



統合分析與森林圖

Meta-Analysis and Forest Plot

以 Cochrane Library 戡科覽醫學資料庫為例

戡科覽（Cochrane）組織是一個提供統合校勘（Meta-Analysis）、醫藥科學，供成員閱覽應用的資料庫，一般譯考科藍。唯似尚無官方中譯，故據其意義與特徵，譯為「戡科覽」醫學資料庫。

戡科覽（Cochrane）於1993年在英國發起，當前由超過三萬七千名的志願者所組成，分布國家超過170個。目的是以「systematic reviews 系統化文獻研究」的方式，歸納醫學研究資訊，提供醫護專業人員、病人、醫療政策致訂者等人選擇使用。

「systematic reviews 系統化文獻研究」就是製作「文獻研究摘要表」，是統合分析（Meta-Analysis）的前置必要工作，再將表格轉化成資料庫格式以進行統何分析。

戡科覽（Cochrane）特色有二：

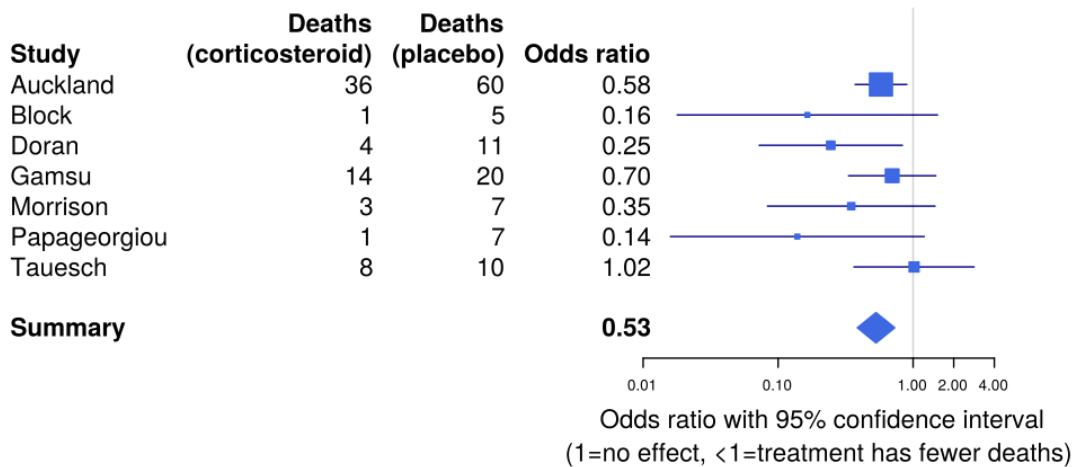
1. 提供統合分析（Meta-Analysis）的各種基本分析。
2. 提供線上資料庫並配合統計森林圖 Forest Plot 與相關視覺化圖形。

（1）成果、服務、呈現方式

以下為戡科覽（Cochrane）官網提供的範例：「對過早分娩女性給予皮質類固醇治療效果」的7項研究統合分析，與其 Forest Plot 森林圖。

這張圖是戡科覽（Cochrane）開始營運、並創建統合分析觀念時的「初版圖表」，優點是簡明，但也可以看出一些不足。

戡科覽（Cochrane）後來發展，也多方彌補改善不足，故比較此「初版圖表」與現在增訂的圖表，亦有觀察演化的意義。



左第一欄：研究者，是文獻摘要表的標準欄位，當前會加入發表年。

左第二欄：實驗組，給予皮質類固醇治療，仍造成嬰兒死亡的人數。

左第三欄：控制組，沒有給予皮質類固醇治療，但給予安慰劑，造成嬰兒死亡的人數。

左第四欄：[Odds Ratio \(OR\) 優勢比](#) 之值。

倒數第二列：**Summary** 合計，即統合結果。

最右欄：表現統合分析之Forest Plot 森林圖。

(2) Forest Plot 森林圖解讀

最下方橫軸：[Odds Ratio \(OR\) 優勢比](#)

OR = 1，表示兩組相同，沒有差異。通常會在此畫 1 縱軸。

如果優勢比值在左，表示左方、實驗組效果好；如果在右方表示右方、控制組效果好，亦即給予實驗處理、皮質類固醇治療，反而導致反效果。

各列之點，表示各文獻之優勢比值，其大小表示相對樣本數大小；線的長度表示 95% 的 **Confidence Interval (CI)** 出入區間（常譯信賴區間，在中文上並不達意，因為其真實意義是「不確定」、類不可信賴。）亦即研究報告的只是樣本優勢比值，而真正的優勢比值，可能在線上的任 1 點。樣本數與抽樣出入呈平方反比關係，所以，點愈大、線愈短；點愈小，線愈長。

