



臨床試驗的隨機化設計

陳珍信

人們處理日常事務，遇到對兩個方案的取捨模稜兩可時，往往會擲個銅板或抽個籤，藉無法預測的公平機會解決難題。探求某一新方案效益的科學推論，若採取無對照方案的觀察研究，在因果關係的說明上就難有準繩。有對照的實驗中各方案裏接受實驗的個體若具相似的品質，實驗的誤差才不易產生。隨機化配置各個體到各方案，就可提供各受驗方案立足點的平等。

統計學一派宗師，英國的斐變（Ronald Aylmer Fisher）的成就之一是在一九二〇年代發展出以隨機化為基本原理的實驗設計。這學域為他服務的農場的育種實驗與統計推論都留下重大貢獻。其實，追溯十九世紀末葉，美國著名的科學哲學家斐珥思（Charles Sanders Peirce）已闡釋隨機化在科學推理上的重要角色。他這觀念直接影響了實驗心理學的統計方法。近年來，日本及美國等先進國家的工業界也大量引進實驗設計，獲致品質改良在研究發展上的突破。

臨床試驗

臨床試驗是醫學界藉前瞻性計劃所從事的

科學研究工作，目的在評量被用於人體的預防、診斷或治療上的藥劑或方法的效益。一九四〇年代，黑鰲（Austin Bradford Hill）鼓吹「對照隨機化試驗」（Controlled Randomized Trials），領導英國醫學研究院的統計工作，樹立了成功將隨機化原則引用於醫學研究的典範。二次世界大戰後，臨床及預防醫學研究多借重臨床試驗所累積的經驗。

電子元件可同時大量製造且規格整齊，而人體間生理及心理的差異大，所以臨床試驗不同於工業界實驗可將元件同時配置到不同的實驗組。病人的徵集常是成年累月的，而且病人就診時須儘速治療，「循序配置」是臨床試驗的特徵之一。為了達成有效的統計分析著想，臨床試驗在設計階段也須預估所需病人的數目。然而，有時為作期中分析報告；或更惱人的是由於藥的副作用毒性過多，有時得提早結束試驗；這些時候，各治療組之病人數若相差過多，統計分析的準度就差。因此，各組人數的平衡也是設計上的重點之一。

另外，臨床試驗易形成「選取偏差」。假若某位參與試驗的醫師偏袒新藥，又預知下個配置組號是新藥組，這時來個病情稍重的人，醫師可能會主觀的排除這病人。這樣的安排可能產生新藥較有效的結果，因為病情輕的較多服用新藥。若採取系統性變化的配置，例如

$ABABAB \dots$ ，就更易導致「選取偏差」。

通常，臨床試驗可多於兩個治療組，但為說明上的方便，以下的討論都以 A 、 B 兩組為例。我們接著介紹臨床該驗上常用的四種隨機化設計。

簡單隨機化設計

「簡單隨機化」(Simple Randomization) 就是每當病人進入試驗時，拋擲一個正反面出現機率相同的銅板，若出現正面，這病人即配至 A 組，否則他即進入 B 組。當然這設計也可利用隨機數表來執行，逢 0 至 4 的數字即配 A 組，5 至 9 則配 B 組。此法的好處是執行簡單又無法事先窺知配置，最重要的是合乎傳統統計推論的原理。在兩組治療真無差別的前提下，分析所得的兩組反應的差異就可全歸因於純粹的隨機誤差。

當病人總數趨近無窮大，機率論的大數法則保證這設計的組間人數呈平衡狀況，而且各組皆有相似分佈的病人特質（如性別、年齡及預後因素）。「完全隨機化」(Complete Randomization) 是簡單隨機化的別稱，正是這性質的最佳說明。

實際應用上，病人不多時，採用本法偶會滋生不平衡的困擾。下表列出在各固定的病人總數下所對應的兩組人數比。按二項機率分佈來計算，表列的兩組人數與較其更不平衡的配置（如病人總數為 10 時，2:8、1:9 及 0:10 三個配置）出現的機率都至少是百分之五。

