

臨床試驗

報告者：莊筱婷

參考來源: Clinical Translational Science: Principles of Human Research, David Robertson, Gordon H. Williams, Chapter 2 Clinical Trials

摘要

本篇文章提供臨床試驗的簡介，包含研究的設計或會面臨到的問題等。

大綱

- 什麼是臨床試驗？
- 臨床試驗結果的表示
- 臨床試驗設計可能面對的潛在問題
- 臨床試驗設計
- 研究假設及統計分析
- 系統性文獻回顧
- 倫理議題
- 研究組織

前言

臨床上的實踐講求整合科學上參考依據的實證醫學，最基本的工具為**隨機分派的臨床試驗**(*randomized clinical trial, RCT*)。隨機分派(randomization)的概念最早是在 1926 年由 Fisher 應用在農業實驗研究(Fisher and Mackenzie, 1923)，他提出在統計分析變異時，各個研究觀察值必須獨立並且沒有干擾，才不會影響研究分析的結果。而在 1931 年，Amberson 將隨機分派的概念應用在人類的試驗研究，評估結核病的治療(Lilienfield, 1982)。在 1948 年英國醫學研究會(British Medical Research Council)針對肺結核的抗生素治療之研究中，建立了大樣本隨機分派的準則，並提供實驗治療及結果評估的指導方針(Medical Research Council, 1948)。

隨機分派的臨床試驗主要是依照嚴謹的研究計畫將研究受試者隨機分配至一種或另一種介入方案，以準確地評估介入後的影響。這個方法漸漸地被用來評估臨床試驗的品質，經隨機分配可以將受試者的潛在危險因子隨機分布，使研究結果有可比較性；另外，研究結果可以應用於藥物的製造、醫療的處置或是健康政策上。雖然臨床試驗仍有部分的爭議，但是像這樣科學化的工具依然目前已被廣泛地使用。

什麼是臨床試驗？

臨床試驗的分期

一個臨床試驗結果的評估最重要的就是需要有明確目標。臨床試驗分為四期(如表一)：前兩期主要是新治療的初期評估，了解新治療的安全性、生理上變化的量、結果是否符合治療機制的假說。第三期則是最重要的一期，評估新治療是否可以應用在臨床。此期評估的內容可能包含存活、主要診斷上的變化、生活品質或成本效益、新治療實踐的可能性等。第四期則是新治療實施後的監測，除了監測療效或不良反應，同時也要發展使用的指示，評估不同程度的治療或是與其他治療方式比較。詳細內容可以參考衛生署公告的藥品臨床試驗一般基準 (<http://www.fda.gov.tw/files/list/藥品臨床試驗一般基準.pdf>)

表一 臨床試驗分期

| 分期 | 特性 | 目的 |
|-----|-------------------|------------------------------------|
| 第一期 | 以少數的志願者評估新治療 | 了解新治療用於人體是否合適。 |
| 第二期 | 以病人做新治療的初期試驗 | 了解劑量反應效應，評估不良反應及其他對疾病上的生理影響。 |
| 第三期 | 大規模的新治療試驗與一般治療的比較 | 評估新治療是否可以取代舊有的一般治療。 |
| 第四期 | 監測新治療在臨床上的實踐 | 上市後的監測可以獲得多的藥物影響、使用率、不良反應的發生等相關資訊。 |

臨床試驗種類

臨床試驗分為 *解釋性(explanatory)* 及 *實用性(pragmatic)* 兩種。解釋性臨床試驗是利用則足夠大的樣本來驗證研究假設的正確性，主要是衡量一種治療的理想成效。實用性臨床試驗是使用不同性質的大樣本族群來推估一種治療的實際臨床成效。一種治療的實施即是依據解釋性及實用性的研究結果判斷，實用性試驗目標或架構可以利用解釋性研究方法與實用性研究方法的比較來做更進一步的了解 (Tunis et al., 2003)。

