

Profiles, not Metrics

Jonathan Adams, Marie McVeigh,
David Pendlebury 及 Martin Szomszor 合著

2019 年 1 月

Web of Science

 **Clarivate**
Analytics

科睿唯安

作者簡介

Jonathan Adams 教授是科睿唯安旗下 *Institute for Scientific Information (ISI)* 的負責人。他是倫敦國王學院 (King's College London) 政策研究所的客座教授。由於在高等教育和政策研究領域的卓越貢獻，Jonathan Adams 在 2017 年被艾希特大學 (University of Exeter) 授予榮譽理學博士學位。

Marie McVeigh 是 *ISI* 編輯團隊的編輯倫理主管。她最初是賓州大學 (University of Pennsylvania) 的細胞生物學家。自 1994 年以來，Marie McVeigh 在 *ISI* 及其前身機構任職，致力於期刊管理和文獻資訊工作並發表了一系列相關文章。她近期負責的 *JCR* 的提升工作，增加了文章層面的具體資訊和資料透明度，以支援正確使用期刊引用指標。

David Pendlebury 是 *ISI* 的研究分析主管。自 1983 年以來，他一直致力於使用 *Web of Science* 資料來探尋科學研究的結構和動態。他與 *ISI* 創辦人 Eugene Garfield 共事多年，並與 Henry Small 共同開發了 *Essential Science Indicators* 資料庫。

Martin Szomszor 博士是 *ISI* 研究分析主管。他曾是資料科學 (Data Science) 負責人，以及 *Global Research Identifier Database* 創始人，他將機器學習、資料整合和視覺化技術領域的廣泛知識應用到相關工作中。他因與英格蘭高等教育資助委員會 (Higher Education Funding Council for England) 合作建立了「REF2015 具有影響力案例研究資料庫」(REF2015 Impact Case Studies Database)，榮膺「2015 年英國資訊時代 50 強資料領袖」(2015 top-50 UK Information Age data leader) 稱號。

關於 Institute for Scientific Information (ISI)

ISI 是科睿唯安旗下 *Web of Science* 事業群的附屬「研究院」，維護著構建 *Web of Science* 及相關數據、分析、產品和服務的知識庫。*ISI* 積極參加外部會議與發表相關報告以進行知識傳遞，亦致力透過研究分析讓知識庫可以持續擴展和更臻完善。

關於 Web of Science

科睿唯安 *Web of Science* 事業群提供全球一流的科研資訊和數據，協助推動學術界、企業、出版機構和政府加快研究步伐。其旗下擁有全球最大、最值得信賴、中立於出版機構的引文索引資料庫及獨立的研究資訊平台—*Web of Science*，以及 *EndNote*、*Converis*、*ScholarOne*、*Publons*、*Kopernio* 和 *ISI* 等諸多知名品牌。

更多資訊請瀏覽：clarivate.com.tw/products/web-of-science

本報告強調：當有關科研人員及其機構的資料被壓縮為簡單的指標或排名時，部分資訊將會因此而遺漏。本報告闡述了四類常見分析，如若誤用將掩蓋真實的科研表現；我們提出了四種視覺化選項，用於解讀每個度量指標下蘊含的更豐富的資訊，以及支援開展全面的、負責任的科研管理。

我們身邊依然存在著聲稱可通過簡單分析來評估論文、科研人員和機構表現的現象。儘管資深分析專家提出反對意見，諸多研究人員對此深表憂慮，但是大學管理人員依然樂此不疲。有關大學排名可信度的爭議無休無止，但大學排名並不會因此而停止發佈。我們不禁要問：諸如單一指標和線性排名之類的簡單分析為何如此受到歡迎？

摘要性統計資料（summary statistics）和排名表（league tables）擁有與生俱來的魅力。我們希望「看看誰做得最好」，就像觀看體育比賽一樣。但是，聯賽積分表是既定成員之間通過進行一系列比賽而衍生的產物，誰在公開比賽中的綜合成績更好，誰的近期排名就越靠前。聯賽積分表基於單方面的比賽名次，能夠一目了然地顯示出選手通過博弈而取得的排名。

