

## 学術情報をめぐる新たな評価指標：Impact Factor, *h*-index, Eigenfactor, Article Influence, Usage Factor

### New Indices for Research Assessment : Impact Factor, *h*-index, Eigenfactor, Article Influence, Usage Factor

佐 藤 翔\*

**[抄録]** 学術情報をめぐる評価指標の近年の動向について、Impact Factorと5年Impact Factor, *h*-indexをはじめとする指標群, EigenfactorとArticle Influence, Usage Factor等の利用に基づく指標群を中心に紹介する。また, その中でJournal Citation Reportsに2009年のバージョンアップで新たに導入された指標とImpact Factor, 総被引用数の関係を7つの分野を例に分析し, 新規導入指標の意義について検討した。分野によって新規導入指標と従来からある指標の関係は異なり, いずれの指標を用いる際にも評価対象の傾向とそれぞれの指標の特徴を理解しておく必要がある。

**[キーワード]** 研究評価, 引用分析, 利用分析, 計量書誌学, Impact Factor, 5年Impact Factor, *h*-index, *h*-indices, Eigenfactor, Article Influence

**[Author Abstract]** I introduced some recent studies on indices for research assessment including Impact Factor, five-year Impact Factor, *h*-index and other *h*-indices, Eigenfactor and Article Influence, Usage Factor and other usage-based indices. In addition, to reveal importance of the new indices, I examined relations between Impact Factor, Total Cites and indices which were introduced in Journal Citation Reports in 2009 in seven subjects. The relations between new indices and others are different depending on subjects and we should understand tendencies of subjects and features of indices before using any indices.

**[Keywords by Author]** research assessment, citation analysis, usage analysis, bibliometrics, Impact Factor, five-year Impact Factor, *h*-index, *h*-indices, Eigenfactor, Article Influence

#### 1. はじめに

2009年1月, トムソン・ロイター社は学術誌評価分析データベース, Journal Citation Reports (JCR) のバージョンアップを発表した<sup>1)</sup>。JCRは学術雑誌評価のスタンダードとなっているImpact Factor (IF) の算出・公表元であるが, バージョンアップでは新たに5年Impact Factor (5年IF), 引用への重み付けを行った指

標であるEigenfactor, Article Influenceなどの指標が導入され, 従来の指標と見比べることができるようになっている。

JCR新規導入指標に限らず, 研究者の評価指標として注目を集めている*h*-indexをはじめ, 学術研究の新たな評価指標を探る研究は近年特に盛んである。この背景には研究者や研究機関の評価(およびそれに基づく資金配分や雇用昇進の決定)に関する活動が盛んになる中で, 計量的な評価指標が求められるようになってきていることがある。

しかし指標にはそれぞれ特徴があり, これを理解せずに用いることは難しい。本稿では近年の研

\* Sho SATO

筑波大学大学院図書館情報メディア研究科  
〒305-8550 つくば市春日1-2  
E-mail: min2fly@slis.tsukuba.ac.jp

究にかかわる指標を巡る動向について、特に JCR 新規導入指標を中心に紹介し、その特徴と従来の指標との関係についての分析結果を述べる。なお研究評価指標についての研究は膨大であり、ここで取り上げるものはその一部に過ぎないことをあらかじめお断りしておく。

## 2. 研究評価指標を巡る動向

### 2.1. Impact Factor

IF とはある雑誌の掲載論文が平均してどれくらい引用されているかを示す指標である。より正確に言えば、ある雑誌の 2007 年の IF は 2007 年に出版された雑誌掲載論文が当該雑誌の 2005～2006 年の論文を引用した回数を、2005～2006 年の当該雑誌掲載論文数で割ったものであり、ある年から過去 2 年分への引用が評価の対象となる。この IF の特徴と問題点についてはすでに多くの論考があるのでここで紙数を割くことは避けるが、近年の動向としては国際数学連盟等による合同委員会が引用統計の研究評価への導入の問題点を指摘する報告『Citation Statistics』を公表し、その中で IF の問題や誤用についても指摘している<sup>2,3)</sup>。問題点の 1 つに、IF は年ごとにかなり変化する（にもかかわらず特定年の IF で雑誌を比較してしまうことが多い）ということが挙げられている。これについては逸村らも主要科学誌の IF の変遷を示し、被引用数の増減以外の理由でも少なからず変動することを指摘している