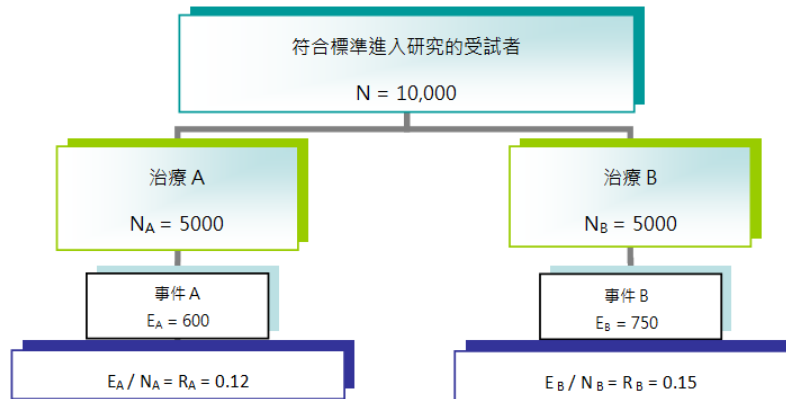
臨床試驗結果的表示

研究中治療可能帶來的效益或傷害可以以相對或絕對危險影響來表示，除了點估計還要有其 信賴區間(confidence interval) 的估計，用以表示治療效果的估計有多精確。許多的研究比較常用相對危險(relative risk, RR)，而非使用絕對危險差異(absolute risk difference, RD)(Bobbio et al., 1994; Naylor et al., 1992)，主要是因為以相對危險降低度表示結果效果會比較明顯。*治療的相對效益(relative benefit of therapy)* 是表示一般治療影響的最好計算值，包含危險對比值(odds ratio) 或相對危險降低度(relative risk reduction, RRR)。此計算值包含信賴區間可以表示統計上的顯著差異，且有方向性。

然而在解釋性臨床試驗比較講求此治療是否有帶來直接的影響，研究可以使用 *益一需治數(number needed to treat, NNT)* 或 *害一需治數(number needed to harm, NNH)* 來表示，數值皆是以絕對差異值來計算。NNT 是研究必須治療多少個病人，才能預防一個額外的不良結果；而 NNH 是 NNT 的反向思考，考慮治

療的可容忍接受度，即研究治療多少個病人會造成一個額外的不良結果。

以上的相關指標以兩種治療為例，如下：



統計描述

| | 發生事件 | 無事件 | |
|---|----------------------|-------|--------|
| A | E _A = 600 | 4,400 | 5,000 |
| B | E _B = 750 | 4,250 | 5,000 |
| | 1,350 | 8,650 | 10,000 |

相對危險(relative risk, RR) = $R_A / R_B = 0.80$

相對危險降低度(relative risk reduction, RRR) = $1 - RR = 0.20$

危險對比值(odds ratio) = $\frac{R_A / (1 - R_A)}{R_B / (1 - R_B)} = 0.77$

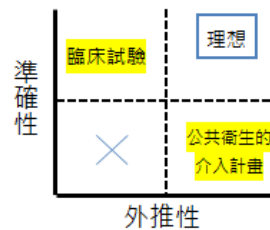
絕對危險差異(absolute risk difference, RD) = $R_B - R_A = 0.03$

益一需治數(number needed to treat, NNT) = $1 / RD = 33$

臨床試驗設計可能面對的潛在問題

1. 這些結果可信嗎？結果可以推估於母群嗎？

研究試驗的內部有一致性嗎？如果研究能重複同一試驗，能得到相同的結果嗎？一個研究試驗的結果，其準確度是我們關心的，即研究的信度、效度是很很重要的。另外，研究的結果是否可以推估在廣大族群使用後的成效呢？在這裡，研究的準確度和外推性的關係如圖。在臨床試驗中，為了提高研究的準確性，通常會先篩選特定條件的族群進入研究，但是這樣會減低研究結果的外推性。而一般公共衛生的介入計畫雖然放寬了進入研究的條件，有較高的外推性，但是受試者的性質變複雜了，可能會影響研究結果的準確性。當然，外推性及準確性差是我們最不想見到的，最理想的狀態就是能同時提高外推性及準確性。



