

# 淺談統計與臨床試驗

林 慧

## 一、前言

一些頭痛或牙痛的患者經常會到藥房裡買一些止痛藥來服用。過了約半個鐘頭之後，疼痛的症狀減輕了許多。他們於是又可以恢復平常的生活步調。藥房裡琳瑯滿目的藥品可以減輕病痛，又可以醫治各種疾病，如高血壓、心臟病等等，這些“救命神丹”到底是什麼？他們又是如何產生的呢？本文將對這些問題作一簡略的答覆。大略描述藥品的研究開發過程，以及統計在這過程中所扮演的角色。

## 二、藥品研究開發過程

一個新藥從研究開發到上市需要冗長的時間。根據「今日美國」(*USA Today*)的報導，這個過程平均費時十二年又八個月。實驗室中藥品成分的開發佔據了前面的三年半。再來是為期六年八個月的人體試驗。而後藥廠即將這些試驗結果送至美國食品藥物管理局 (Food and Drug Administration, 簡稱 FDA) 審核。美國食品藥物管理局一般的審核時間為二年六個月。如果審核通過，廠商即能將此新藥推出市面。於是這家廠商的股票也跟著大漲。

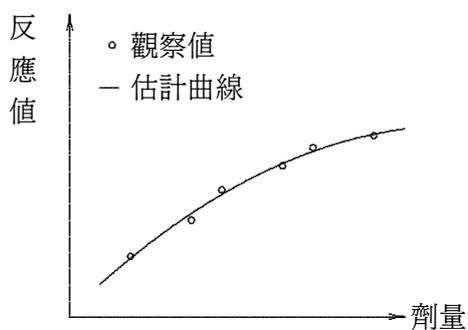
爲了確保藥品能符合美國食品藥物管理局的審核及藥品的安全性 (safety) 和有效性 (efficacy)，統計方法在研究發展的各個階段爲極重要的工具。在下面各小節中，我們敘述了研究開發的各個階段及統計所扮演的角色。其中包括了臨床前的試驗和臨床試驗。在第三節中我們將討論設計臨床試驗及分析實驗結果時，統計上的一些考量因素。

### 2.1 臨床前的試驗 (Preclinical Trials)

開發新藥的第一個步驟即是開發組成此新藥的化合物。此過程稱爲藥品開發 (drug discovery)。在這個階段，化學家及其他研究人員針對特定疾病找出治療此疾病的一個新的化合物。然後變化分子結構 (molecular structure) 以期找到最佳效果的組合。在這階段中，我們需以動物 (如老鼠) 作試驗，以測試其反應值，觀察對藥理 (Pharmacological) 方面的活動，及找出不同劑量與反應值的關係。

爲了確保試驗結果具有科學上的效力，此時我們需用到多種統計方法來設計實驗及分析結果。例如在開發化合物的過程涉及到

多種因素。為了能有效地找出此化合物及達到最佳的分子結構組合，一般常用的統計方法有篩選設計 (screening design)，多變量分析，如群聚分析 (cluster analysis)，區劃分析 (discriminant analysis)，因素分析 (factor analysis) 等。又如，在找出不同劑量與反應值之間的關係時，我們使用迴歸 (regression) 方法以實驗室的數值為依據，找出最能描繪此關係的曲線 (見圖一)。



圖一

如果上述的試驗結果良好，藥廠會決定繼續研究開發此新藥。接下來需做此藥物毒性 (toxicology) 的測試。我們知道藥品雖然可以治療疾病，但其本身亦含有毒性。尤其劑量高時雖然療效更加，但相對地，毒性亦跟著增加，此時我們需找出毒性與劑量間的關係 (如用迴歸方法)。若從這些實驗中此藥品顯示出為安全的，且有潛在的能力治療此疾病，則此時藥廠會向政府機關提出調查新藥的申請 (Investigational New Drug Application, 簡稱 IND)，將藥品的試驗推展至人體。即為所謂的臨床試驗 (clinical trials)。在2.2節中我們會更進一步地介紹此階段。

