

## 第二章 實驗 Experimentation

實驗有兩個目的：(1) 觀察迄今未知或未加解釋的新事實；(2) 檢證為某一理論提出的假設是否與可視察到的事實相符合，但兩者往往是互不相干的。

- 雷內·杜博斯 -

生物學實驗·實驗的部署與估價·給人錯誤印象的實驗·提要

### 生物學實驗 *Biological experiments*

我們今天所認識的科學，可說是從文藝復興時期採用實驗方法開始的。然而，儘管實驗對於大多數學科都很重要，卻並非適用於一切的科學研究。例如，在描寫生物學、觀察生態學或者各種類型的醫學臨床研究中，都不用實驗。但即便如此，後一類型的研究也利用了很多同樣的原則。其主要不同點在於：假設的檢驗是從自然發生的現象中收集資料，而不是從人為地實驗條件下的現象中收集資料。在寫上一章最後部分和本章第一部分時，我的對象是實驗人員。但是，對於純觀察的研究工作者，這些章節可能也有幫助。

通常，實驗在於使事件在已控制的條件下發生，儘量消除外界不相干因素的影響；並能進行密切的觀察，以便揭示現象之間的關係。

『對照實驗』是生物學實驗中最重要的概念之一。在『對照實驗』中有兩個或兩個以上的相似組群（除了一切生物體所固有的變異性外，其餘的條件完全相同）：一個是『對照』組，作為比較的標準；另一個是『試驗』組，要通過某種實驗步驟，以便人們確定它對試驗的影響。人們通常使用『隨意抽取樣品』的方法來編組，即用抽籤或排除人為挑選的方法，把樣品分別編入甲組或乙組。按照傳統的實驗方法，除要研究的那一個變數外，各組其它一切方面都應儘量相似，而且實驗應該很簡單。『一次變化一個因素，並把全部情況進行記錄。』這一原則現仍廣泛採用，特別在動物實驗方面。但有了現代統計方法的幫助，現在已有可能設計同時試驗幾個變數的實驗了。

在研究工作的開頭應該盡可能進行一項關鍵性的簡單實驗，以判斷所考慮的主要假設是否成立。細節的計劃則可稍後作出。因此，在對各部分作試驗之前先對整體作試驗，往往是明智的。例如：當你想用純細菌培養物再次引起疾病之前，最好先試著用帶病組織傳染；在試驗化學分餾物的毒性、抗原性及其它影響前，應先試驗其原始提取物。這一原則看來似乎簡單、明顯，但常被忽視，從而浪費了許多時間。同樣這個原則，在初步試驗某個定量因素影響時，通常最好在一開始就斷定在極端條件下，例如使用大劑量的條件下，是否會影響實驗的結果。

與此十分相似的另一條原則是 逐步排除 的方法。有種猜謎遊戲，諸如提出『動物、植物、還是礦物』等一連串問題，就很傳神說明了這種方法。用逐步縮小可能性

的方法，常常比直接但是盲目的猜測能更快地找到未知的事物。該原則應用於稱重時，先試驗過重和過輕兩個重量，在實驗中所會產生的影響，然後使這兩個極端重量逐步接近。在用化學方法尋找一種未知的物質時，這種方法特別有用；但是這種方法也同樣經常應用於生物學各個分支領域中。例如，在研究某種疾病的起因時，有時，我們排除各種可能的選擇方案，最後只剩下一個可能性最高的方案。以便集中精力進行研究。

在生物學上，開始的時候進行一種小規模的初步實驗往往是一種好方法。除了經濟上的考慮以外，在最初階段就進行複雜的實驗，試圖對所有的問題作出全面的回答，往往很難得出理想的結果。不如讓研究工作分階段逐步進展，因為後面的實驗可能要根據前面實驗的結果加以修訂。『試點實驗』是初步實驗的一種，常用於以人或家畜為對象的實驗中。這是一種小規模的、往往是在實驗室進行的實驗，旨在確定是否值得進行全面的現場實驗。另一種初步實驗是『觀測實驗』，目的是為主要實驗的部署尋找指導原則。讓我們以傳染因子或毒性因子的活體滴定分析為例。在『觀測實驗』中，稀釋度間隔很大(如一百倍)，用於每一稀釋度的動物很少(如兩頭)。取得結果以後，在可能的滴定終點兩側再選擇間隔小(如五倍)的稀釋度，同時使用較大的動物群(如五頭)。通過這種方法，我們便可用最少量的動物而獲得準確的結果。