但科學研究並非單方面事物：其過程很複雜，沒有兩個項目是完全一致的。研究機構也並非只有一項任務：他們既要教學，又要研究；他們的研究內容可能是基礎性、分析性、應用性、協作性、社會性或行業性的；他們的研究活動可能跨越多個學科，每個學科都有各自的學術特點。

單一指標在某些類型的比較中是具有價值的，例如各大學類似的院系中每名研究人員的相對產出，我們從中可以瞭解到「類似」研究中的真正差異。但是，如果用單一指標替代全面的科研管理，例如在缺乏補充資訊的情況下進行學術評估，甚至將單一指標視為招聘標準，那麼，這類資訊就具有一定的局限性，而且單個（或孤立）指標可能會被誤用。

通過使用分佈在科研活動和學科領域中的替代性指標，大學排名採用一系列變數來「描繪」一所大學。每個變數都被標引：與計數、資金、影響力、時間及其他不相容的項目相掛鉤而進行量化；然後再通過加權將不同的專案組合在一起，從而得出最終分數。如果沒有合理的資料管理機制，該分數將遠遠不能體現大學生命的豐富性與多樣性。

每一個被過度簡化或誤用的指標，其實都有更好的替代選項。通常是先進行合適的、負責任的資料分析，再以圖形方式來顯示多個互補的維度。當我們把資料置於更廣泛的背景中，將能夠看到新的特點，瞭解更多的資訊。下面的範例將充分顯示這個做法是多麼簡單易行，而且能大幅提高我們解讀科研活動的能力。



個體：h-index 與 beam-plot

物理學家 Jorge Hirsch 於 2005 年建立了 h-index，h-index 是一種被廣泛應用但卻不一定能夠全面反映科研人員論文和引文影響力的指標。它將一系列論文及其被引次數縮減為單個數字：一個具有指數 h 的研究人員（或團隊，甚至是國家）至少已發表了 h 篇論文，並且每篇論文至少已被引用了 h 次。

h-index 取決於職業生涯的長度和學科。因為隨著時間的推移，不同研究領域之間的論文，其被引次數的積累速度各不相同。因此，h-index 不適用於對個體進行比較；它通常不涵蓋非期刊出版物；從數學的角度看，也不具有一致性（Waltman 和 van Eck，2012）。

來自德國馬普學會（Max Planck Institute）的 Lutz Bornmann 和 Robin Haunschild 提出的另一種方法（Bornmann 和 Haunschild，2018）：是將研究人員的文章放在適合進行比較的背景下。每篇論文的被引次數均按與其具有相同學科和出版年份的期刊的平均值進行「正規化」（normalized），並將該值轉換為百分位數。相對單純的平均值而言，這種方法因為引用分佈偏斜度很大，從而能更準確地衡量集中趨勢。百分位數為 90 意味著該論文位於引用率最高的前 10% 之列，另外 90% 則是引文影響力較低的論文。中位數為 50：論文的篇均影響力範圍基本在 0 到 100 之間。

Beam-plot 可用於展開公平的、有意義的評估。它可迅速傳達 h-index 永遠不能傳達的資訊。如圖所示，該位研究人員的平均百分位數明顯高於 50，表明其發表的論文在其所屬領域具有核心影響力。雖然論文的年度中位影響力在早年低於 50，但我們可以看出，這一數值隨著時間的推移逐漸超過了平均水準。

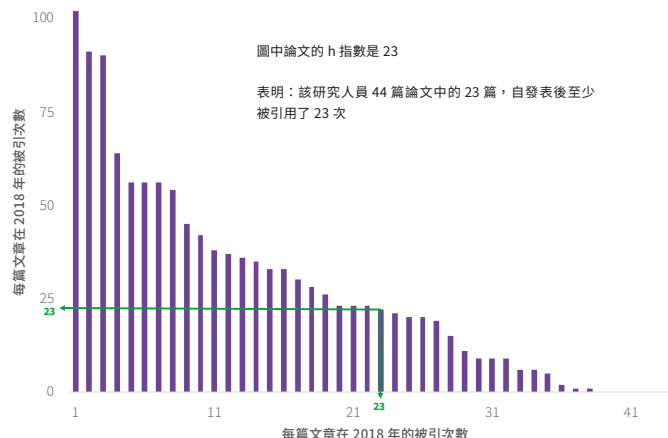


圖 1. 在此示例中，h-index = 23 的科研人員在 15 年間共發表了 44 篇可被引用的期刊文獻（單獨編著或與其他人合著），其中包括不能以這種方式進行分析的報告和會議論文。通過對資料製圖，我們可以發現隱藏在 "h" 值下的相對高被引文章的分佈、偏斜和存在情況，及消失的未被引用的文獻。