在人體臨床試驗以前，除了動物體內的毒性測驗，我們還需觀察動物對藥品吸收 (absorption)，分配 (distribution)，新陳代謝 (metabolism) 及排泄 (elimination) 等情形，即一般熟悉的 ADME 研究。藥品劑型 (formulation) 的問題亦需在這階段考慮。我們希望選擇的劑型，可最有效地將此藥品送至體內的活動區 (site of action)，而產生最佳的治療效果。一般常用的劑型有錠劑 (tablet)，膠囊 (capsule)，藥粉 (powder)，及藥水 (liquid)。如果藥品為外用藥，一般則考慮乳霜 (cream)，乳液 (lotion)，及乳膠 (gel) 等等。一般國家的政府機關對藥品的劑型規定需有嚴格的試驗，才能將此劑型用於人體上作實驗。

## 2.2 臨床試驗 (Clinical Trials)

這個階段的試驗一般可分為四期，即第一期 (phase I) 至第四期 (phase IV)。

### 第一期 (Phase I):

第一階段人體試驗的主要目的是監控藥物的安全性，對其特性獲得初步的瞭解，我們藉由此階段的試驗，觀察了藥物動力學 (Pharmacokinetics) 及藥物靜力學 (Pharmacodynamics) 上的反應。藥物靜力學主要在於探討藥物對人體所起的生化及物理作用。觀察其反應的時間，過程 (time course) 及獲得一些藥物在人體應用的劑量及服藥期間 (dosing interval) 的經驗。藥物動力學則是觀察人體對藥品作用的過程，其中包括如何吸收，分配，及新陳代謝等。我們可藉由一些參數來觀察此結果，例如：達到

最高血液濃度所需的時間 (time to maximum concentration, 簡稱 T-max), 最高血液濃度值 (maximum concentration, 簡稱 Cmax), 反應曲線下的面積 (area under the curve, 簡稱 AUC) 等等。除此之外, 在第一期的臨床試驗中, 我們亦作生體相等 (bioequivalence) 及生體可用 (bioavailability) 的試驗, 以測試同一種藥品的不同劑型或不同藥品是否可達到相同的療效。

如上所示, 這個階段的試驗, 我們欲確保藥品的安全性及討論人體對此藥物的反應, 爲了保守起見, 此時的試驗個體不爲病患者, 而是一些符合條件的男性。例如, 健康情形良好, 無心臟病, 無高血壓, 身高體重在一一定的範圍之內等, 年齡層約在二十至四十歲之間, 每次的試驗個體數大約在二十至八十之間。

#### **第二期 (Phase II):**

如果試驗的藥品在第一期沒有出現安全及其他方面的問題, 此藥品即可用於有特定疾病的患者。臨床試驗的第二期即是以此藥欲治療的病患爲試驗個體。在此試驗中, 除了藥品的安全性, 我們亦觀察其有效性及更進一步地收集藥理, 藥力及藥物動力學方面的資料。同時, 我們亦作藥量範圍 (dose ranging) 的試驗, 找出此藥品的有效藥量範圍以用於第三期的試驗中。

#### **第三期 (Phase III):**

這個階段包含了美國食品藥物管理局規定的“至少兩個充分及控制良好 (adequate and well controll) 的試驗”。在此階段中, 我們更進一步地評估此藥的有效性與安全性。除了短期 (如數週) 的試驗外, 亦觀察長

期 (約一至二年) 的行爲。此階段包括了評估藥品有效性爲主的重要安慰劑試驗 (pivotal placebo-controlled study) 及和別種已在這方面所使用的藥物之比較試驗 (active-controlled studies)。試驗個體亦包含了一些特定的母體如年長者 (通常爲 65 歲以上), 或肝臟, 腎臟有疾病者。

如果第一期至第三期的臨床試驗結果良好, 藥廠會向美國食品藥物管理局提出新藥申請 (New Drug Application, 簡稱 NDA)。若獲核准, 廠商則得以將此研究成功的新藥推出市面。美國食品藥物管理局在審核新藥申請時, 會嚴格地檢閱所有有關文件。包括臨床試驗前及臨床試驗的一切過程及結果。此審核小組的主要成員包括醫學、藥學、化學, 及統計方面的專業人士。統計審核人員的職責在於審查整個研究開發過程所做的實驗及分析是否具有統計的效力 (statistically valid), 以確保整個試驗結果的品質。

