



不難發現在科學研究的報告中，
針對相同的研究議題，
不同的研究者，
獲得的結論卻不一致，
或同一研究者前後所獲得的
結果常不一致。
再者許多研究的假說，
經過一段時間的熱烈追尋後，
就從文獻中漸漸消失了。

王新基／公衛所
劉明毅／環醫所

書評

大部分發表的研究結果其結論的真實性都需要被存疑



今年諾貝爾生醫獎得主Dr. Marshall，當年還是實習醫師時，就提出

胃潰瘍是由幽門桿菌所引起，而不是當時被學術界視為顯學的胃酸所致，當年

Dr. Marshall的文章在澳洲根本無法被發表，然後來結果證明Dr. Marshall是對的。事實上，許多研究者當面對研究結果與主流的假說不一樣時，而不敢發表自己的研究成果，或者投稿出去常常被期刊拒絕。但是，根據歷史的經驗法則，研究者應該對自己的研究發現要有信心。面臨別人的質疑或冷落不需要氣餒，因為有可能你的發現才是正確的。

基本上，目前科學研究發表的主流，是選擇發表統計上具有顯著意義的

發現，更甚者研究期刊論文審查者，常會拒絕接受與自己研究結果相反的發現；但是，這種科學的慣性操作常會延緩研究的進步。許多改變人類歷史的偉大發現，當是歷盡千辛萬苦才被認定，主流的認定價值很不幸的常淪落為阻礙人類文明進步的頑石。

針對目前持山倒海似的科學文章中，到底其研究結果的真(true)或偽(false)發生的情形，長久以來有諸多的討論，另外，對於科學研究是否帶給人類了解更多事實真相的契機，或只是終其一生白忙了一場，不但被別人誤導了也誤導了別人。這些年來Dr. Ioannidis針對「發表的研究結果其結論的真與偽」這個話題，發表了許多令人省思的論文。本次書評我們就來探討篇名是「Why most published research findings are false」這篇文章。

Dr. Ioannidis列舉了臨床、基礎、基因體、流行病學等相關的研究，發現在Published Research Findings中，真正可採信的研究結果，機率超過百分之五十的只有Phase III randomized control trials and meta analysis，而流行病學與基因體相關研究結果真實性則遠低於50% (10-20% for epidemiological studies... <1/1000 for genomic research findings)。作者最重要的貢獻是提出分析架構，並用此架構評估各研究領域研究發現的可信度，並且根據各研究領域研究的結果做相互之間的比較和排序。作者用簡單醫學診斷工具評估的概念，來說明他得到的結果。當一個被診斷者，如果具有診斷工具所評估的medical condition，則他的true state定義為positive；如果沒有此medical condition，則他的true state定義為negative。診斷的結果(test result)，根據

診斷結果，被診斷者如果被診斷出有此 medical condition，則 test result 定義為 positive；不然 test result 定義為 negative。作者用類比的方式，將研究假說是真實的情況當作 true state is positive，研究假說不是真實的情況當作 true state is negative。將收集資料後統計分析的結果進一步顯示，研究假說成立的情況視為 positive test results，研究假說不成立的情況視為 negative test results。

診斷工具的優劣取決於敏感度與特異度，敏感度指的是在 true state is positive 的個案中診斷結果是 positive 的比例，特異度指的是在 true state is negative 的個案中診斷結果是 negative 的比例。高敏感度與高特異度代表高準確度的診斷工具，用統計的方式檢定研究假說是否成立時，我們同樣的希望所用的統計方法具有高敏感度及特異度。以假設檢定術語來說就是檢定力 (statistical power) 要高及型 I 誤差機率 (probability of type I error) 要低。通常的要求是型 I 誤差機率要小於 0.05，統計檢定力則要超過 80%，研究的樣本數基本上要達到上述的標準。診斷工具的陽性預測值，指的是診斷結果是陽性的個案中實際情況也是陽性的比例；這統計顯著意義的研究假說，其真實性的機率就是此研究結果

的陽性預測值。陽性預測的高低與診斷的 medical condition 的盛行率 (prevalence)，敏感度及特異度高低有關，盛行率越高或診斷工具的準確度越高，陽性預測值就越高；也就是說，研究結果的陽性預測值與研究假說，在收集資料前真實性的比例，以及收集資料後統計分析所用方法的統計檢定力及型 I 誤差的機率大小有關。

