

ほらいずん

研究計量に関するライデン声明について

科学技術・学術基盤調査研究室 客員研究官 小野寺 夏生、室長 伊神 正貴

【概要】

論文の被引用数等の計量データは、適切に利用されれば専門家（ピア）による評定をより妥当、公正にするための補完となり得るが、データに主導された評価や、指標の意味・性質の不十分な理解による誤用がしばしば見られる。このような状況に対して、研究評価における計量データの利用についてのベストプラクティスを示した「研究計量に関するライデン声明」（“The Leiden Manifesto for research metrics”）が、2015年にNature誌上で公表された。本稿では、このマニフェストの成立の経緯を述べた後、10項目の原則の全訳を紹介する。このマニフェストは、研究者、管理者、評価者の全てにとって、計量データに立脚した研究評価のガイドラインとなるものと考えられる。

1. はじめに

ICT技術の発展とネットワーク環境の進展により、多くの研究評価において、論文の被引用数等の計量データ利用が加速している。Web of Science Core Collection（以下Web of Scienceという）やScopusなど引用索引を備えたデータベース、そして、それらのデータ群を組み合わせたInCitesやSciValなどの情報分析ツールの発展により、かつては専門家が手間をかけて行っていた計量データの利用は比較的容易になった。一方、ツールやデータベースの利用が適切であれば、計量データは、専門家（ピア）による評定をより妥当、公正にするための補完となり得る¹⁾。しかしながら、現実には、データを補完材料として利用するのではなく、データに主導され、引きずられた評価が往々にして行われている。データや、それに基づいて計算される種々の計量的指標の意味・性質が十分理解されないまま、誤って利用される例もしばしばである。また、幾つかの機関から出される大学ランキングには、用いられている指標が恣意的である等の指摘がある^{2,3)}にもかかわらず、多くの大学や関係機関が毎年この順位の変動に極めて敏感になっている。

このような状況に対し、科学計量学の研究者は

これまでもしばしば警告を発し、計量データの適切な利用の在り方を論じてきた^{4~7)}が、それらが結実した形で、2015年に「研究計量に関するライデン声明」（“The Leiden Manifesto for research metrics”）（以下「ライデン声明」という）が公表された⁸⁾。ライデン声明は10項目の原則(principles)から成り、研究評価における計量データの利用についてのベストプラクティスや注意点を示したものであり、研究者、管理者、評価者の全てに対する、計量データに立脚した研究評価のガイドラインと考えられる。

2. ライデン声明ができるまで

ライデン声明の基礎となったのは、2014年9月にオランダのライデン大学で開催された19th International Conference on Science and Technology Indicators (STI 2014)⁹⁾において、Dr. Diana Hicks (Georgia Institute of Technology)が行った基調講演である。ここで彼女は、研究者、研究機関、研究プログラムの評価への計量データの使用に関して次の七つの原則を示し、NatureやScienceのようなトップレベルの雑誌でこれを公表することを提案した^{10,11)}。

- ① 計量は評定の代替物ではない。
- ② 高品質のデータを得るために時間と金を費やせ。
- ③ 計量は透明かつ受け入れやすいものでなければならぬ。
- ④ データは被評価者により確認される必要がある。
- ⑤ 研究分野による違いに敏感であれ。
- ⑥ 分野と時期による違いを考慮してデータを規格化せよ。
- ⑦ 計量は戦略的目標と連携すべきである。

このスピーチに対して活発な議論がなされた。計量データ・指標の責任ある利用のガイドラインとなる原則を科学計量学コミュニティが共同して発表すべきであるという多くの意見があり、研究評価のための計量データの適正な利用のために科学計量学の研究者は積極的役割を果たすべきという声も上がった。その結果、この Hicks の 7 原則を基にして、ライデン大学科学技術研究センター (CWTS) の Dr. Paul Wouters が中心となって、このコミュニティが合意できるマニフェストをまとめることとなった。

こうしてまとめられたのが、Nature 誌で Hicks、Wouters ら 5 名の著者により発表されたライデン声明⁶⁾ である。以下の 3. は、参考文献 8) の記事からマニフェストの部分 (“TEN PRINCIPLES”) を日本語訳したものである (注 1 参照)。

なお、ライデン声明自体のホームページ¹²⁾ から、Nature 記事へのほか、各国語への翻訳記事やビデオへのリンクが張られている。

3. ライデン声明－10の原則

原則 1 定量的評価は、専門家による定性的評定の支援に用いるべきである。

定量的計量は、ピアレビューで生じやすいバイアスについて異なる見方を提示し、考察を深めるのに役立つ。同業研究者について判定することは広範な関連情報なしには難しいので、これによりピアレビューは強化されるはずである。しかしながら、評定者は意思決定を数字に任せてはならない。指標は情報に基づく判定を代替してはならない。評定者はそれぞれが行う評定に責任を保持している。