#### **第四期 (Phase IV):**

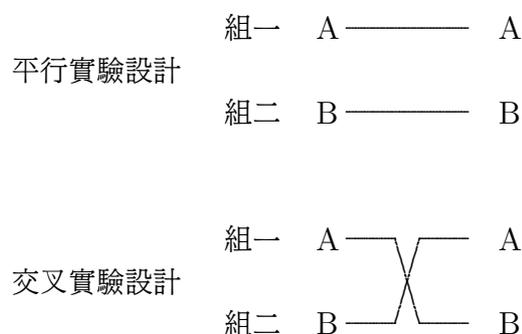
在藥廠將新藥推出市場後, 藥廠方面仍需繼續地觀察新藥的安全性, 注意是否有副作用事件 (adverse event) 及一些特定藥理方面的效應產生。對一些不曾包含在前試驗階段的病患如老年, 幼童, 孕婦等, 因其亦可能爲服用此藥品的病患, 因此會在此階段中加入試驗。除此之外, 這個時期亦對罹病率 (morbidity), 死亡率 (mortality) 及藥品可能產生的新用途 (indication) 作深入的研究。

### **三、臨床試驗中統計實驗設計與分析之一些考量因素**

在整個藥品研究開發過程中，從化學成份的發現至人體的試驗，統計均扮演了極重要的角色以確保此藥品具有成份 (identity)，效力 (strength)，品質 (quality)，及純度 (purity)。尤其美國食品藥物管理局在審核新藥申請時對試驗的設計與分析有嚴格的要求。因此統計在整個藥品研究開發過程中是不可或缺的。在這一節中，我們列舉了一些臨床試驗中，統計實驗設計與分析上的考量因素。

### 3.1 平行 (Pararell) 與交叉 (Cross -Over) 實驗設計

臨床試驗的前期大都使用交叉的實驗設計，而中後期則以平行實驗設計為主。在平行的實驗設計中，每個試驗個體在整個試驗期間只接受一種藥 (亦稱處理)。而交叉實驗設計中，每個試驗個體可接受一種以上的處理。例如有四十名病患參與為期四週之試驗。若將此四十名病患隨機分為二組，組一和組二，每組各由二十名組成。如圖二所示，在平行實驗設計中，組一的試驗個體在為期四週的試驗均接受處理 A，而組二亦只接受處理 B。反之，若採用交叉實驗設計，組一的成員在前面二週先接受處理 A。經過一段洗滌時間 (washout period) 讓可能的殘存藥效排出體外後，組一的成員於是被交叉至處理 B。相同地，組二先接受處理 B，而後交叉至 A。



圖二. 平行及交叉實驗設計

交叉與平行實驗設計各有其優缺點。例如在交叉實驗中，每個試驗個體均為自己的控制 (control)。因藥的差異可由人體內的差異來評估，所以交叉實驗大致上來說比平行試驗所需的試驗個數為少。因此交叉實驗設計較為經濟。而且在交叉實驗中，試驗個體可免去長期使用藥效較低之處理。因其在某些階段，即可將試驗個體交叉至不同的藥品，所以從這個觀點上來看，交叉實驗是比平行試驗為道德 (ethical)。除此之外，由於試驗個體有機會接受一種以上的處理，他們於是有可能表達對某些處理的喜好程度。然而在交叉實驗中如兩種處理間的洗滌時間不夠長，前一期的藥效即可能延續 (carry-over) 至後一期。此即所謂的延續效應 (carry-over effect)。此時，我們則無法辨認所觀察的差異是由於處理效應 (treatment effect) 或是延續效應。此現象即為一般通稱的混淆 (confounding)。若有這種現象發生，不同處理被使用的先後順序需加以考慮。若處理在不同期間 (period) 有不同的效應。結果則容易