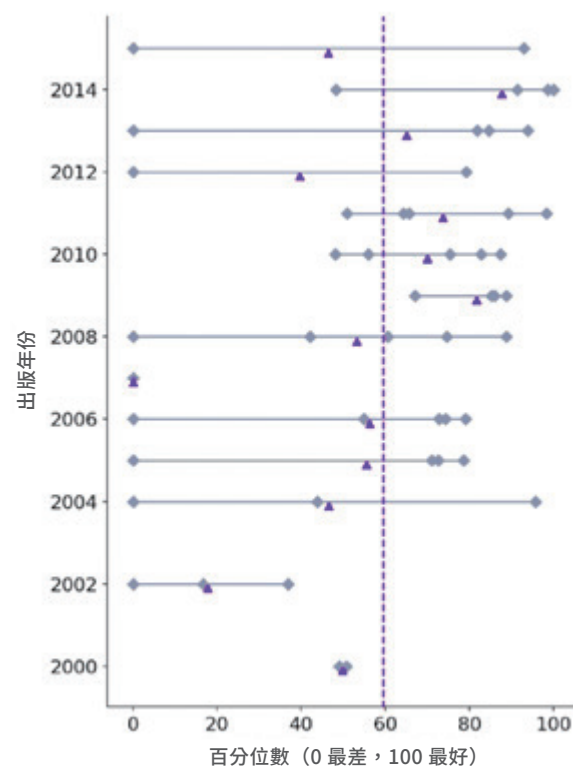


圖 2. 圖 1 中資料的 beam-plot，將每篇文章與相應年度與學科的參考資料集進行比較，都使用 0-100 百分位數。該圖顯示了每年的百分位元數範圍（射束灰色標記）及其年度中位數（樞軸上的紫色標記）。基準線是該研究人員的總體平均值：百分位 59。

期刊：期刊影響因數（JIF）與期刊引證報告（JCR）的對比

量化科研評估通常會關注出版物的集合，並將平均被引次數與基於學科的基準進行比較。評估人員還可以關注發表該文章的期刊。

Journal Impact Factor (JIF) 是我們所熟悉的常用指標。JIF 由 *ISI* 創辦人 Eugene Garfield 建立。Garfield 於 1955 年提出了出版物「影響力」的概念並創建了「期刊」影響因子 (Garfield and Sher, 1962)，旨在幫助全新的 *Science Citation Index (SCI)* 選擇新期刊。第一份 *JCR* 隨後於 1975 年正式出版。

JIF2（即基於兩年的期刊資料）有兩個基本要素：分子和分母。分子是指前兩年在期刊上所發表的任何論文在當前年份中的被引次數；分母是指這兩年所發表的實質性的科研論文 (article) 及綜述 (review) 的數量。這兩個基本要素可通過調整來適應更短或更長的時間間隔。僅以前一年論文情況為考量的 *JIF* 可凸顯快速變化的領域；而以前面 5 年或 10 年的論文情況（數量及被引次數）為考量的 *JIF* 則更能彰顯某個特定年份的期刊被引用情況。

預算有限且需管理諸多圖書訂閱工作的圖書館員通常使用量化期刊比較工具，出版商則使用該工具來跟蹤系列出版物的績效表現。問題在於，旨在科學管理期刊的 *JIF*，卻被不負責任地應用於更廣泛的研究管理之中。

為解決這一問題，2018 年的 *JCR* 通過更為豐富的資料背景修訂了期刊略述。例如：以同一學科中百分位數來表達 *JIF* 值的橫條圖將能夠快速顯示其分區；此外，對研究人員至關重要的是，每篇文獻對引文的貢獻都將在覆蓋所有文章的引用頻率分佈圖中進行顯示。