原則 2 機関、グループ又は研究者の研究目的に照らして業績を測定せよ。

プログラムの目標はその開始時に明示されるべきであり、また、業績を評価する指標は、それらの目標と明確に関係付けるべきである。指標の選択やその活用の際には、より幅広い社会経済的及び文化的な状況を考慮すべきである。科学者の研究目的は様々である。学術的知識の最前線を進める研究と、社会的問題の解決を目指す研究とは目標が異なる。学術的なアイデアの卓越性よりも、政策、産業、あるいは公衆への貢献に基づく評価もある。全ての状況に適用できる単一の評価モデルはない。

原則 3 優れた地域的研究を保護せよ。

世界の多くの地域で、優れた研究は英語で発表されると見なされている。例えば、スペインの法律は、同国の学者が高インパクトの雑誌に発表することを望ましいとしている。インパクトファクターは、米国中心で、いまだにほとんどが英語である Web of Science 収録の雑誌を対象に計算されている。こうしたバイアスは、国・地域についての研究が多い人文・社会科学において特に問題が大きい。他の多くの分野でも、国・地域という側面を持つ。例えば、サハラ以南アフリカにおける HIV の疫学などの例がある。

しかし、このような多元性や社会的関連性は、高インパクトのゲートキーパーたる英語雑誌の関心を得るような論文を創出するために抑制される傾向がある。Web of Science で高引用を得ているスペインの社会学者たちは、抽象モデルに長年取り組んでいるか、米国のデータの研究を行っている。高インパクトのスペイン語論文では、地域の労働法、高齢者のための家族健康管理、移民の雇用などのトピックについての社会学者の独自性が失われている^{a)}。優れた地域的研究の発見・それらへの報奨の付与のためには、高品質の非英語文献に基づいた計量が有用であろう。

原則 4 データ収集と分析のプロセスをオープン、透明、かつ単純に保て。

注 1 本レポートの和訳は著者が独自で行ったものであり、和訳に当たっての原文の解釈に対する全責任を有する。原文では「評価」の概念に含まれる語として “evaluation”、“assessment”、“review”、“judgement” が使われているが、本稿ではそれぞれに対して「評価」、「評定」、「レビュー」、「判定」という訳語を当てた (それらの派生語についても同様)。“metrics” の訳は「計量」に統一した。また、[] で示したのは著者による補足である。和訳に際しては、可能な範囲で正確を期しているが、和訳が定まっていない表現も多いことから、より正確な表現については元となる論文を参照願いたい。

評価のために要求されるデータベースの構成は、明確に表現された規則に従い、研究が終了する前に設定されるべきである。これは、数十年にわたり計量書誌学的評価の方法論を確立してきた学術グループと商業グループに共通の経験である。これらのグループは、査読論文に公表されたプロトコルを参考としてきた。この透明性は精密な検討を可能とした。例えば、2010年に、我々のグループの一つ（ライデン大学の科学技術研究センター（CWTS））が用いていた重要な指標の技術的性質について公開の討論が行われ、この指標の計算法の改訂に結び付いた^{b)}。最近参入している商業グループも同様な標準に従うべきである。また、ブラックボックスの評価マシンを受け入れるべきではない。

指標が単純であることは、その透明性を増すことであり長所である。しかし、単純化した計量は記録をゆがめることもある（原則7参照）。評価者は、バランス（研究過程の複雑性に忠実である単純な指標）を得ることに努めなければならない。

原則5 被評価者がデータと分析過程を確認できるようにすべきである。

データの品質を確かなものにするため、計量書誌学的調査の対象となる全ての研究者が、自分の成果が正確に同定されていることをチェックできるようにすべきである。評価過程の指揮・管理者は全て、自己確認又は第三者の検査によりデータの正確性を保証すべきである。大学は、その研究情報システムの中にこれを実装することができるだろうし、それは、これらのシステムの提供者の選択の指針であるべきである。正確で高品質なデータの照合・処理には時間と資金を要する。そのための予算を惜しんではならない。

原則6 分野により発表と引用の慣行は異なることに留意せよ。

ベストプラクティスは、一揃いの指標候補を選び、分野によってその中から選択できるようにすることである。数年前のことだが、欧州のある歴史学者のグループが、その国のピアレビュー評価において比較的低い評点を得たことがあったが、それは、この