產生偏差 (bias)。再則，由於試驗個體在不同期間交叉至不同處理，所以當試驗要比較多種處理時，因需較長的時間，容易產生試驗個體退出 (drop-out) 試驗的情形。對這些退出的個體，我們則無他們退出試驗後的資料。此情形會造成統計分析上的困難。其他缺點如所欲治療的疾病在短期內即能痊癒 (如感冒)。病患可能在痊癒時僅接受了部分的處理。在此情況下，我們即失去了比較處理間差異的一些重要訊息。在上述的各情形中，使用交叉實驗設計則為不智之舉。

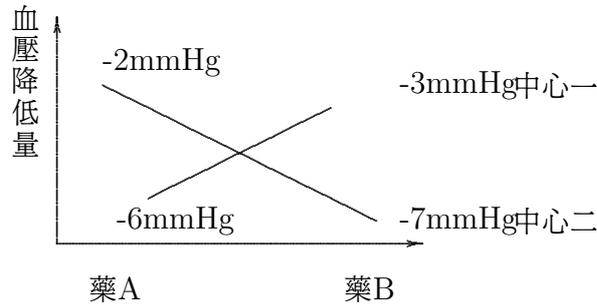
交叉實驗或平行實驗設計的選擇應取決於實驗的目的及性質。一般來說臨床試驗第一期大都採用交叉實驗。而在第二期至第四期則以平行實驗為主。美國食品藥物管理局對此二設計的選擇並無硬性規定。但一般建議在大多數的試驗中採用平行實驗設計。

### 3.2 單一及多元試驗中心 (Single Versus Multicenter Trials)

多數的臨床試驗使用一個以上的試驗中心，因其可在較短的時間內徵募到所需的試驗個體。尤其對一些不易徵募到病患的試驗 (如心臟衰竭 (congestion heart failure)) 若採用單一試驗中心，則試驗將會非常的費時。在工業界，時間即為金錢。若能愈早讓新藥上市，對公司的相對利益也就愈大。因此單一試

驗中心在實行上有實際的困難。此外，參與多元試驗中心的個體來自不同的區域，故較單一試驗中心所徵募到的個體較能代表整個群體。且在單一試驗中心裡，處理的效應將和此中心的效應混淆，無法辨認觀察到的結果是由於中心或是處理的效應。因此單一試驗在臨床試驗中大都不被採用。

然而多元試驗中心亦有其缺失。如中心的效應為顯著時，將會影響處理效應的估計，及在試驗中可能有中心與處理交互作用 (treatment-by-center interactions) 的產生。所謂的交互作用即是在不同的中心產生不同的處理效應。例如藥 A 與 B 為兩種治療高血壓的處理。如圖三所示，在第一中心裡藥 A 與藥 B 的平均舒張壓減低量分別為 6mmHg 及 3mmHg。故藥 A 在第一中心的表現比 B 為佳。然而在第二中心，藥 A 與 B 的平均血壓降低量為 2mmHg 及 7mmHg。因此我們宣稱藥 B 在第二中心的結果比 A 為佳。在此情形下，我們無法結論究竟藥 A 與 B 何者具有較高的療效，因答案隨中心而異。此時試驗中心與處理即有交互作用的產生。若在試驗中有此情形發生，我們應盡力找尋交互作用產生的原因 (例如由於地域差異或其他因素所造成) 以合理地解釋此結果。



圖三. 交互作用

### 3.3 隱蔽 (Blinding) 及開放式 (Open-Label) 試驗

臨床試驗為了避免試驗個體及臨床研究者 (investigator) 心理和在實行時有偏差 (bias), 試驗的過程通常是採隱蔽式的。由於隱蔽對象的不同, 又可分為雙隱蔽 (double-blind) 及單隱蔽 (single-blind) 兩種。

在雙隱蔽的試驗中, 臨床研究者及試驗個體均不知試驗個體接受何種處理。換句話說, 執行及接受試驗的雙方均不知隨機指派計劃表 (randomization schedule), 即不知處理的分配。若試驗為單隱蔽, 只有試驗個體受到隱蔽, 調查者則知整個處理的分配。採用單隱蔽主要是基於道德上 (ethical) 的考量。臨床研究者於是可監控病患對藥物的反應。尤其對一些病情嚴重的患者, 更需要緊密的監視。然而在單隱蔽的試驗中則易產生研究者執行實驗時的偏差。例如臨床研究者若知某處理對病患無益。雖然病患被隨機分至這處理中, 基於職業道德, 臨床研究者可能避免讓病患接受此處理, 而以別種處理取而代之。如此一來, 極可能造成試驗的混亂。

