

百農嘗一草——

實驗設計之應用

羅夢娜

國立中山大學應用數學系

「二重隱瞞」的實驗設計，使測量反應者與受驗者都不知曉受驗者到底接受何種處理，以避免主觀的期望或偏見而影響實驗的結果。

現代的人生活越來越緊張及忙碌，感受到各種不同的精神上的壓力，於是頭痛不再是陌生的名詞。假設有人宣稱他發明了一種新的藥可治療頭痛，我們應怎樣設計一個實驗來評估它的效果？基本上應採用比較的方法，將接受實驗者分為兩組，一組服用這種新藥，稱為處理組（treatment group）；一組則不服用這種新藥稱為控制組（control group）。然後再比較兩組接受實驗者的反應。根據經驗顯示，實驗者應屬何組，需採隨機化方式決定，以消除一些外來的偏差。

此外在測量反應時，因為牽涉到各種主觀的評估或臨床的判斷，應盡量使測量反應的醫生或受驗者都不知道受驗者所接受的處理為何，以避免因主觀的期望或偏見可能會影響實驗的結果。這種實驗的方式稱為「二重隱瞞」（double blind）。而為了達成二重隱瞞，即使對於控制組，實際上不接受任何藥物，仍應給予不含任何藥劑但與新藥形式完全相同的藥片或注射，一般稱控制組中所接受的藥片或注射為安慰劑（placebo）。在美國曾經實地做過一個歷史上規模最大的醫學實驗，我們可用它來說明這種想法。

沙克疫苗之實驗

1916 年，美國首次發現急性骨髓灰白質炎（poliomyelitis）的傳染病。在其後的 40 年中，數以千計的病人因受到感染而死亡或身體部分變麻痺，其中受害者以兒童居多，故又稱為小兒麻痺症（polio）。

到了 1950 年代時，已發現數種對抗小兒麻痺的疫苗，其中又以沙克（J. Salk）所發現的疫苗看來最有希望。在實驗室的試驗中，已證實了它的安全性，並能產生對抗小兒麻痺的抗體。為了證實它對兒童的確有產生抗體的效果，需要大規模實驗室外所做的臨床試驗結果來支持。在 1954 年，美國公共衛生署組織了一個歷史上規模最大的醫學實驗。大約有 2 百萬兒童參與這實驗。有 50 萬左右的兒童接受了疫苗，約 100 萬兒童有意使其不接受疫苗，約 50 萬兒童拒絕接受疫苗。總共花費超過了 500 萬美元。從全美國較易受到小兒麻痺感染的一些學區中，選擇了最易受到傳染的小學一、二、三年級的兒童來做實驗。接受疫苗的兒童就形成所謂的處理組，未接受疫苗的兒童即形成所謂的控制組。雖然兩組人數不同，但可以得病的比率來做比較。

一個令人困擾的問題是，為何不讓所有的兒童都接受疫苗，然後再看之後的得病比率是否比以前降低即可？原因是小兒麻痺是一種流行性的傳染病，它的

DBRC NFIP 設計

得病率每年都不一樣。如 1952 年在美國大約有 6 萬人得病，但 1953 年大約只有一半，若無控制組，則 1954 年的低得病率可能代表兩種意義：沙克疫苗的確有效，或那年小兒麻痺並不十分流行。所以需要處理組及控制組來比較是很明顯的。

但另一問題又產生了，即兒童接受疫苗，需要家長的同意；所以一種可能的設計方式是凡有家長同意的兒童即形成處理組，家長不同意的形成對照組。但如此的設計又有新的問題產生。一般高收入家庭的家長比低收入家庭的家長願意讓自己的孩子接受疫苗，但如此的設計對疫苗而言是較不利的安排。乍看之下似乎十分奇怪。因為小兒麻痺是一種與衛生環境有關的疾病，對於那些父母收入較低，衛生環境較差的兒童，幼年時還具有由母體而來的抗體，很可能即感染過輕微的小兒麻痺，那些存活者以後就免疫了；但高收入家庭一般比較注重衛生，所以來自這些家庭的兒童比較沒有機會在小時候感染而產生抗體，日後感染得病的比率也會較高。

為避免因這些其他外來因素與處理效果混淆不清而產生偏差的結果，應當盡量使處理組及控制組除了處理不同外，其他都愈相似愈好。如此才能在看到反應結果不同時，下結論說差異是來自處理的因素。

NFIP DBRC 設計

在 1954 年的實驗中，美國小兒麻痺基金會（NFIP）建議使用一種設計，即將所有獲得家長同意的小學二年級學生注射疫苗，一年級和三年級學童則不予注射，當作控制組。許多學區都採用 NFIP 的設計做實驗，但是它有兩大缺點：一小兒麻痺是一種傳染性疾病，由接觸傳染，所以有可能產生二年級學童中的病例特別多，對疫苗效果有不利的傾向，或二年級學童中的病例特別少，對疫苗較有利；二在處理組中之學童因均獲得家長同意，所以家庭因素可能影響結果，對疫苗不利。