新的期刊略述清晰地表明，*JIF* 將更大、更複雜的資料池匯總在一起。*JIF* 對期刊經理而言可能是個很有用的工具。但對研究管理人員而言，*JIF* 只能提供他們需要的關於期刊或論文價值的部分資訊。



圖 3。左圖：EMBOReports 的期刊影響因數趨勢圖顯示了該期刊的影響因數變化趨勢及其在相應學科全部期刊中的影響因子排序百分位變化趨勢。右圖：2017 年引文分佈圖顯示了中位數和整體分佈情況。

研究機構：篇均引文影響力與影響力全貌 Impact Profile™

當我們將考察標的從個人和期刊轉移到研究團隊和機構時，將科研活動的資料通過單一指標進行分析所產生的局限性變得更加明顯。

我們對主要從事生物醫學研究的兩家研究機構的學術出版物進行了考察。雖然他們的具體研究任務並不相同，但其總體研究方向大致相似；由類似的科研資助機構提供支援，並擁有相似的歷史背景。為了幫助考察小組對這兩個機構開展分析，我們匯總了其最近 5 年的學術論文出版記錄，發現他們一共在期刊上發表了約 1,250 篇論文：A 機構 845 篇，B 機構 403 篇。

由於被引次數的長期累積速度因學科而異，因此，這些單位的論文被引次數需要針對相應學科和出版年份的世界平均值進行「正規化」，從而得出「學科正規化引文影響力」值（CNCI — 也稱為「基準重構影響力」Re-Based Impact RBI）。A 機構的平均 CNCI 為 1.86，B 單位為 2.55（與世界平均值 1.0 相比）。

雖然這些 CNCI 值不具有統計功效，但在管理實踐中我們通常可以假設：體量越小「影響」越大。然而，某一機構 CNCI 的平均值是由數百篇文章各自的 CNCI 加權平均而來。因此，平均影響力指標可能具有一定的欺騙性。