為了要獲得長期的安全及有效性的經驗, 藥廠在短期的雙隱蔽試驗之後, 實行為期一、二年的開放性試驗。在此期間基於安全的考慮, 試驗個體及臨床研究者雙方均免於隱蔽。因此容易產生雙方面的偏差。有鑑於此, 在此階段所採用的統計方法大都只限於敘述性 (descriptive) 而非推論性的 (inferential) 的分析。

### 3.4 隨機指派 (Randomized) 及非隨機 (Nonrandomized) 試驗

在試驗中將病患隨機地指派至各處理中, 我們稱之為隨機試驗 (randomized trials)。反之則稱為非隨機試驗 (nonrandomized trials)。在非隨機的試驗中, 病患可選擇自己的處理。因此會造成選擇上的偏差。故實驗結果較難以推論。除非有特殊原因, 一般臨床試驗多不採用非隨機分配。

隨機指派在統計的實驗設計中扮演了最重要的角色。其目的在於將人為或非人為所造成的偏差減至最低以保護實驗的結果。藉由隨機分配, 每個試驗個體有相同的機會被分配至各個處理中。因此我們可以將非處理

的差異平均分配。例如，一個臨床試驗的目的為測試是否咖啡因能增加解字陣方塊的能力。試驗的處理為咖啡因及非咖啡因飲料。試驗的過程首先將試驗個體分配至二種處理中。在服用飲料半個鐘頭後，計算他們答對的題數。若此時臨床研究者將智商較高的試驗個體分配接受咖啡因而將較低的分配接受非咖啡因，在此情況下發生的試驗結果將無法用於判斷是由於處理或人體智商差異所造成。此即為 3.1 節中所稱的混淆現象。為了將混淆現象發生的機率減至最低，我們採用隨機分配，或更進一步地採用限制性的隨機分配以完全地避免其發生。

臨床試驗中常採用的隨機分配方法有完全隨機試驗 (completely randomized design) 及隨機集區試驗 (randomized block design)。其中又以後者為甚。在完全隨機試驗中，所有的試驗個體有相同的機率分配至任一處理。其做法是將所有試驗個體視為一群體，隨機分配是對整個群體而言。而隨機集區試驗則是以試驗中心為單位，在每個中心內執行隨機分配。例如，某個試驗由三種處理 A,B,C 及三百位病患來自五個中心 (每中心六十人) 所組成。這個試驗若是採用隨機集區設計，則在每個中心內六十位病患被隨機分配至三處理，每個處理二十人 (表一 a)。而若採用完全隨機實驗設計，三百位病患將隨機地被分配至三處理，每個處理一百人 (表一 b)。由此可見，採用完全隨機試驗仍可能發生有些中心的病患完全被分配至某種處理。因此，處理效應和中心效應無法辨認。若採用隨機集區設計，則可避免此情

況發生。病患的分配可保持較平衡的狀態。然而隨機集區試驗亦有其缺失。例如，若試驗的處理數量為小時 (如二種處理 A 與 B)，臨床研究者可能猜測處理分配順序 (如 A, B 或 B,A)，且猜對的機率極高 (1/2)。此時我們若將從藥廠運送處理至試驗中心的單位加大，臨床研究者則較不易揣測分配順序。如兩種處理的試驗以四為運送單位，分配的順序有  $\frac{4!}{2!2!}$  種 (AABB, ABAB, BAAB, ABBA, BABA, BBAA)。調查者猜對此順序的機率則減低至 1/6。藉此，我們可儘量保持試驗雙隱蔽的特性。

表一a 隨機集區設計				表一b 完全隨機設計		
處理				處理		
中心	A	B	C	A	B	C
1	20	20	20	100	100	100
2	20	20	20			
3	20	20	20	(不以中心分層)		
4	20	20	20			
5	20	20	20			