グループが、Web of Scienceに収録される雑誌よりもむしろ図書に成果を発表しているためであった。この歴史学者は不運なことに心理学の学科に属していた〔歴史学者が心理学の学科に属していたため、雑誌論文によってピアレビュー評価がなされたという意味だと思われる〕。歴史学者や社会学者は、成果のカウントに際して図書や自国語の論文が含まれることを要求するし、計算科学者は会議論文がカウントされることを要求する。

分野により引用傾向は異なる。トップにランクされる雑誌のインパクトファクターは、数学ではおよそ3、細胞生物学ではおよそ30である。〔この差を埋めるための〕規格化した指標が必要である。最も頑健な規格化法はパーセンタイルに基づくものであり、各論文は、それが属する分野の被引用数分布中のパーセンタイル位置（例えばトップ1%、10%、20%）に従って重み付けされる。非常によく引用される論文1件は、パーセンタイル指標に基づくランキングでは、大学の位置を僅かに上げる程度だが、平均被引用数に基づくランキングでは、中位から一挙にトップまで押し上げることがあり得る^{c)}。

原則7 個々の研究者の評価は、そのポートフォリオの定性的判定に基づくべきである。

h指数^{注2}は、新しい論文がなくても年齢を重ねるほど高くなる。h指数は分野によっても異なる。トップレベルの研究者の場合、生物学では200、物理学では100、社会科学では20-30程度である^{d)}。この値は、〔h指数の計算に使う〕データベースにも依存する。計算科学分野では、Web of Scienceではh指数が10前後であるが、Google Scholarでは20-30である研究者がいる^{e)}。研究者の成果物を読んで判定する方が、一つの数字に頼るよりもずっと適切である。多数の研究者を比較する場合でも、個々の専門性、経験、活動及び影響に関するより多くの情報を考慮するやり方が最良である。

原則8 不適切な具体性や誤った精緻性を避けよ。

科学技術指標は、その概念が曖昧で不確かになりがちであり、また、普遍的には受け入れられない強い仮定に立っていることがある。例えば、被引用数の

注2 h指数は、2005年にJ. E. Hirschによって提案された研究者の業績を示す指標で、ある研究者の発表論文中h回以上引用された論文がh件以上あることを満たす最大のhを、その研究者のh指数とする。例えば、10回以上引用された論文が10件以上あるが11回以上引用された論文は11件未満なら、h指数は10である。研究の生産性（論文数）とインパクト（被引用数）を一つの数値で表すことが特徴である。当初は研究者に対する指標として提案されたが、研究グループや雑誌に対しても使われている。

意味も長らく論争されてきている。したがって、ベストプラクティスは、より頑健で複眼的な描像を与えるように複数の指標を用いることである。もし不確かさや誤差が定量化できるのであれば（例えばエラーバーの形で）、その情報を公表される指標値とともに示すべきである。それができない場合、指標の作成者は少なくとも誤った精緻性を避けるべきである。例えば、[Journal Citation Reports では] インパクトファクターを小数点以下3桁まで表示して同点の雑誌の出現を避けるようにしている。しかし、被引用数の概念上の曖昧さやランダムな変動性を考慮すれば、このような僅かなインパクトファクターの差によって雑誌を区別する意味はない。誤った精緻性は避けよ。小数点以下1桁で十分である。

原則9 評定と指標のシステム全体への効果を認識せよ。

指標は、それがもたらすインセンティブによってシステムを変化させる。これらの効果を予期しなければならない。このことは、一揃いの指標を用いることが常に望ましいことを意味する。単一の指標は、ゲーム化や目標の取り違えを招く（指標の測定自体が目標になる）。例えば、1990年代のオーストラリアでは、機関からの発表論文数に大きく依拠する数式を使って大学の研究への資金配分を行った。大学は査読制雑誌の1論文あたりの「価値」を計算することができた。2000年時点でのその価値は800豪ドル（当時のレートで約480米ドル）の研究資金に相当した。予想されたように、オーストラリアの研究者が発表する論文数は増加したが、それらは被引用数の低い雑誌に集中し、論文の質の低下を示唆した^{f)}。

原則10 指標を定期的に吟味し、改善せよ。

研究の目的と評定の目標は変化し、それに伴って研究システム自体も共進化する。かつて有用であった計量が不適切になり、新しいものが現れる。指標のシステムも見直しが必要であり、適時修正しなければならない。[原則9で述べた]単純な数式の影響に気付いて、オーストラリアは2010年に、より複雑で質の面を強調した Excellence in Research for Australia イニシアティブを導入した。