病人總數	兩組人數
10	2:8
20	6:14
50	18:32
100	40:60
200	86:114
500	228:272
1000	469:531

我們可以看出，在病人總數超過 200 時，有些統計學家建議不必擔心此設計不平衡趨勢的理由了。以下的三種設計都著重在平衡的配置。

隨機排列區集設計

隨機排列區集設計 (Random Permuted Block Design) 的著眼點在求組間人數的平衡。每隔 $2S$ 個病人的一區集內是 S 個 A 及 S 個 B 的隨機排列，而 $2S$ 就稱為區集大小。這設計的優點是每隔固定人數，組間必達完全的平衡；任何時段的組間人數差在 S 之內。假如病人型態隨時節變化的話，本法易形成相似的治療組。而且，試驗提早結束時，組間平衡仍存在。

不幸的是若醫師知道設計的區集大小並記住區集內前 $S-1$ 個病人的組號，那第 S 個病人的組號就暴露了，「選取偏差」就可能產生。

以條件機率計算，每個病人進入 A 或 B 組的機率還是相等，祇是不成為獨立事件的序列。傳統統計分析的基本假設是每個隨機變數是獨立的來自相同的機率分佈，與這設計的配置並不切合。若一試驗需 200 名病人，我們做好 100 支 A 籤及 100 支 B 籤置於箱內，病人一到就抽支籤，這是唯一一區集而其大小為 200 的排列區集設計，而不是簡單隨機化設計。

偏頗銅板設計

偏頗銅板 (Biased Coin) 是正反兩面出現機率不同的銅板，設正面出現的機率是 $p > 0.5$ 。這設計是兩組平衡後的下一個配置採簡單隨機化；不平衡後的處理是擲銅板得正面則配病人至人數較少的那組，得反面則至人數較多的那組。這是葉符隆 (Bradley Efron) 折衷前述兩個設計所提出的構想，在強制達成

平衡的同時，又顧及隨機性。他並以數學證明這設計的若干優點，包括使用馬可夫鏈的定理。以醫師用最佳方法去猜中病人配置組號次數的期望值來度量「選取偏差」時， $p = 2/3$ 的偏頗銅板設計與排列區集大小為18的設計相似，可見偏頗銅板設計具有相當好的隨機性。

袋球模式設計

偏頗銅板設計無視於組間不平衡度的大小，皆以相同的 P 來強行平衡調配。我國旅美統計學家魏立人教授有鑑於此瑕疵而研究其改良型式，「袋球模式」(Urn Design) 即為此型的一個例子。設一袋原有 α 個白球及 α 個黑球。隨機取一球，若是白球，病人配至 A 組；反之配至 B 組；再放該球入袋，並添 β 個異色球入袋。這設計即重複取球，記為 $UD(\alpha, \beta)$ 。當 $\alpha = 0$ 時，初次取球以相等機率決定，再添 β 個異色球入袋，這種 $UD(0, \beta)$ 顯然與 β 大小無關，是個好設計。當然， $UD(\alpha, 0)$ 是簡單隨機化設計。

在不同設計下的統計分析

排列區集、偏頗銅板及袋球模式三種設計，尤其是後兩者，折衷考慮平衡度與隨機性，裨益病人不太多的臨床試驗之實施。若以當今盛行的個人電腦來配合，執行上更加方便。反對這類循序配置設計的最主要理由是它們導致複雜的隨機分佈，也就產生統計分析的艱難，因為這類設計所得的數據並非獨立而分佈相同的隨機變數了。目前幾乎所有統計推論都建立在簡單隨機化的架構上，所以我們舉個例子來談分析上的問題。

假設有二十名病人參加臨床試驗，他們的循序配置組號及治療一年的結果(以 \bigcirc 表生， \times 表死)如下圖：

ABAB BBAA BBAA BAAB ABBA
 $\bigcirc\times\times\times \times\times\times\bigcirc \times\times\bigcirc\bigcirc \times\bigcirc\bigcirc\times \bigcirc\bigcirc\bigcirc\bigcirc$ (圖一)