### 3.5 安慰劑控制 (Placebo-Controlled) 及活動性控制 (Active-Controlled)

臨床試驗的主要目的是觀察新藥有效性及其他一些行為。為了達到這個目的，在試驗中除了新藥，亦需加入其他的藥劑作為實驗的控制 (control) 以為比較。一般最常用的控制有安慰劑控制，及其他活動性的控制。安慰劑不含化學活性成份，對疾病本身並無療效。而它仍可能有某種程度的影響。正面的如

止痛,負面的如副作用等。活動性的控制是以其他能治療相同疾病的藥劑為控制,將新藥與其比較,藉以評估新藥的有效性及安全性。

臨床試驗的第二期大都採用安慰劑控制。雖然第二期也做藥量反應 (dose-response) 試驗,用以找出藥品有效範圍及觀察不同劑量所產生的反應值以評估其有效性及安全性。但是藥量反應試驗通常比安慰劑控制試驗需要較多的試驗個體以刻畫出藥劑反應曲線。若用安慰劑為控制以評估新藥,則不需如此多的樣本數。在第三期時,雖然有些試驗仍以安慰劑為控制,然而第三期的試驗期間一般較第二期為長。若讓病患 (尤其是病重者或一些致命疾病的患者) 使用不能真正治療疾病的安慰劑,實為不道德的 (un-ethical) 行為。在此情況下,我們採用可以治療此疾病的其他藥物為控制。將新藥與其他活性藥物作比較有助廠商瞭解將來新藥推出市場後,如何將新藥定位 (如價格) 的問題。使用活性控制的目的通常有兩種: 一是證明新藥與其在有效性方面是相等的 (equivalence), 但新藥有效好的安全性,或廠商願意以較低的價格將新藥推出市場。二是證明新藥比較有效。然而活性控制要達到這兩個目標,一般來說並非易事。此外使用活性控制比安慰劑控制往往需要較多的試驗個體。一般試驗中使用數百個個體並不足以證明兩者是相等或不等。此乃使用活性控制的缺點。

#### 四、結語

一個新藥的產生必須結合了龐大的物力,財力,及融合各方面研究人員的專業知識,如醫學家,藥學家,藥理學家,化學家,及統計學家等等。其中統計在整個過程中扮演了不可或缺的角色。從藥品化學成份的開發,實驗室中的試驗,到人體的臨床試驗,每個階段均需使用統計以確保整個計劃的有效性 (validity)。尤其在臨床試驗階段,有效地設計實驗,及分析才能使試驗結果更具科學效力。統計的實驗設計目的是使估計值 (如藥效) 的偏差及變異 (variability) 減低至最小。換言之,一個良好的實驗設計應具有高度的準確 (accuracy) 及精確 (precision) 度。在第一期的臨床試驗中,常用的試驗為交叉實驗設計。而在第二期至第四期中,常用的試驗則為平行實驗設計。其中一些重要的實驗因素,例如試驗中心數的選擇,應採單隱蔽,雙隱蔽或開放式的考量,應使用安慰劑或其他可治療相同疾病的藥品為實驗之控制,如何隨機分配病患等等問題,我們在第三節中討論了不同選擇的優缺點及考慮的因素。

在設計實驗或分析結果時,統計人員必需和其他部門人員 (如資料管理部門,政府法規部門,藥物安全部門及主持試驗之醫學家等等) 緊密合作,以求最有效的實驗結果。尤其和醫學家更需經常地討論,以確定實驗目的及設計最有效的實驗方法。統計從業人員和醫學從業人員難免有意見相左的時候。一般而言,前者考量問題以科學為出發點,期望試驗結果具有科學的效力。而後者往往從醫學的經驗及道德的觀點上出發。例如,從統計的觀點,我們希望藥品的分配是採雙隱蔽的,

因其可避免人為產生的偏差。然而在某些情況下，醫學家則主張採單隱蔽或完全開放（見 3.2 節）。此時，兩方需溝通及協調，以尋求解

決之道，確保試驗結果的有效性。

—本文作者任職於國立政治大學統計學系—