所謂的『篩選』試驗也是一種初步實驗。這是用大量物質進行的一種簡單的試驗，目的在於找出其中那一種東西值得進行進一步試驗，譬如藥物之類的東西。

偶爾，可以安排一種小型實驗或者試驗，以便在短時間內指出某一想法中某部份的佐證不夠，不值得進行大型的實驗。這類性質的簡略實驗有時可以如此安排：若是得出這樣的結果就有價值，若是得出那樣的結果就沒有價值。然而，這裏有一個最低的限度，即使是初步實驗，把『規模』降低到這個限度之下也就沒價值了。假如實驗確實有進行的價值，則應將其安排得至少很有可能取得有用的結果。由於急躁或是缺乏資金，年輕科學家往往魯莽從事，進行計劃不周的實驗，而這種實驗幾乎不可能取得有意義的成果。當簡略實驗被預計可能會得出可靠結果時，它才有資格成為複雜實驗的簡略實驗，也必須這樣我們才有理由進行這種簡略實驗。研究工作要在每一初步問題確定無疑以後才能進展到下一步，否則全部工作可以被稱為『草率馬虎』，這樣說是很恰當的。

作為一次成功的實驗，其最基本條件是要能再現(重新安排條件相同的實驗而得出相同的結果)。在生物學實驗上，這一條件經常很難滿足。在已知因素未變的情況下，如果實驗的結果不同，我們通常認為是由於某個或某些未知的因素影響著實驗的結果。我們應該歡迎這種情況；因為尋找未知的因素可能導致有趣或意外的發現。正像我的一位同事最近對我所說：『在實驗出毛病的時候我們得出了成果。』然而，我們首先應該知道是不是實驗技術上出了錯誤，因為最常犯的是技術上的錯誤。

做實驗的時候，在技術重點上採取極其審慎的態度是非常值得的。一種新技術方法的發明人，由於他能勤勉刻苦，對重要的細節小心重視，有時能夠獲得一些結果；相反的，對該課題不夠熟悉、不夠刻苦用心的研究人員是難以獲得的。就這一點而言，

卡萊爾 (Thomas Carlyle) 所說『天才就是無止境刻苦勤奮的能力』一語千真萬確。賴特 (Almroth Wright) 爵士選用羅林斯 (Rawlings) 傷寒桿菌菌株作為預防傷寒的接種疫苗，就是一個很好的例證。直到最近，由於利用了某些新技術，人們才發現羅林斯菌株是一種製作疫苗的極好菌株。賴特當初慎重地選擇了這種菌株，其所依據的理由在大多數人看來都是微不足道的。西奧博爾德·史密斯是一位難能可貴的大細菌學家，他談到研究工作時說：『決定實驗結果的，正是我們對於表面上微不足道、枯燥乏味、而且不勝麻煩的細節時，所採取的謹慎小心的態度。』

然而在這方面我們應該要有辨別能力，因為可能在無關緊要的細節上大做文章，以致浪費了時間。

對實驗工作的全部細節作詳盡的記錄，是一條最基本而重要的規則。人們經常需要回過頭來參考以前的某個實驗細節，而該細節所具有的意義在進行實驗的時候還意識不到。這種情況發生之頻繁是令人驚訝的。巴斯德保存的筆記就是這種精心記錄全部細節之重要性的出色例證。除了為所做的研究和所觀察到的現象提供可貴的記錄外，做筆記也是促使自己進行更為細緻觀察的一種有用方法。

實驗人員必須正確認識自己所使用的方法，認識這些方法可能受到的限制及其所能達到的精確程度。他們必須非常熟悉實驗用的方法，才能把它用於研究工作，並且必須能夠取得穩定可靠的結果。任何方法都難免要出差錯，難免會得出使人誤解的結果，實驗人員應能迅速發覺這類問題。如果可能，關係重大的定性和定量分析都要用另一方法加以核對。科學家對於自己的儀器也必須有所了解。現代的複雜儀器常常很便利，但也並不都是穩妥可靠的。所以有經驗的科學家常常避免使用這種儀器，以免產生給人錯誤印象的結果。

