更多需要注意的重點。但是隨著新治療方法數量與日俱增，治療方式也益趨複雜，期待每一項藥品或是治療方法，與治療該相同疾病的其他方法，都能找到高品質直接比較之實證資料，將有實際上的困難。從研究資源分配的角度來看，對每種治療都做相互的直接比較研究，不只曠日費時，也不一定是有效益的作法。

雖然使用間接比較分析，比直接比較研究需要更多的注意，但不可否認的，臨床實務上有關不同治療方法的選擇，許多時候是非正式的採用間接比較的形式。如果能夠有系統地執行間接比較分析，且大家能夠掌握間接比較結果的正確解讀，應該是提升現有實證資料應用效率的積極策略，以期能夠有效地協助增進對病人的照護。

### 參考資料

1. 邵文逸。實證醫學中的統計原理。醫療爭議審議報導，系列 17。
2. Simpson EH. The Interpretation of Interaction in Contingency Tables. *Journal of the Royal Statistical Society, Ser. B* 1951;13:238-241.
3. Heiner C Bucher, Gordon H Guyatt, Lauren E Griffith, and Stephen D Walter. The Results of Direct and Indirect Treatment Comparisons in Meta-Analysis of Randomized Controlled Trials. *J Clin Epidemiol* 1997;50:683-691.
4. AM Glenny, DG Altman, F Song, C Sakarovitch, JJ Deeks, R D'Amico, M Bradburn and AJ Eastwood. Indirect comparisons of competing interventions. *Health Technol Assess* 2005;9(26).
5. Eric Lim, Ziad Ali, Ayyaz Ali, Tom Routledge, Lyn Edmonds, Douglas G Altman, Stephen Large. Indirect comparison meta-analysis of aspirin therapy after coronary surgery. *BMJ* 2003;327:1-5.
6. Wells GA, Cranney A, Boucher M, Peterson J, Shea B, Robinson V, Coyle D, Tugwell P. Bisphosphonates for the primary and secondary prevention of osteoporotic fractures in postmenopausal women: a meta-analysis [Technology report no 69]. Ottawa: Canadian Agency for Drugs and Technologies in Health; 2006.
7. Wells GA, Sultan SA, Chen L, Khan M, Coyle D. Indirect treatment comparison [computer program]. Version 1.0. Ottawa: Canadian Agency for Drugs and Technologies in Health; 2009.
8. Wells GA, Sultan SA, Chen L, Khan M, Coyle D. Indirect Evidence: Indirect Treatment Comparisons in

Meta-Analysis. Ottawa: Canadian Agency for Drugs and Technologies in Health; 2009.

9. Fujian Song, Douglas G Altman, Anne-Marie Glenny, Jonathan J Deeks. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. *BMJ* 2003;326:472.
10. Fujian Song, Yoon K Loke, Tanya Walsh, Anne-Marie Glenny, Alison J Eastwood, Douglas G Altman. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. *BMJ* 2009;338:b1147.