其中的 A 、 B 分別為新藥組與對照組。這試驗設計時即想求證新藥對病人存活比對照的舊藥有效。數據顯示 A 、 B 兩組的平均存活率為 0.8 與 0.2 我們想檢定 A 組比 B 組有效是否具有統計學上的意義。這個待求證的命題就是統計檢定術語裏的「對立假設」(Alternative Hypothesis)。要否定「對立假設」的命題，即新藥不比舊藥有效，稱為「虛無假設」(Null Hypothesis)。在這有關存活率比較的問題上，我們若將「虛無假設」縮小為兩種藥一樣有效，或兩組存活率相等，可得同樣的檢定模式。這可用數理統計來證明，我們不再說明。在虛無假設成立的條件下，所謂的極端便是對立假設的方向。我們可計算在虛無假設下， AB 兩組存活率比所觀察到的數據還趨極端的機率，來判定「虛無假設」成立的可能性。這機率簡稱為 P 值。若 P 值極小，表示我們觀察的結果似不可能發生，以反證法來說，虛無假設便不成立了。反之，若 P 值不小，則無充分證據判定虛無假設不成立。

例設這試驗是簡單隨機化設計(簡記為 SR)，圖一的結果可歸納成下表：

	A	B	
\bigcirc	8	2	10
\times	2	8	10
	10	10	20

比較兩種比率的問題，斐變的精確檢定(Fisher's Exact Test)是極佳的統計檢定法。在虛無假設下，兩組死亡的機會是均等的。在固定兩組人數 a 與 b 及死亡病人總數 d 的條件下，先計算 A 組有 y 個病人死亡的機率。

	A	B	
\bigcirc			$(a+b)-d$
\times	y	$d-y$	d
	a	b	$a+b$

由組合的基本觀念，我們知道隨機變數 Y 有個超幾何機率分佈 (Hypergeometric Distribution)

$$Pr(Y=y) = \frac{\binom{a}{y} \binom{b}{d-y}}{\binom{a+b}{d}},$$

當 $y=0, 1, 2, \dots, d$ 。這例子裏， $a=b=d=10$ ，觀察到 $y=2$ 。因此， $y=0$ 及 1 是比 $y=2$ 更趨極端的結果。斐變建議計算 P 值如下：

$$P = Pr(Y=2) + Pr(Y=1) + Pr(Y=0) = 0.0112,$$

表示 A 組的治療優於 B 組具統計意義。

若這試驗是採取排列區集設計如圖一所示，五個區集而各區集大小為四 (簡記為 $PB[4]$)，則斐變的精確檢定無法適用。各區集的存活情況可由下表看出：

(區集一)			(區集二)		
○	A	B	○	A	B
×	A	B	×	A	B
	2	2		2	2
	1	0		1	0
	1	2		1	2
	2	2		2	2
(區集三)			(區集四)		
○	A	B	○	A	B
×	A	B	×	A	B
	2	0		2	0
	0	2		0	2
	2	2		2	2

(區集五)

	A	B	
○	2	2	4
×	0	0	0
	2	2	4

在固定各區集的 a 、 b 及 d 的情況下， A 組至少死兩名病人，祇有由區集一與二各死一名的唯一途徑。利用超幾何機率計算五個區集依序死 1、1、0、0、0 個病人的機率，相乘後即得 P 值為

$$P = \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) \left(\frac{1}{6}\right) \left(\frac{1}{6}\right) (1) = 0.0069.$$

若圖方便，將此設計的結果用建立在簡單隨機化基礎的統計方法來分析，所得的上段 P 值為 0.0112 顯然高估了。雖然兩個分析都結論 A 組的治療優於 B 組具統計意義，但正確的使用以區集設計考慮的分析，對此結論有更大的把握。