在實驗對象是人或珍貴家畜，只能對他們進行有限的對照實驗時，常常出現一些困難。如不能滿足對照實驗的基本要求，則最好放棄這種嘗試。這種說法看起來似乎簡單明瞭，但是，研究人員常常覺得困難太大，而採取妥協辦法而作了一些無用的安排。具備一個令人滿意的對照組是必要的，而這種必要性決不是用極多的實驗數量能代替的。兒童接種卡介苗的經過就是一個極好的例子。卡介苗的接種是二十五年前採用的，當時認為它能使人們免於感染結核病。儘管二十五年來作了大量的實驗，但是直到今天，卡介苗在歐洲人種免疫問題上是否有價值仍有爭論。由於對照組中實驗的對象都經不起嚴格的比較，大多數的實驗都說明不了問題。威爾遜 (Graham Selby Wilson) 教授有關卡介苗接種的評論很好地說明了實驗工作的困難和可能的危險。他的結論是：『這些結果表明，在從事以人為對象的對照實驗時，非常重要的一點是保證受接種的兒童和作為對照的兒童，在包括下列各種因素在內的各方面都十分相像：年齡、種族、性別、社會、經濟以及居住條件、智力水準、父母合作程度、感染疾病的可能性、享受幼兒保健等機構的福利，以及生病時所得到的治療。』〔譯註：嚴格說來並沒有絕對的對照組，而只有令人滿意的對照組，因為二個對照的實驗（例如一個施加某種藥物，另一個不施加這種藥物，把這兩個實驗結果加以比較）中被測試的對象無法完全相同，我們只能盡量要求那些可能影響實驗結果的因素相同，當然這只是一種假設而已，雖然科學實驗留有這個破綻，但也由於如此，我們才能繼續不斷的縫補它，如果一切都很完美，那我們也不可能

再進步了。]

威爾遜教授在談話中向我指出：一種據稱是對人類有效的藥物，如果不在實際使用之前做過有決定意義的實驗，那麼想在以後用未經此種治療的對照組進行比較的實驗，簡直是不可能的。日後，這種據稱是有效的藥物就被普遍採用了，至於它是否真正有效，則無從得知。例如，巴斯德的狂犬病治療法從未經過充分的實驗，以證明這種治療法對於被狂犬咬過的人們有防止狂犬症的效能。因而，一些權威人士懷疑該療法是否真有價值。但是，現在已經不可能再去進行一種試驗：那就是不給受狂犬咬過的對照組的人進行此種治療。

使各組處於不同環境之中，有時是現場實驗中不可缺少的組成部分。在這種實驗中，人們難以確定觀察到的差異，究竟是由觀察的特定因素所造成的，還是由與環境有關的其它變數所造成的。有時，我們用這樣的方法克服這個困難，照樣再編成試驗組和對照，從而，發現甚至消除環境造成的任何影響。如果不能排除那些觀察到、但認為是外來的變數，那麼，也許有必要使用一系列的對照組，或進行一系列的實驗，以便把所比較的兩個群體中，每一已知的差異從實驗上分離出來。

如有可能，應該用某種客觀的尺度來評定實驗結果。然而這一點有時做不到，比如，當實驗結果是關係到臨床症狀或是對有機體組織結構變化進行比較時，對實驗結果的評定可能會受到主觀因素的影響。這時，應該保證判斷實驗結果的人確實不知道每一個個體屬於那一組，以求其判斷的客觀性，這一點是重要的。不論科學家確信他自己是如何的客觀，如果在進行判斷時他知道那些病例屬於那一組的話，就很難保證他的判斷沒有下意識地受到偏見的影響。一個責任心強的實驗人員，由於意識到這種危險，甚至也有可能犯那種使他的判斷偏向到與預期結果相反方向上的錯誤。當然，智力活動上完完全全的忠實態度是實驗工作的首要基本條件。

在完成了實驗，並且必要時借助生物統計學評定了實驗結果以後，應將實驗結果與已知有關該課題的一切加以聯繫，作出解釋。

### 實驗的部署與估價 *Planning and assessing experiments*

生物統計學，即把數理統計的方法應用於生物學，是一門比較新的學科。它對研究工作的重要性直到最近才得到普遍的承認。第一章已經提到了有關這個學科的著作，這裏我不想多說，只想提醒讀者注意幾個普遍性的規律，並強調指出研究人員要至少必須懂得這些一般原理。對統計方法略知一二，在於涉及到數值的任何形式實驗或觀察性研究都是不可少的；而對有一個變數以上的複雜實驗，則更是如此。