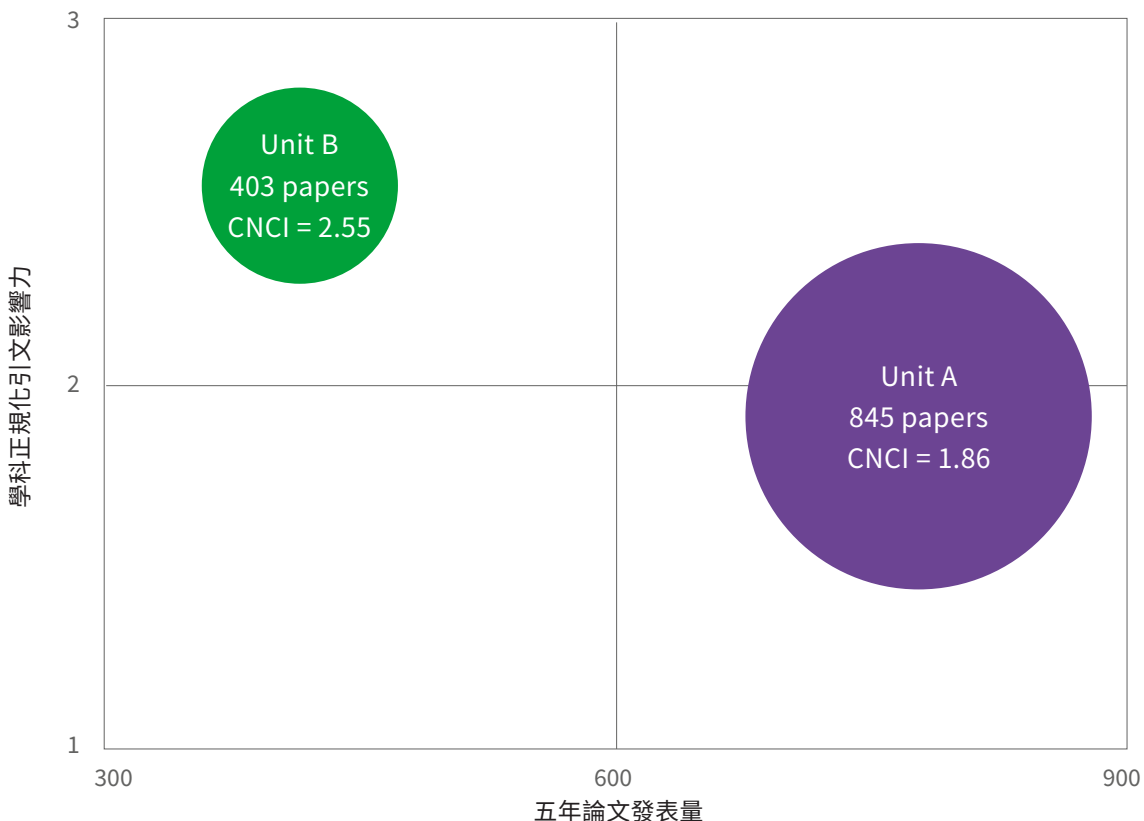


圖 4. 兩家英國生物醫學研究機構的五年論文發表量和學科規範化引文影響力（CNCI）。B 機構的論文發表量約為 A 機構的一半，但規範化引文影響力卻遠高於 A 機構。

被引次數很容易出現偏斜分佈，幾乎任何樣本都有很多低值和幾個高值。因此，為了使論文被引情況視覺化，我們以世界平均值為基準，對被引次數進行分類：首先，對於超過世界平均值的文章，我們分為四個區間，報告引文影響力為世界平均值 1 到 2 倍，2-4 倍、4-8 倍和 8 倍以上。

同時，我們針對引文影響力為世界平均值 1.0 到 $\frac{1}{2}$ 倍， $\frac{1}{2}$ 到 $\frac{1}{4}$ 倍，依此類推，又得出影響範圍低於世界平均值的四個區間。我們將未被引用的論文單獨放在第九個區間中，揭示出每個資料集的整体「影響力全貌」(Impact Profile™)，顯示那些高被引和低被引論文的真實分佈情況 (Adams、Gurney and Marshall, 2007)。