4. おわりに

ライデン声明が念頭に置いているのは、主に雑誌論文の引用に基づくデータや指標であると思われる。しかし最近、論文のインパクトを測る別のデータとして、種々のソーシャルネットワークサイトにおける論文の利用や言及によるオルトメトリクス¹³⁾データも注目されつつある。Bornmannは、ライデン声明はオルトメトリクスにも適用可能であり、オルトメトリクスの利用者はこれらの原則を十分考慮すべきであると論じている¹⁴⁾。この10原則が、科学計量学関係者のみならず、多くの研究者、研究機関の管理者、研究行政担当者の注意を引くことを期待したい。

謝辞

本レポートをまとめるに当たって、第1研究グループ客員総括主任研究官の伊地知寛博氏から貴重な助言を頂いた。また、原論文について和訳の許可を下さった、Diana Hicks氏、Nature誌に感謝申し上げます。

参考文献

a)~f)は、文献8)に挙げられている参考文献である。

- 1) Waltman, L. A review of the literature on citation impact indicators. *J. Informetrics*, 2016, 10(2), 365-391.
- 2) van Raan, A. F. J. Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics*, 2005, 62(1), 133-143.
- 3) 小野寺夏生. 大学ランキングは信頼に値するか? *化学と工業*, 2010, 63(10), 810-811.
- 4) Seglen, P. O. Causal relationship between article citedness and journal impact. *J. Am. Soc. Inf. Sci.*, 1994, 45(1), 1-11.
- 5) Garfield, E. The history and meaning of the journal impact factor. *J. Am. Med. Assoc.*, 2006, 295(1), 90-93.
- 6) Leydesdorff, L. Caveats for the use of citation indicators in research and journal evaluations. *J. Am. Soc.*

- Inf. Sci. Technol., 2008, 59(2), 278–287.
- 7) Glänzel, W and Moed, H. F. Opinion paper: thoughts and facts on bibliometric indicators. *Scientometrics*, 2013, 96(1), 381–394.
 - 8) Hicks, D., Wouters, P., Waltman, L., de Rijcke, S. and Rafols, I. The Leiden Manifesto for research metrics. *Nature*, 2015, 520(7548), 429–431 (23 April 2015).
<http://www.nature.com/news/bibliometrics-the-leiden-manifesto-for-research-metrics-1.17351>
 - 9) STI 2014 Leiden. <http://sti2014.cwts.nl/Home>
 - 10) Halevi, G. Reporting Back: STI 2014 Leiden, The Netherlands. *Research Trends*, 2014, (39), <https://www.researchtrends.com/issue-39-december-2014/reporting-back-sti-2014-leiden-the-netherlands/>
 - 11) de Rijcke, S. The Leiden manifesto in the making: proposal of a set of principles on the use of assessment metrics in the S&T indicators conference. 2014, (Sep), <https://citationculture.wordpress.com/2014/09/15/the-leiden-manifesto-in-the-making-proposal-of-a-set-of-principles-on-the-use-of-assessment-metrics-in-the-st-indicators-conference/>
 - 12) Leiden manifesto for research Metrics. <http://www.leidenmanifesto.org/>
 - 13) 林和弘. 研究論文の影響度を測定する新しい動き—論文単位で即時かつ多面的な測定を可能とする Altmetrics —, *科学技術動向*, 2013, 134, 20–29. <http://hdl.handle.net/11035/2357>
 - 14) Bornmann, L. and Haunschild, R. To what extent does the Leiden manifesto also apply to altmetrics? A discussion of the manifesto against the background of research into altmetrics. *Online Inf. Rev.*, 2016, 40(4), 529–543.
 - a) López Piñeiro, C. & Hicks, D. Reception of Spanish sociology by domestic and foreign audiences differs and has consequences for evaluation. *Res. Eval.*, 2015, 24(1), 78–89.
 - b) van Raan, A. F. J., van Leeuwen, T. N., Visser, M. S., van Eck, N. J. & Waltman, L. Rivals for the crown: Reply to Opthof and Leydesdorff. *J. Informetrics*, 2010, 4(3), 431–435.
 - c) Waltman, L. et al. The Leiden ranking 2011/2012: Data collection, indicators, and interpretation. *J. Am. Soc. Inf. Sci. Technol.*, 2012, 63(12), 2419–2432.
 - d) Hirsch, J. E. An index to quantify an individual's scientific research output. *Proc. Natl Acad. Sci. USA*, 2005, 102(46), 16569–16572.
 - e) Bar-Ilan, J. Which h-index? — A comparison of WoS, Scopus and Google Scholar. *Scientometrics*, 2008, 74(2), 257–271.
 - f) Butler, L. Explaining Australia's increased share of ISI publications—the effects of a funding formula based on publication counts. *Res. Policy*, 2003, 32(1), 143–155.