圖一 實驗設計之準確性及外推性

2. 在解釋性臨床試驗中會高估治療的期望影響？

在研究設計時很難去假設其他對治療結果也很重要的生物或行為等相關因子的影響。在小規模的試驗中觀察到的影響結果經常會在之後的大規模的研究中被發現是高估的。其原因有可能是：

(1) 刊登偏差(publication bias)

通常研究的結果達統計上顯著性時比較容易被期刊接受並刊登，被刊登的研究中約有 85%的結果評估是有效的(Dickersin and Min, 1993)。但是 simes(1987)回顧過去已刊登的研究知道複合式化療對治療卵巢癌有益處，而將已刊登和未刊登的試驗一起評估時卻發現其實治療是沒有顯著性的影響。

在小樣本的試驗中可能有刊登偏差，所以在小樣本的試驗結果可以發現過去的研究結果可能被高估了。而現在的臨床試驗已被要求要先上網登記，如此可以將所有的研究結果匯集並公開。

(2) 一個類同 regression to the mean 的概念

在小樣本的研究中，由於樣本較少比較容易受到極端值的影響，導致結果可能被高估或低估。而在大樣本的研究中，極端值的影響會因樣本數變大而減少，使研究結果趨向於平均值。

研究的樣本數僅研究的樣本數及進入研究的條件限制都要在研究設計時嚴謹地被定義，增加研究結果估計的可信度。

3. 質性交互作用的影響(differences in kind)

研究的結果可能會受到受試者的個人因素而影響，因此，整體試驗結果通常會分組來呈現更明顯的結果。*次組分析(subgroup analysis)*是常見的臨床試驗之作法，舉例來說，使用基因型(genotype)或基因表現分析(gene expression analysis)作為評估工具，一樣都可以發現化療和放射線的治療是有差異的(Dressman et al., 2006; Potti et al., 2006; Schwarz et al. 2008)。但是分組的定義及結果是有爭議性的，因為在各組中有偽陽性或偽陰性的可能。

另外，治療所帶來的影響是否只對研究中的小樣本族群有意義，研究除了需要有足夠的檢力來做分析，樣本的代表性、品質也是很重要的。

4. 量性交互作用的影響(differences in degree)

一般研究評估量性的交互作用比質性的容易，所謂量性的交互作用是治療的相對效果可能會因為疾病的嚴重度不同而有不同的結果，舉例來說，相較於年

輕人，老年人進行搭橋手術(bypass surgery)的治療效果較佳。

不過研究通常為了減少分析的成本或是複雜度，所以會針對研究樣本的條件做限制，一般會排除低危險族群和高年齡層或腎功能有缺損者，如此也可以解決交互作用的影響。

5. 研究所未預期的結果發生

(1)未預期的生理機轉

一個治療的發展要有明確的目標及生理機轉，但在研究中通常不會納入全部的人做試驗，研究分析也只是驗證假說，要特別注意的是，治療也可能有其他不同的生理機轉正在同時進行著。舉例來說，一個減肥藥物 phentermin-fenfluramin combination 除了會抑制食慾，同時還會導致心臟異常肥大。

(2)多重治療的交互作用

有許多疾病會使用一種以上的治療方法，然而一般臨床試驗很少同時評估多重治療的影響，這樣可能會因為治療間的交互作用而有誤差。舉例來說，血小板受體阻斷劑 abciximab 和肝素 heparin 在體內會產生交互作用，增加出血量。

(3)治療後長時間的影響不可預期

治療後短時間和長時間的影響在一般手術治療評估是容易被理解的，病人在術後會再評估其死亡率或發病率而得到長時間的影響結果。這樣的觀念在急性醫療處置也是一樣，舉例來說，在使用血栓溶解治療(fibrinolytic therapy)的 24 小時內會增加死亡的風險，但是對長期死亡率的影響是有益的。

以上試驗的追蹤時間、試驗可能帶來的影響等都有可能會被遺漏，所以在實驗設計中有適當的追蹤時間及要觀察影響的定義是很重要的。

臨床試驗設計

依照臨床試驗申請計畫書摘要第五至九點(http://www.fda.gov.tw/files/list/藥品臨床試驗申請須知_9609.pdf)，在研究開始前，先要有一個種的研究設計，如此才能給予研究最終的結果，研究設計時應考慮：