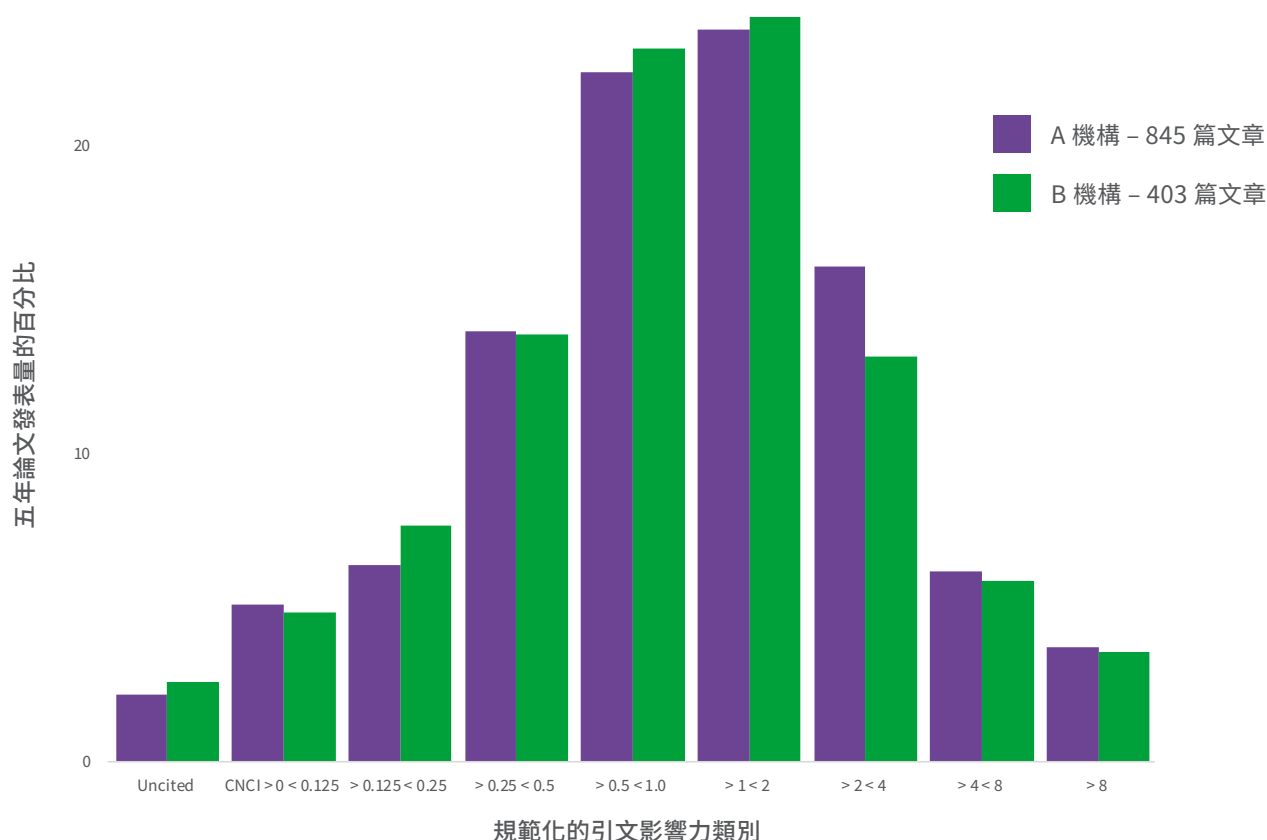


圖 5. 兩家英國生物醫學研究機構的五年影響力全貌 (Impact Profile™)。每篇論文的被引用次數均按該篇論文出版年份和期刊所屬學科 (CNCI：見文本) 基於世界平均值進行了「正規化」處理，並分配了圍繞著該平均水準進行分組的一系列區間 (世界平均值 = 1.0；未被引用的論文居於最左)。計數顯示為每家機構的發文量百分比。

這種方法所生成的圖像為我們提供了遠比圖 4 總平均值更加豐富的資訊。這個畫像看起來更像是分佈在世界平均值兩側的正態 (高斯) 曲線。我們可以「定位」每個機構的整体平均值，並檢查其科研產出高出或低於該平均值的比例：對於兩家機構來說，低於機構平均 CNCI 的論文占比更高。

最重要的是，我們可以立即發現這兩個機構的影響力全貌 (Impact Profile™) 並沒有實質性差異，都客觀地呈現出兩家機構的研究表現。事實上，通過回顧原始資料，我們發現 B 機構之所以擁有較高的平均影響力，主要是因為其有一篇綜述文章，被一家權威性期刊高頻率引用。

大學：排名與多因素研究足跡

諸如「篇均引文影響力」之類的總結性指標僅能精簡一類資料。即使如此，與影響力全貌 (Impact Profile™) 中的圖片相比，科研管理人員在使用平均值時仍會錯失大量資訊。

當排名列表將關於不同大學活動的多類別資料彙集成一個數字排行榜時，資訊遺漏問題將變得更加嚴重，並且有用的內容將更易錯失。

對於我們所熟知的年度排名體系，如泰晤士高等教育 (THE) 的世界大學排名，資料收集相對廣泛，演算法容易理解並且經過了平衡處理。此外，排名還考慮了機構規模等因素，以避免榜單被最大規模的老牌大學所統治。儘管如此，大多數讀者仍「知道」哪些大學會排在前列。

在這些大學求學或任教的學者們都知道該校科研活動的複雜性和多樣性，及學校產生的成果。通過對以下兩所英國倫敦的著名大學進行比較，我們便可揭示這種差異。他們是擁有以理工傳統的帝國理工學院 (Imperial College, London)，以及在社會科學領域享譽全球的倫敦政治經濟學院 (London School of Economics)。

通常情況下，沒有人會對這兩所大學進行直接比較，更不用說將其歸於一類進行排名了。

即使是對同樣擁有大型醫學院的老牌知名綜合大學進行比較，例如愛丁堡大學 (第 29 名) 和曼徹斯特大學 (第 57 名)，我們也無法解讀排名結果。這些排名位次究竟意味著什麼？