如果線跨過中央縱軸，亦即「未達 95% 顯著水準」。

以本圖為例，只有 **Auckland**、**Doran** 2 篇研究全部在左方，亦即其他 5 篇尚不可確認。

在「合計」列的菱形，表現統合分析的結果，亦即按照各研究樣本的比例，將 **Odds Ratio (OR) 優勢比** 與出入區間，共同加權



合計，菱形上下形成直線與橫軸的交點，為統合Odds Ratio (OR) 優勢比，菱形的寬度為統合出入區間。

本例顯示統合 Odds Ratio (OR) 優勢比為 0.53，出入區間未跨越縱軸，「達到95%顯著水準」，亦即認為：過早分娩女性給予皮質類固醇治療，會發生減少嬰兒死亡的效果。

(3) 詮釋與辯難：此圖表結論尚有可討論待決之處

此「圖表初版」反映忽略了「邏輯實證論」與「科學知識產生程序」的最後一步：詮釋與辯難。

文獻研究就是二次資料收集，並不能直接對原始資料作處理與結論。還是必須作資料淨化 Data Screening。尤其統計視覺圖，通常最容易讓研究者發現資料取捨的問題。

資料分布圖，通常有 3 種情形，與其應有之決策。

1.趨勢一致型：可能有結論。

如果資料全部在一側、或大多數在一側，可以宣告可能正確的結論。

2.資料對峙型：可能尚無結論

如果資料分布在兩側、無大多數在一側，則無可能正確的結論，甚至研究的自變項有可能是無關的。

3.出現特異資料：尚待追究

特異資料 outlier 即在一側僅出現 1 個資料，或在資料量很大時，出現在 1 側少於 5% 的資料。本圖表就出現只有 1 個 Tauesch 在右側。

特異資料 outlier 通常是具備非正常因素，或有錯誤而產生，應該刪除。否則他已不當影響統合 Odds Ratio (OR) 優勢比值，降低了此治療法的有效證明。

但歷史上也已多次出現，特異資料 outlier 是對的，而其他全部資料是錯的。而且，若出現這個現象，通常可打破社群內已習以為常的錯誤，或創造新典範。

(4) 特異資料的特殊貢獻；選舉行為研究為例

2012 總統選舉預測

<http://tx.liberal.ntu.edu.tw/~BlackPool/2012Pres/!PresidentialElection2012.htm>

吳統雄預測：

馬英九：蔡英文 = 53.5 : 46.5，馬勝 7%

民進黨委託5家機構均預測蔡英文勝，標準化平均數字為：

馬英九：蔡英文 = 45.16 : 54.84，蔡勝 9.68%



中選會公布票數與相對得票率為：

馬英九 (6,891,139)：蔡英文(6,093,578)=53.07 ：46.93

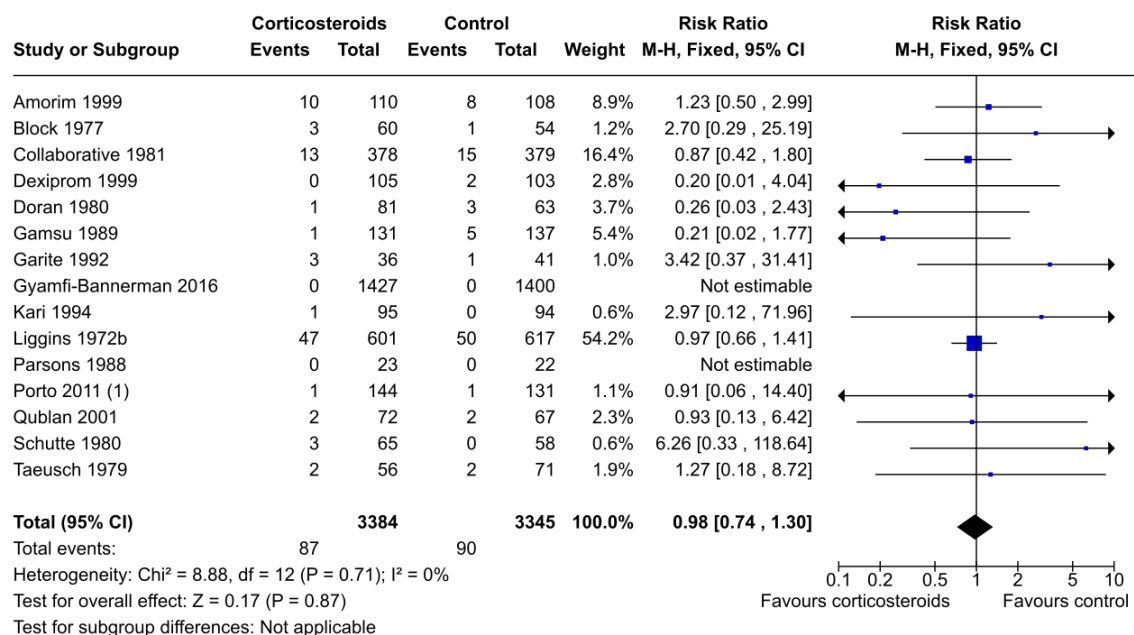
實際馬勝 6.57% 與吳統雄預測相同，誤差僅為：0.43%

在 6 項預測資料中，僅吳統雄與其他 5 項不同，是為特異資料，但特異資料才是正確預測。

(5) 戲科覽 (Cochrane) 的改善發展

戲科覽 (Cochrane) 應也注意到以上的問題，經過數十年的改善發展，雖仍以 Forest Plot 森林圖統合分析之主要表現，但其報告結構與取向已有很大改變。