1. 進入研究的條件限制(criteria)

在解釋性臨床試驗中，能進入研究的樣本應要有完整的限制條件，以減少干擾。而實用性臨床試驗則是依照解釋性臨床試驗中的限制，將條件放寬，納入



更多受試者，使研究盡可能可以代表母群。

2. 收集資料的方式

提供研究的資料要如何收集要事先建立。在解釋性臨床試驗中，由於樣本數不多，收集資料上遇到的困難比較容易解決。然而在實用性臨床試驗則要考慮一個龐大的樣本數，研究收集資料的速度盡可能要快，所以收集資料的方法或是工具也應盡可能的簡單、方便。

3. 療效指標(endpoint)

臨床試驗的結果評估要最初應有完整的定義，即**療效指標(endpoint)**，研究分析者也可能在盲化的設計下進行臨床試驗結果的相關校正，提高研究精確度。有時候為了減少研究的時間及成本，所以研究會使用與療效指標有相關的**替代指標(surrogate endpoint)**來做評估。但是使用這種指標有可能會面臨到幾個問題：疾病的形成有很多種原因，選擇的治療介入可能只會影響在眾多病因中的替代指標而已；另外，所選擇的替代指標與疾病的形成沒有因果相關。舉例來說，抗心律不整藥物可以降低心室跳動速度，但是 Cardiac Arrhythmia Suppression Trial(CAST)的研究卻比預期的早就結束，原因是使用這樣的治療會有比預期高的死亡率。因此研究若要使用替代指標，一定要有足夠的文獻證實與主要的療效指標有相關性。

4. 其他輔助治療的措施

其他研究中未考慮的治療亦有可能會對研究的結果造成影響。研究時應要考慮可能有相關的輔助治療情形，避免影響結果的有效性或外推性。

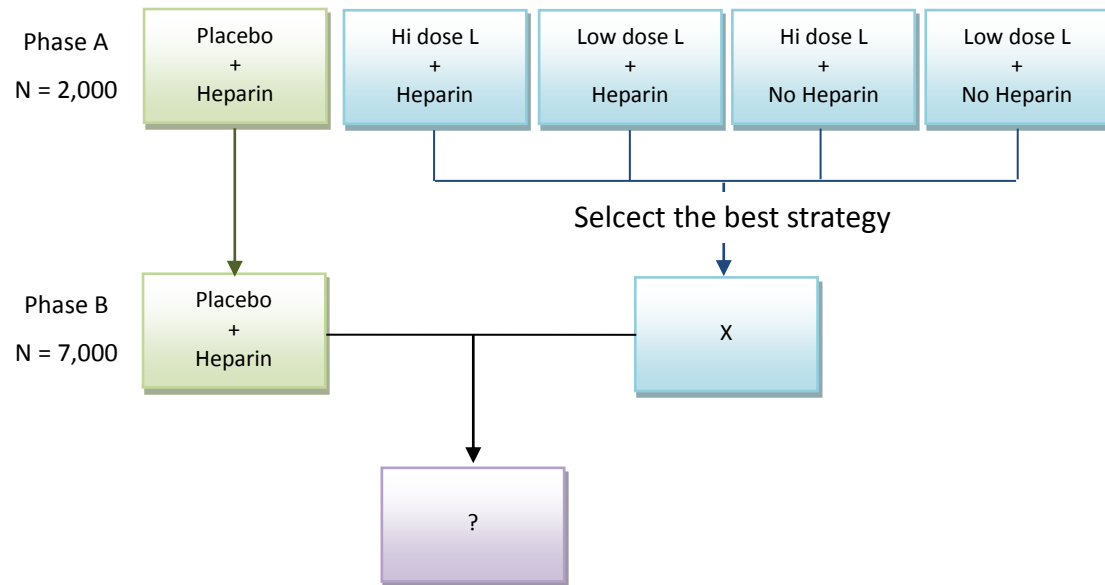
5. 多重隨機分派

當研究有多種治療法需要同時評估時，研究會因要處理的資料變多而變得比較複雜。一個**多因子矩陣型試驗(factorial trial)**的研究設計即是適用於兩種以上治療方法的評估，可觀察兩種治療間交互作用的影響。