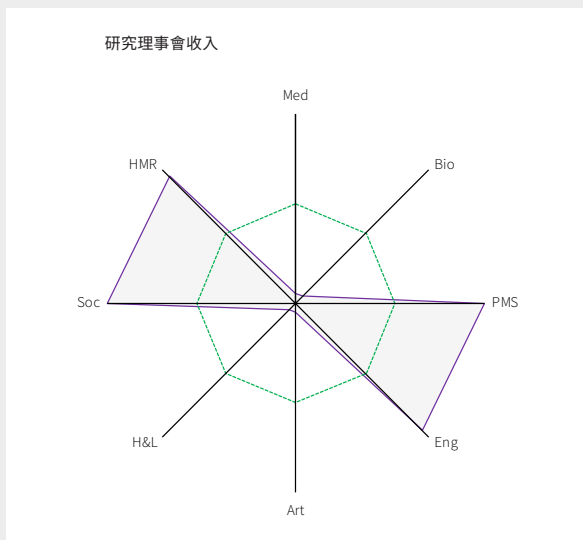
實際上，任何大學都有高分項和低分項，導致其相對位置不斷變化。全球大學排名可能很引人矚目，但只是一個參考點。即便是針對學生的、經過精心篩選的排行榜，也會隱藏太多細節，更不用說作為管理層的決策工具了。

Evidence Ltd 於 2002 至 2014 年間在「英國高等教育研究年鑒」(UK Higher Education Research Yearbooks) 中使用的「研究足跡」(Research Footprint) 更具資訊性。研究足跡是一個雷達圖，它使用多個坐標軸來顯示多個指標，以便將每個實體與一個基準參考研究足跡進行比較，或者由一系列的機構「足跡」共用一幅概述。

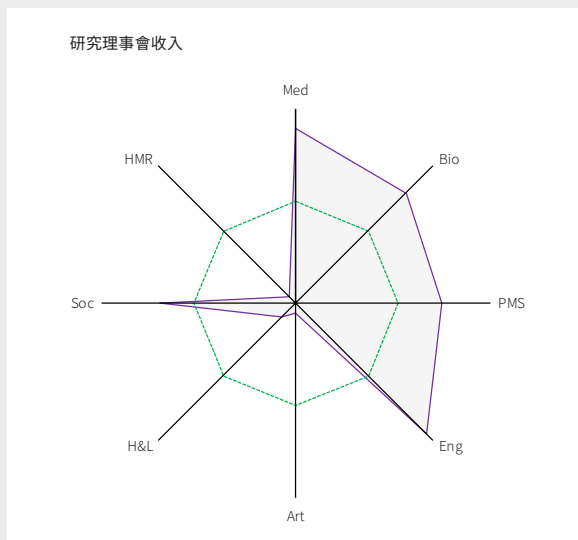
全球大學	WUR 位次		英國大學
University of Oxford	1	1	University of Oxford
University of Cambridge	2	2	University of Cambridge
Stanford University	3	9	Imperial College London
MIT	4	14	University College London
CalTech	5	26	London School of Economics
Harvard University	6	29	University of Edinburgh
Princeton University	7	38	King's College London
Yale University	8	57	University of Manchester
Imperial College London	9	78	University of Bristol
University of Chicago	10	79	University of Warwick

表 1. 在 2018 年 *Times Higher Education* 世界大學排名 (WUR) 中排名最高的全球大學位次

帝國理工學院



倫敦政治經濟學院



上圖：英國研究理事會基金對兩個機構八個學科領域的資助情況：醫學 (Med)；生物學 (Bio)；物理與數學科學 (PMS)；工程 (Eng)；藝術與設計 (Art)；人文與語言 (H&L)；社會科學 (Soc)；以及健康和醫療相關學科 (HMR)。研究足跡顯示出這些機構如何不同，但仍可進行比較。

下圖：五家著名的生物醫學實驗室在六個 *Web of Science* 學科中所發表出版物的學科正規化引文影響力 (CNCI)：



圖 6. 兩家英國高等教育機構（上排）的研究足跡顯示出他們在主要學科的基金資助情況（類似圖表可用於論文量、學生和工作人員人數，或引文影響力）與來自相應對比組參考基準的對標情況。下排的研究足跡比較了幾家著名生物醫學實驗室在其活躍的特定學科中的學科正規化引文影響力：這種情況下無需基準對標。