根據作者的分析，陽性預測值超過百分之五十的研究，發現只有樣本數較大的 phase III randomized control trials 和 meta analysis，能夠進入 phase III 人體試驗的介入措施，由於已完 phase I and II trials，所提出的介入措施有效的機率很高，再加上在實驗過程控制產生偏差的因素，因此研究結果的真實性也就很高。所謂 meta analysis 則是針對某個特定研究假說，彙整相關研究的數據加以驗證，由於彙整後樣本數增加的结果，提高統計檢定力，增強研究結果的真實性。至於基因體相關的研究，通常與研究假說可能有關係的基因數量很多，但是其中與假說真正有關的基因可能只有個位數。當提出某些基因與研究假說有關，真實性的機率可能很低，再加上樣本數較小，因此陽性預測值會很低。作者認為每一千篇發表的基因研究成果，可採信的不足一篇。流行病學相關研

究，通常是探討危險或保護因子與疾病或死亡之間關係的研究，由於是觀察型的研究，存在許多影響結果的偏差，譬如選擇偏差、干擾因素等，發表的結果危險或保護因子與疾病或死亡之間相關程度通常不大，雖然研究樣本數通常不小，其研究結果的真實性亦遠低於百分之五十。

在診斷疾病時，通常會使用不只一個診斷工具，如果只要其中一項診斷結果是陽性，則視為陽性個案的話，則敏感度會增加，但是偽陽性 (false positive rate) 也隨之上升，而陽性預測值在一般情況下 (false positive rate < 0.05; true positive rate > 0.8) 會下降。作者用此概念討論熱門研究領域，因競爭性對陽性預測值的影響，針對某個熱門的研究假說，文獻報告如果出現許多迷離著意義的結果，我們要小心這些結果的真實性可能不高，因為研究者可能為取得研究經費，想盡辦法強化其研究成果。事實上，不難在文獻裏發現有些研究假說在熱烈追逐下，最後聲聲匿跡了。

診斷上的偏誤 (bias)，會造成錯誤的診斷，陽性預測值可能因此下降。除了研究設計及測量上造成的偏誤，研究者選擇性的發表研究完成結果，或者論文及計畫審查者個人學術上的偏見也可造成偏誤。審查論文者，對於與自己不同的研究結果，有可能會採取壓抑策略，這

或發表的結果偏袒一方，其真實性值得存疑。

作者根據他觀察及分析的結果，提出以下的結論：

- 一、樣本數越小，發表的研究結果真實性越低。(The smaller the studies conducted in a scientific field, the less likely the research findings are to be true.)
- 二、分析出來的差異值越小，發表的研究結果真實性越低。(The smaller the effect sizes in a scientific field, the less likely the research findings are to be true.)
- 三、檢驗假說越多且沒有精心過濾，發表的研究結果真實性越低。(The

greater the number and the lesser the selection of tested relationships in a scientific field, the less likely the research findings are to be true.)

- 四、研究設計、變項定義、結果評估、分析方式可塑性越大，發表的研究結果真實性越低。(The greater the flexibility in designs, definitions, outcomes, and analytic modes in a scientific field, the less likely the research findings are to be true.)
- 五、有過多贊助單位或其它利益考量，發表的研究結果真實性越低。(The greater the financial and other interests and prejudices in a scientific field, the less likely the research findings are to

be true.)

- 六、研究領域競爭越激烈，發表的研究結果真實性越低。(The better a scientific field with more scientific teams involved, the less likely the research findings are to be true.)

總之，綜觀Dr. Ioannidis的分析結果，在追求科學真理時為提高研究的真實性，我們應該強化下列的目標：

- (一) 研究的樣本數要足夠；
- (二) 研究設計、變項測量要求的標準要提高；
- (三) 研究假說要深思熟慮，不要在一個研究裏驗證過多的假說；
- (四) 研究要避免受到非學術因素之影響。

後記

多年來我倆常常在一起討論，到底我們在科學領域中忙的不可開交，到終了時會不會只是白忙一場，對人類在科學的建樹上是零，這是多麼令人恐慌的結局。之前我倆只是站在以互勉的方式，激勵彼此要能完成堪稱有建樹的研究工作，當讀到這篇文章看到作者不但提出分析架構，更把我倆多年來的疑惑給了數據上的剖析；從開始閱讀此文之日開始，乘著是隔壁鄰居之便，在下班後多次討論，文章修改再修改，就如同在共同執行一個研究計劃，如今將文章呈現給成大醫訊的讀者，期望能得到大家熱烈的回應。



參考文獻

- Ioannidis JPA. (2005) Why most published research findings are false. PLoS Medicine 8:596-701.