首先，初做研究的人必須懂得：在實驗部署階段就必須考慮到統計學。否則，實驗的結果可能沒有進行統計學處理的價值。因而，生物統計學不僅涉及實驗結果的解釋，而且也用於實驗的部署。人們通常認為：除純粹的統計方法外，生物統計學還把這些方法應用於實驗所涉及到的各種更為廣泛的問題，諸如實驗設計的一般原則以及有關的邏輯性問題等。費歇爾爵士對生物統計學方法的發展作了很大貢獻。他的著

作《實驗的設計》一書就討論了這些問題。

選擇對照組和試驗組時，首先要顧及邏輯和常識。舉例說，人們通常犯這樣的錯誤：加以比較的各組不是在同一時間進行的，即把今年獲得的資料同前幾年獲得的資料加以比較。這樣獲得的證據，雖然可以作出一些有益的啟示，但是決不可能證明什麼結果。『你拿小桶淘海水，如果趕上落潮，你和月亮都有大功勞。』在生物學研究中，可能有許多意想不到的因素，影響著處在不同時間、不同地點的群體。在考慮了一般因素以後，應用統計方法決定必需的組群的大小，根據重量、年齡等等選擇動物。並且，在考慮到這些因素的同時，將動物分入各組，而不失隨意選擇樣品的原則。

由於生物特有的變異性，從來沒有兩個動物組或植物組是完全相似的。即便費盡苦心保證兩組中所有的個體在性別、年齡、體重及品種等各方面近於相同，總還是會由於一些迄今尚未為人所知的因素所造成的差異。我們必須體認到，想要獲得完全相同的組群是不可能的。對付這一困難的方法是，估計這種變異性，並在評定實驗結果時加以考慮。在合理的範圍內，最好為實驗選擇彼此差異很小的動物，這個做法的目的是應可以增加實驗的敏感性。但也不必千方百計去做到這一點。因為我們亦可增加組群中個體數量等其它方法來達到這一目的。在一定情況下，可用數學技巧來校正組群之間的差異。

解決實驗用動物差異問題的另一個方法是『配對』，即把彼此酷似的動物一一對(如用一對孿生子或同胎動物)。把每一動物與其『對偶』加以比較。即可獲得一系列的實驗結果。採用一胎雙生的動物，在數量上常常是非常經濟的，當實驗對象是購買價格與飼養費用高昂的動物時，這一點就很重要。紐西蘭進行的乳脂產量實驗表明：每對同胎雙生母牛提供的資料不少於一百一十頭配對母牛所提供的資料。在生長率實驗方面，同胎雙生的小牛比普通小牛的用處要大二十五倍。

在第一次試驗某種實驗步驟時，常常很難事先估計需要多少動物才可以保證得到明確的結果。在使用價格昂貴動物的情況下，可先用少量動物進行試驗，然後重覆這個試驗，直至累積的結果足以滿足統計要求，這種做法也許比較經濟。統計學的一個基本概念是：被觀察的組群中的個體，是一個無限大的假想群體中的一個樣品。我們現在有專門的方法能隨意抽取樣品，並能估計樣品所需具有的廣泛性使其足以代表整體。我們所要求的樣品數量取決於物質的變異性，並取決於實驗結果所能容許的誤差大小，即所要求的精確度。

費歇爾認為：從前，過分強調了實驗時一次只能變化一個因素的重要性。他指出：部署實驗時，同時檢測幾個變數是有顯著好處的。採用適當的數學方法，能使一個實驗同時包括幾個變數。這種做法不僅能夠節省時間和精力，而且比分開處理每個變數能夠提供更多的資料。所以如此，是由於每個因素都從不同情況的角度受到考察，而且可以觀測到各因素間的任何相互作用。實驗上孤立地處理單一因素的傳統方法，常常意味著對該因素的限定有些主觀武斷，並且是在受到限制、過於簡化有條件下進行試驗。但是，動物實驗卻不如植物實驗那樣常用複雜的、多因素的實驗方法，雖然在做幾種不同成分飼料組合的營養試驗時，採用這種方法很有好處。