結論

本報告中討論的單一指標（h-index、JIF、篇均引文影響力）和大學排名都能提供一定資訊，但都存在被普遍誤解、不可靠，並且經常被誤用的現象。我們可通過研究活動的「圖片概述」（picture profile）這種直觀的分析方法來替代簡單指標。作為圖解說明方法，「圖片概述」具有以下特點：製作相對簡單；可釋放更多有價值的資訊；並支援適當的、負責任的科研管理。

beam-plot，而不是 h-index

beam-plot 是顯示研究人員發文量和影響力的單一「概述」，可顯示研究人員的發文量和影響力在一年內如何變化，以及如何隨著時間而改變。beam-plot 使用百分位數，意味著引文影響力是高度偏斜分佈的，可以認為是同時考慮了論文所屬學科以及論文發表時間。若只使用單一值的 h-index，我們僅能瞭解概況，但無法正確展開科研評估工作。

期刊概述頁面（A Journal Profile Page），而不僅是 Journal Impact Factor（JIF）

JIF 經常被誤用。它所關乎的並不是科研評估，而是期刊管理。若 JIF 所在的使用場景是將單點值設置為概述或更為廣泛的科研活動，那麼研究人員和管理人員將能夠看到 JIF 可在文章層面提供極為多樣性的績效資訊。JIF 可能是一個指南，但是如果我們要想獲取真實資訊，就需要先獲得超越圖書館和出版機構的完整背景。

影響力全貌（Impact Profile™），而非孤立的學科正規化引文影響力（CNCI）

學科正規化引文影響力（CNCI）的平均值也可能具有誤導性，因為它會淹沒多樣化的資料傳播，導致某篇論文或某個期刊會出現高度偏斜，並且很容易受到異常值的影響。影響力全貌（Impact Profile™）可將該偏斜轉變成更容易理解的形式，並揭示潛在分佈情況。通過顯示世界平均值和機構平均值的分佈，它可揭示出論文的相對被引情況（超出還是低於）。匯總值只能告訴我們 X 的平均值高於 Y，但影響力全貌（Impact Profile™）不僅能夠提出一系列問題，而且還能指引我們找到科研管理相關問題的答案：在哪有合作論文；同一個人是否同時發表了高被引和低被引論文；以及我們是否隨著時間的推移而進步等等。

研究足跡（Research Footprint），而非大學排名

大學排名所壓縮的資訊量遠遠超過了絕大多數的分析指標。研究足跡可按學科或資料類型顯示績效。它可以用於比較兩個機構或國家，也可將一系列的目標群組織與相應的基準進行對標。至關重要的是，它證明了：僅僅通過一個數值，便能對兩個複雜研究體系進行比較的方法根本不存在 實際情況要更加複雜！

西方諺語說的好，一張圖片勝過千言萬語。一張資料分佈視覺化圖形勝過一千個單一指標。

“ 一張圖片勝過千言
萬語。一張數據分
布視覺化圖形勝過一千個
單一指標。 ”

參考文獻 (References)

Adams J, Gurney K A and Marshall S. (2007). Profiling citation impact: a new methodology. *Scientometrics*, **72**, 325-344

Bornmann, L and Haunschild, R. (2018). Plots for visualizing paper impact and journal impact of single researchers in a single graph. *Scientometrics*, **115**, 385-394. DOI <https://doi.org/10.1007/s11192-018-2658-1>

Bornmann, L and Marx, W. (2014). Distributions instead of single numbers: percentiles and beam plots for the assessment of single researchers. *Journal of the Association for Information Science and Technology* **65**, 206–208. DOI: 10.1002/asi

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *PNAS*, **102**, 16569–72.

Garfield, E. (1955). Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas. *Science*, **122**, 108-111.

Garfield, E. (2006). The History and Meaning of the *Journal Impact Factor*. *Journal of the American Medical Association (JAMA)*, **293**: 90-93, January 2006.

Garfield, E., & Sher, I. H. (1963). New factors in the evaluation of scientific literature through citation indexing. *American Documentation*, **14**, 195-201

Waltman, L., & Van Eck, N. J. (2012). The inconsistency of the h-index. *Journal of the American Society for Information Science and Technology*, **63**(2), 406-415.

Wang, J. (2013). Citation time window choice for research impact evaluation. *Scientometrics*, **94**, 851–872. doi: [10.1007/s11192-012-0775-9](https://doi.org/10.1007/s11192-012-0775-9)

Web of Science



科睿唯安