6. 適當的研究

研究進行有一個很重要的觀點：**不確定原理(uncertainty principle)**，即研究中所介入的治療效果好壞是不確定的，如此才不會有倫理道德的問題存在。一般發展一個治療都希望是能療效，所以在做研究時一定要盡可能的選擇一個最好的治療方法(pick-the-winner strategy)進行研究，如圖二。另外，研究的分組一定要

有適當的隨機分派，如此可以將實驗中其他可能會影響研究結果的因子均勻的分佈在各組，減少個人偏差。



圖二 pick-the-winner strategy

7. 盲化(blinding or masking)

在解釋性臨床試驗中，為了降低偏差，研究常會有盲化的設計，即研究中的部分人員不知道治療的方式是哪一種，依照這些部分人員分為：

- 單盲(Single blinding)：受試者不知。
- 雙盲(Double blinding)：受試者及調察員不知。
- 三盲(Triple blinding)：雙盲再加上資料監測者等不知。

一種研究設計：*double-dummy*，即是為了保持盲化的設計，而在藥物試驗的對照組中，使用與治療組相同外觀的藥物，但實質對研究結果沒有影響的安慰劑。

不過要注意的是，不是每一個試驗都可以有盲化的設計，如手術，不能因為不讓受試者不知道是否有治療而製造一個手術切口，這樣只會增加受試者危險的風險，在道德倫理上也是不被允許的，也增加試驗的成本。

8. 介入治療的強度

研究給予的治療其強度要適當，這樣的概念在藥物劑量的決定上特別重要，舉例來說，最近有研究指出使用紅血球生成素治療(erythropoietin-stimulating agents)會導致血紅蛋白濃度較高的病人出現較多的心血管不良事件。所以在解釋

性臨床試驗時一定要先有適當的治療強度，避免在實用性臨床試驗有更多不良事件發生的可能。

研究假設及統計分析

臨床試驗申請計畫書摘要第十點(http://www.fda.gov.tw/files/list/藥品臨床試驗申請須知_9609.pdf)即是研究統計的部分，包含：

研究假設

研究設計最重要的依據就是為了回答問題，所以設立一個假說，在研究最後才能明確地回答問題。臨床試驗的研究可以觀察目前臨床上治療還未釐清的問題，另外也可以比較治療的效能。通常研究會使用 *不劣性試驗(non-inferiority trials)*，主要是要證實新治療的療效至少不輸舊治療，要注意的是，此種假設為了在統計上達到推翻假說的顯著意義，研究需要有足夠大的樣本。

研究會先設兩種治療的效果是相同的虛無假說(null hypothesis)，當虛無假說為真時，但研究統計結果卻拒絕虛無假說，則犯了 *型一誤差(α error)*。一般 α 訂為 0.05，即研究容許有 5% 的機率會犯錯。注意若研究不斷重複分析會增加 α error，因此研究檢定次數也是很重要的。

當研究虛無假說不正確時，但研究統計卻接受虛無假說，則犯了 *型二誤差(β error)*。研究的統計 *檢力(1- β)*，即正確地拒絕虛無假說的機率，研究各組間的結果差異以及研究的人數會影響檢力。檢力一般能在 80% 就非常好了。

研究利用文獻探討估計研究結果可能帶來的 *最小差異值(minimally important clinical difference, MID)*，使用此值計算研究達到所訂的檢力下所需的最小樣本數，並考慮研究中間可能遺失的比率，最後得應納入研究的人數。

研究結果評估方式

依照受試者未完成研究的情形，有兩種分析方法：一為 *治療意向分析法(intent-to-treat)*，研究將所有符合條件且經隨機分派的受試者皆納入統計分析。若研究只納入完成追蹤治療者，此為 *符合計畫書分析法(per protocol)*。