當然，同任何其它研究技巧一樣，統計學有其有用之處，也有它的局限性。因此，必須認識它在研究工作中所扮演的地位和作用。其主要價值在於對假設的檢驗，而不在著手進行新發現。新發現的獲得可能是由於考慮到了最細微的暗示，不同組群之間統計數字上最細小的差異，從而想到有可以深入研究的東西；而統計學則往往關係到事先精心安排實驗，用以檢驗一個已經形成的觀念。此外，研究人員決不能為了給統計分析提供足夠的數據，而在觀測及處理實驗細節方面時有所犧牲。

有一點是人們容易忘記的：在解釋實驗結果的時候，要用到常識，而利用統計學並不會減少這樣做的必要性。在處理兩組間存有顯著差異的現場實驗數據時，特別容易產生謬誤。差異並不一定是由已知道的因素所造成，因為可能在某一個實驗裏存有某種變數，其影響及重要程度還未被人認識而已。這不僅是學理上的可能性，從預防結核病、普通感冒和牛乳腺炎的許多接種實驗中出現的差異結果，都可作為例證說明這一點。在接種疫苗的同時，常伴隨有良好的保健措施和其它條件，這些都可能影響實驗的結果。從統計中可以看出：吸煙者的平均壽命不如非吸煙者的平均壽命長，不過，這並不一定說明吸煙就會縮短壽命。因為不吸煙的人可能在一些更重要的方面更加注意自己的健康。在精心設計的實驗中，由於最初是隨意抽取樣品的，因而保證了組群間能正確比較，這時，這樣的謬誤就不會產生。

統計學家對於供他分析的數據，在可靠性和準確度方面，容易估計過高。如果他同時又不是一個生物學家的話，尤其如此。所以，實驗人員應該聲明，測量僅僅是進行到厘米、克或其它某個單位的量級。統計學家能有一些生物實驗方面的親身體驗是對他有幫助的。而且，統計學家應該充分熟悉他所協同的實驗之各個層面。統計學家和生物學家之間的密切合作常使他們能夠運用受過教育的人所具有的常識，而不致陷入一大堆艱深的數據之中。

有時，由於實驗者將其實驗結果僅作為平均數提出，而致使科學報告受到損害。平均數往往提供不了多少資料，而且可能給人錯誤的印象。我們應該提供資料出現次數的分布，並且，與個體有關的數字常常有助於形成一幅完整的畫面。圖解有時也給人錯誤印象，繪製圖解所依據的數據須加以批判地審查。如果圖解上標繪的點不夠密集，也就是說觀測的次數不夠多，那麼，用直線或曲線把這些標繪點連結起來有時是缺乏根據的。這些線段也許並不代表真實位置，因為我們不知道兩點之間究竟出現了什麼情況，例如，很可能出現了意想不到的上升或下降。

### **給人錯誤印象的實驗** *Misleading experiments*

有關在研究工作中運用推理、假設和觀察可能會出現的危險將在本書適當章節加以討論。為了防止對實驗過分信賴的傾向，這裏也應提醒讀者：實驗有時也會給人非常錯誤的印象。出現差錯的最常見原因是技術上的錯誤。實驗人員必須對自己使用的技術極為熟練，否則就不能信賴實驗的結果。即使是出自專家之手的技術方法，也要經常對照已知的『肯定』、『否定』樣本來進行檢驗。除了技術上的疏忽外，還有更難以捉摸的原因使一些實驗『出毛病』。

亨特 (John Hunter) 故意讓自己傳染上淋病，以觀察淋病是不是一種與梅毒有顯著區別的疾病。但不幸他用以接種的物質同時含有梅毒菌，結果，兩種疾病他都傳染上了。從而長時間內形成了一種錯誤的概念：二者都是同一種疾病的表面症狀。尼達姆 (Needham) 用肉湯罐所作的實驗使他自己和別人都相信自然發生是可能的。當時，還沒有足夠的知識來證明這種謬誤究竟是由偶然的污染，還是由於加熱不足未能保證完全消毒而造成。近幾年，我們見到一次顯然是成功的實驗，證明『棒麴黴素 patulin』對於普通感冒有治療效果。統計學上的要求得到了滿足。但是，迄今為止從未有人證明得益於棒麴黴素，而第一次的實驗何以獲得成功至今仍是個謎。