另有許多學區在看到 NFIP 設計方式之缺點後，採用另一種設計，即所謂二重隱瞞隨機化控制設計（doubly blind randomized controlled experiment, DBRC）。為了公平起見，控制組的學童須從與處理組相同的群體中選出，此群體即為獲得家長同意的兒童。否則家庭因素的影響會與疫苗效果混淆不清。在分配兒童屬於控制組或處理組時，為使二者愈相似愈好，似乎還應考慮其他因素：如家長的收入、兒童的健康情形、性格、交友習慣等。但根據經驗顯示，在做這類選擇時常常會產生偏差。所以最好的方法就是採用隨機化的過程。用一組亂數表碰到奇數代表控制組，偶數代表處理組，或丟一公正的銅板，每人都有 50 % 之機會進入控制組或處理組。以如此客觀、隨機化方式劃分處理組與控制組的實驗，則稱為隨機化控制實驗。

此外如控制組兒童亦須接受看來像疫苗的一種鹽水安慰劑之注射，以避免心理作用的影響。另外亦採用二重隱瞞的方式，即兒童本身與檢查診斷者均不知道兒童是否注射過疫苗，以免診斷時因為主觀的期望及偏見會產生偏差，以致影響實驗的結果。在測試沙克疫苗的實驗中，這是最好的一種設計了。

兩種方法之比較

究竟這次歷史上最大的醫學實驗結果如何呢？表一顯示在隨機化控制的實驗中，處理組的得病率比控

制組的得病率低很多，可做為沙克疫苗的確有效之強而有力的證據。

表：1954年美國沙克疫苗之實驗結果。

二重隱瞞隨機化控制(DBRC)實驗			美國小兒痳痺基金會(NFIP)實驗		
	人數	得病率		人數	得病率
處理組	200,000	28	二年級(處理組)	225,000	25
控制組	200,000	71	一、三年級(控制組)	225,000	54
父母不同意	350,000	46	二年級(父母不同意)	125,000	44

註：表中得病率指每10萬名兒童中得病之人數

二重隱瞞隨機化控制實驗除了使偏差減至最小外，還具有其他技術上的好處。如費雪(R.A.Fisher)在他所著的實驗設計一書中所述，實驗的隨機性使我們可用統計學上的知識做一些分析測驗。在此究竟每10萬人有71與28的差異是因為機運，與處理無關，還是由處理所造成？NFIP設計，其結果受許多其他因素影響，無法分辨出處理的真正效果；事實上亦可由附表看出的確會造成偏差，在前述實驗中處理與控制有 $61\%[(71 - 28)/71]$ 的差異，而NFIP實驗中卻只有 $54\%[(54 - 25)/54]$ 。這些誤差在此尚未對實驗結論有很大的影響。下面我們再舉一例說明不良設計常會將研究者誤導入歧途。

胃凍結法

1958年，王更(Wangensteen)醫生發明了一種新的治療胃潰瘍的方法，稱為「胃凍結法」(gastric freezing)。病人被麻醉後，在他的胃中放入一氣球，然後將冷媒(coolant)由氣球注入，將胃凍結將近一小時。這樣可以暫時停止消化過程而使潰瘍開始慢慢痊癒。王更用這方法治療了24位病人，所有病人都痊癒了。因此有一段時間他的技巧變得十分受歡迎。因為傳統的治療法需要動大手術。他的方法顯得容易得多。但是仍有許多醫生對王更的結果表示懷疑。其中一個主要原因就是他沒有控制組。

1963年時，另一位若飛(Ruffin)醫生設法安排了一個DBRC實驗，來評估胃凍結法。這個實驗安排在5個不同醫院中進行。用隨機方式指定82位病人接受處理，78位病人列為控制組。在處理組的病

人，真正接受胃凍結法治療；屬於控制組的病人則接受假的胃凍結法。其過程與王更治療方式相同，只是在氣球內做了個分流器，使冷媒在還沒凍結胃前就流回去了。這實驗的事後評估工作進行了兩年，由一些並不知道病人是否接受了真正的胃凍結治療的醫生來診斷。若飛醫生的結論如下：

大部分的病人在接受治療後的前六星期都覺得病情有改善(處理組有47%，控制組39%)，或沒有繼續發作。但再過一段時間後，兩組中大部分病人都再度復發而且變得更嚴重(在24個月後，控制組中有39%及處理組中有45%)。由此結果我們可以合理的假設，早期所得到有關病人覺得疼痛減輕，病情改善的報告可能來自於治療過程所造成的心靈作用。

王更的實驗設計不佳，因為它缺乏控制組。它過分誇大了胃凍結法的效果。若飛的實驗經過良好設計——採用DBRC方式，它證明了胃凍結法是無效的。

本文強調在臨床試驗中隨機化地將病人分至處理組及控制組，以及二重隱瞞的重要性。忽略了這些原則，我們便會很容易被導入歧途而得到錯誤的結論，不但浪費金錢、人力，而且延誤了真正有效的治療，對病人健康影響甚巨。我們在做這類實驗時豈可不慎乎！中國自古即有神農嘗百草的精神，可惜未能深入探討改進的方法，否則實驗設計或有可能源自於中國哩。

(本文取材自D. Freedman, R. Pisani and R. Purves, Statistics, Ch. 1, "Controlled experiments".) □