與以上相同主題「對過早分娩女性給予皮質類固醇治療效果」的新版報告改為：



Footnotes

(1) The outcome measured in this trial was stillbirth.

報告的內容豐富了許多，全文請見：

<https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD004454.pub3/full#CD004454-sec-0053>

改善發展的重點如下：

1. 圖表增益

增加了欄位與更多資訊，包括：

文獻來源，多一個 Subgroup，即在某大型研究中，包括此主題的子研究文獻。



將實驗組（Corticosteroids）與控制對照組（Control）的各組總樣本（Total）與死亡樣本（Events），分別列出。

增加：某研究之樣本數，占全部研究樣本數合計的百分比（Weight）。

改用風險比 Risk Ratio (RR) 代替優勢比 Odds Ratio (OR)。兩者為採用分母不同，當前RR為全體樣本，而初版OR為各組樣本。但兩者解讀方法相同。

M-H：採用 Mantel-Haenszel method 方法。

Fixed：採用固定效果模式（fixed effect model），或譯為「固定因子」。

95% CI：95% Confidence Interval 的出入區間（常譯信賴區間，在中文上並不達意，因為其真實意義是「不確定」、類不可信賴。）。

最右仍是森林圖。就本圖的菱形位置，可知統合分析的結論是：不確定有或沒有效果。

2. 辨識研究差異

戡科覽（Cochrane）體會到研究之間的差異會影響結論的不同，以及統合之後的品質。並將差異分為3類：

Clinical diversity 診治差異。由於戡科覽（Cochrane）是醫學資料庫，故強調診治，即一般實驗法中，實驗處理 experimental treatment 的方法。

Methodological diversity 研究方法差異。如：單次研究或長期研究？選擇樣本的方法...等。

Statistical heterogeneity 統計異質性，特簡稱為 Heterogeneity 異質性。即為一般統計中的「同質性檢定」，而戡科覽（Cochrane）反過來檢定異質性。

3. 新增檢定資訊

(1) I^2 ：Heterogeneity 異質性檢定指標，以百分比形式呈現。

0% to 40%: might not be important：異質性低。

30% to 60%: may represent moderate heterogeneity：存在中等異質性。

50% to 90%: may represent substantial heterogeneity：存在中等異質性。

75% to 100%: considerable heterogeneity：存在嚴重異質性。

本例為 0%，即無異質性，統合結論應相同。

異質性檢定指標的產生計算方法，請參：

<https://handbook-5->

[1.cochrane.org/chapter_9/9_5_2_identifying_and_measuring_heterogeneity.htm](https://handbook-5-1.cochrane.org/chapter_9/9_5_2_identifying_and_measuring_heterogeneity.htm)

<https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-015-0024-z>

(2) Test for overall effect 整體檢定



即 Z 顯著性檢定，樣本之間是否有差異。本例為： $Z = 0.17$ ($P = 0.87$)。

傳統上有設定 2 個水準：若 $P < .05$ 或 $P < .01$ ，（即 $Z > 2$ 或 $Z > 2.5$ ）則樣本之間存在差異。

P 實為機率值，並沒有一個必然斷點，傳統的顯著水準，是提供一個人為判斷點。現代電腦可提供明確數據。本例即：各樣本（研究文獻）之間結論相同的可能性為 87%，彼此差異很小，在 0.17 個標準差之內。

Test for subgroup differences 子群組差異分析

本例沒有子群組。

按照研究方法之不同，分別統合

戡科覽（Cochrane）應更感受到研究方法影響結果，所以會將差別很大的研究方法，再分組報告。

其他分組後報表，請詳見原文。

4. 增加「文獻篩選」先行程序

戡科覽（Cochrane）當前對是否要納入統合分析的文獻，先實施了相當嚴密的「文獻篩選」程序，如被判為品質不足的文獻，將不被列入統合分析。本主題的「初版圖表」中，數據影響最大的 Auckland 研究文獻，在當前版被剔除了，也應是「初版圖表」認為「給予皮質類固醇治療效果」偏向有效，而當前版認為效果不明的原因之一。

「文獻篩選」完整程序請見以下全文「研究方法」章中的 Assessment of risk of bias in included studies 節，與「研究發現」章中的相應報告：

<https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD004454.pub3/full#CD004454-sec-0053>