當我看到防止羊群發作肉蠅病的所謂『米爾斯手術』時，我認識到它的重要性。米爾斯的發現所具有的極大的潛力大大地激發了我的想像力。我進行了一次數量達幾千頭羊的實驗，而且，不等結果出來，就勸說研究肉蠅問題的同事們也在別的地方進行實驗。大約一年以後得到結果，手術對我所試驗的羊群完全無效。其他人的試驗，以及所有以後進行的試驗都證明這種手術對羊群具有極有效的保護性。至於我的實驗為何失敗則找不出令人滿意的解釋。所幸的是：我當初對自己的判斷很有信心，所以勸說同事們也在國內其它地方進行試驗。因為，我當時假如更謹慎一些，假如等到結果出來再說，就有可能把這種手術的採用推遲許多年。

美國進行的幾次大規模實驗證明：免疫措施大大降低了一九四三年以及一九四五年流行性感冒的發病率。但在一九四七年，同樣類型的疫苗卻失敗了。後來發現，這次失敗是由於一九四七年的流感病毒菌，不同於前幾年流行的那種，也就是用來製造疫苗的那種病毒菌。

世界上不同地區的科學家，使用近似的生物體卻得出相反的結果，這決不是罕見的現象。這種現象有時可追溯到意想不到的因素，例如，實驗豚鼠對白喉毒素有極為不同的反應，就可追溯到豚鼠飼料的不同。有時，儘管進行了充分的調查研究，仍然無法找出差異的原因：在美國伊頓 (Monroe Eaton) 博士的實驗室裏能使流行性感冒病毒在老鼠中蔓延，而英國安德魯斯 (C.H. Andrewes) 博士的實驗室裏就做不到，即使他使用了同樣品種的老鼠和同樣病毒菌株、同樣的籠子以及完全相似的方法。

我們必須記住：嚴格地說，實驗結果有效，僅僅是對於實驗進行時所擁有的條件而言，在生物學上尤其如此。在必然會受到限制或改變的條件下，所得到的實驗結果究竟有多大的實用範圍，關於這一點，在作實驗結論的時候必須十分小心謹慎。

達爾文有一次半認真地說：『大自然是一有機會就要說謊的。』班克羅夫特 (Wilder Dwight Bancroft) 指出：所有的科學家由親身經歷都知道，使實驗得出正確的結果常常是多麼困難，即使在知道該怎麼做的時候也是如此。因此他說，對於旨在得到資料的實驗，不應過分信任。

上面援引的例子都是一些所得到的結果實際上是『錯誤』的，或是給人錯誤印象的實驗。幸好，這都是個別的例子。然而，更常見的是由於不知道確切的、必要的條件，致使實驗不能證明什麼問題。例如法拉第早期試圖以磁鐵獲得電流實驗就一再失敗。這樣的實驗表明，如眾所周知，證明一再失敗的命題是非常困難的。而且，對

於從這樣的實驗中得出正確結論的愚蠢做法，科學家們往往是能識別的。據說，有些研究機構故意銷毀『一再失敗的實驗』記錄；此外，不予發表那些未能證實所要檢驗的假設之研究結果，是一種很可取的做法。

### **提要** *Summary*

大多數生物實驗的基礎是對照實驗。進行對照實驗時，以隨意抽取樣品的方式將個體編入組群，這些組群除去需要進行研究處理的因素之外，其它各方面都應相同，並考慮到生物體特有的變異性。先進行整體試驗，後進行分部試驗；並按步驟排除各種不影響實驗結果的因素，這是兩項有用的原則。在進行實驗時，密切注意細節，作出詳細的筆記以及客觀解釋實驗結果，都是很重要的。

生物統計學不僅用於實驗結果的解釋，而且用於實驗的部署。生物統計學的一個基本概念是：有一個無限大的假想群體，實驗組群或數據是從中隨意抽取的樣品、應通過對變異性的估計，並在評定實驗結果時將其考慮在內，以克服由於生物體特有的變異性所造成的困難。

如同研究工作所使用的其它方法一樣，實驗並不是萬無一失的。從實驗上不能證明一種假設，並不等於證明這種假設是不正確的。