若此試驗是以偏頗銅板 $p=3/4$ 設計 (簡記為 $BC[3/4]$) 而得如圖一的配置，依此設計考慮的分析，可算出 P 值為 0.004，若錯以簡單隨機化來分析，則 P 值被高估了。偏頗銅板設計在二十名病人的情況下，計算較繁瑣，我們就以一個簡單的例子來說明不同的設計雖得一樣的配置及反應結果，統計分析的結果也會不同。

假設有四個病人，就有十六個可能的配置。下表就各種設計中八個配置可能出現的機率

各種配置出現的機率

配置	SR	$PB[4]$	$PB[2]$	$BC[3/4]$	$UD(0,1)$	$UD(1,1)$
(1) A A A A	1/16	0	0	1 / 128	0	1 / 120
(2) A A A B	1/16	0	0	3 / 128	0	4 / 120
(3) A A B A	1/16	0	0	3 / 128	0	6 / 120
(4) A A B B	1/16	1/6	0	9 / 128	0	9 / 120
(5) A B A A	1/16	0	0	6 / 128	1 / 12	8 / 120
(6) A B A B	1/16	1/6	1/4	18 / 128	2 / 12	12 / 120
(7) A B B A	1/16	1/6	1/4	18 / 128	2 / 12	12 / 120
(8) A B B B	1/16	0	0	6 / 128	1 / 12	8 / 120

列出，其他的八個配置出現的機率可用對稱法求得，如 $ABAB$ 與 $BABA$ 出現的機率相同。若試驗的配置結果是 $ABBA$ ，反應結果為 A 組全活但 B 組全死，則各設計下所分析得的 P 值為

$PB[4]$	$PB[2]$	$BC[3/4]$	$UD(0,1)$	$UD(1,1)$
$1/6$	$1/4$	$1/5$	$1/4$	$2/11$

計算的根據為 A B 兩組各有兩名病人及死亡總數為二的條件是固定，另外的細節是

$$1/5 = 18/128 \div (9 + 18 + 18 + 18 + 18 + 9) / 128,$$

$$2/11 = 12/120 \div (9 + 12 + 12 + 12 + 12 + 9) / 120.$$

若錯用以簡單隨機化為基礎的統計模式，即用斐斐精確法計算，就得 $P = 1/6$ ，導致不同的結果。

針對這設計影響統計分析的問題，葉符隆及魏立人教授在大樣本情況下分別提出機率分佈逼近結果的建議及證明，這些屬於高深統計研究的題材，本文不再討論。

臨床試驗的隨機化設計下的統計分析，本文祇討論兩個比率相較的例子，其他的問題則更繁複。當初祇是直覺的改變簡單隨機化設計，以折衷平衡度及隨機性，隨後的統計分析沒料到須大費周章。另外，試驗設計中有時也須注意重要預後因素平衡的調整，如白血球病應配置兩組病人白血球多寡分布要相似，這是層化 (Stratification) 的考慮。應用上，隨機化與層化可配合使用，我們也不細談了，祇要假設以上的設計所得的兩組病人之特質相似便可。

隨機性是個簡單方便的觀念，也是統計學的基礎，在臨床試驗設計實用遭遇的問題，正可說明科學的發展有賴理論的應用與修正。

參考資料

1. Efron, Bradley (1971). *Forcing a Sequential Experiment to Be Balanced*, *Biometrika*, 58, 403-417.
2. Wei, Lee - Jen (1977). *A Class of Designs for Sequential Clinical Trials*, *Journal of the American Statistical Association*, 72, 382-386.
3. Halpern, Jerry & Brown, Byron Wm., Jr. (1986). *Sequential Treatment Allocation Procedures in Clinical Trials - with Particular Attention to the Analysis of Results for the Biased Coin Design*, *Statistics in Medicine*, 5, 211-229.
4. Wei, L. J., Smythe, R. T. & Smith, R. L. (1986). *K-treatment Comparisons with Restricted Randomization Rules in Clinical Trials*, *Annals of Statistics*, 14, 265 - 274.

— 本文作者任職於中央研究院統計所 —