系統性文獻回顧

在臨床應用上，最常會問：有足夠的證據可以證明此種治療是最好的嗎？此時將不同的研究資料整合進行分析，可以增加統計上的檢力，此稱為**統合分析 (Meta-analysis)**。要注意的是，這些研究的內容相似度一地要高，才能確保分析結果的正確性。同樣的，此種分析會不會有刊登誤差，另外，分析中的各個研究的品質優劣不一，這都可能影響分析結果，不過這些問題都可以用統計方法來解決。

統合分析一開始會先假設各個研究為同質的，為一 *fixed effect model*，然而，一般研究要達到同質是不可能的，所以有 *random effect model* 的發展，同時考慮組間及組內的變異。另外也考慮有其他更新的研究結果會加入分析，而有 *cumulative meta-analysis*。分析也可以使用 *meta-regression* 評估其他共變項的影響，以解釋非同質性的結果。

倫理議題

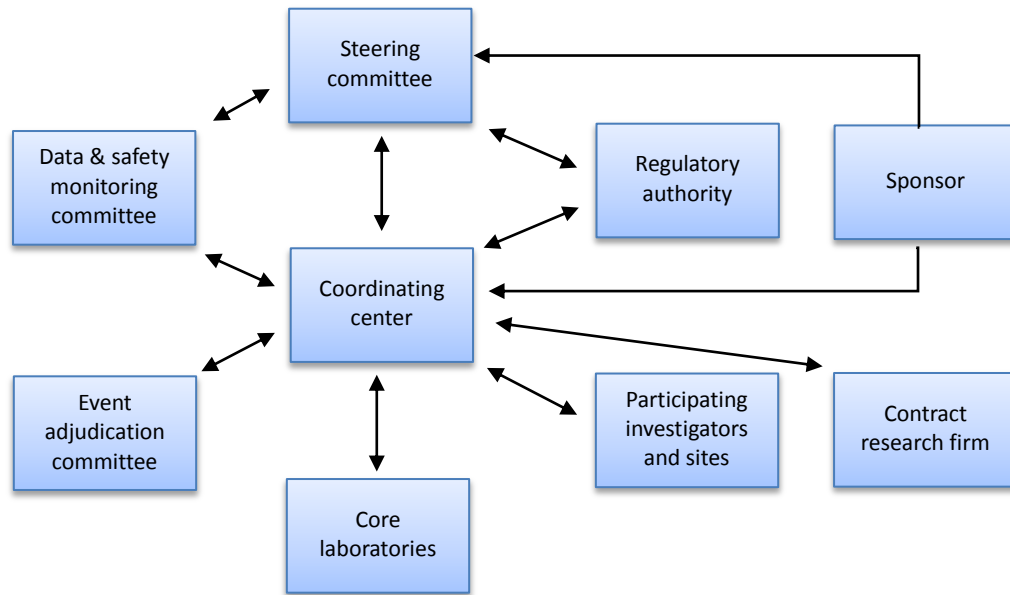
醫療正義、行善、知情同意、不傷害

對於舊有的治療有不適反應者，他們是否要更換另一種新的治療方法，大部分的病人都會聽從醫師的指示，而這新的治療方法通常會被認為是比舊有治療方法好的。然而，此時為了證明新的治療方法的效果，研究設計不免於需要與接受一般治療或安慰劑的比較。

到底需要多少人參與研究，相較於實用性臨床試驗，這在解釋性臨床試驗中比較會碰到的問題，研究若希望結果能達統計上的顯著差異需要有足夠的樣本數，評估研究所需最少的樣本數很重要的。要注意的是，要讓每一位受試者在了解研究內容並且同意的情況下，研究始得開始。

病人的傷害較容易被監測出來，而存活是無法被監測的。舉例來說，使用 β 阻斷劑初期對心臟衰竭的患者會有惡化的情形，但研究也指出此治療對長期的存活影響是有益的。任何治療除了有治療效果，也多少會有些不良反應。總之，無論是哪一種治療方法，以『不傷害』為重要的原則。

研究組織



圖三 臨床試驗的基本組織架構圖