(図 1)<sup>4)</sup>。

また、IF を巡るもう 1 つの大きな動向として、前述のように 5 年 IF の JCR への導入が挙げられる。5 年 IF は文字通り、IF の対象期間を 5 年に延ばしたものである。分野によっては IF の対象期間となる過去 2 年よりも古い論文を多く引用することもあり、IF の分析対象期間を疑問視する声は『Citation Statistics』報告等でも述べられてきたが、トムソン・ロイターは 5 年 IF の導入でこのような議論に 1 つの答えを示した形になる。5 年 IF の導入効果については第 3 章で検討する。

### 2.2. h-indices

Hirsch は 2005 年に、研究者の生産性と論文の質を同時に評価できる新たな指標として *h-index* を提唱した。これは「研究者が発表した論文のうち、被引用数が *h* 回以上の論文が *h* 本以上あることを満たす数値」をその研究者の *h-index* とするというものである<sup>5)</sup>。研究者の論文生産性と質を 1 つの指標で表現できる（とされる）こと、ある研究者の発表論文とそれぞれの被引用数がわかれば算出できることなどからこの指標は注目を集めた。発表から 4 年経った 2009 年 3 月現在で Hirsch が *h-index* を提案した論文は既に 200 回以上引用されている（Web of Science による）。Web of Science をはじめ、Scopus など *h-index* を取り入れたデータベースも既に存在し、研究者の業績評価指標として一定の地位を確立しつつ

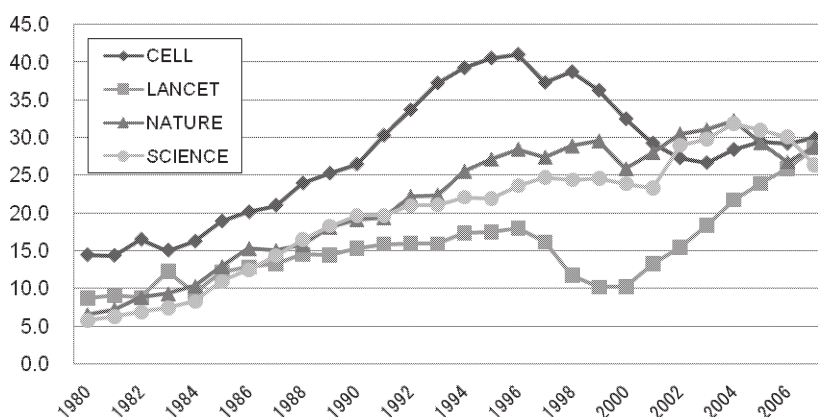


図 1 主要科学誌の IF 変遷 (逸村ほか<sup>4)</sup> に 2004～2007 のデータを追加したもの)

ある。

広く普及すると同時に、*h-index* の特徴を生かしながらその弱点を改良した指標を新たに提唱する研究も次々と出てきている。*h-index* の利点の1つにはその頑健性（1, 2本特別に引用される「大当たり」が出てても *h-index* は大きく変化しない）が挙げられるが、これは同時に *h-index* の問題点であるともされる。例えば被引用数が10回の論文を10本発表した研究者と、被引用数100回の論文を10本発表し、他にはまだ発表していない研究者の *h-index* はいずれも10となる。これを克服しようとして考案されたのが *g-index*（被引用数上位 *g* 位までの論文の被引用数の和  $\geq g^2$  を満たす *g* の最大値<sup>6)</sup>、*A-index*（*h-index* の算出に用いる *h* 本の論文の平均被引用数）、*R-index*（*h-index* の算出に用いる *h* 本の論文の被引用数の和の平方根<sup>7)</sup>）などの被引用数上位論文の被引用数を評価に反映した指標である。それぞれ *g-index* は単独で計算できるが頑健性は失われる、*A-index* や *R-index* は *h-index* と同時にでなければ算出できないが、*h-index* と合わせて用いることで頑健性は保ちつつ被引用数上位論文の影響も併せて評価できる、という特徴を持つ。

*h-index* を改良・修正した指標はそのほかにもあり、また *h-index* の考え方を学術雑誌の評価など研究者以外の評価に用いる例も出てきている<sup>8)</sup>。これらの *h-index* に連なる指標群を *h-indices* と呼び、近年の研究評価の領域ではこの *h-indices* の研究が大きな位置を占めつつある。ただし *h-indices* の評価への利用については疑問視する声もあり、『Citation Statistics』報告の中では「(*h-index* の有用性は) 納得いく証拠によって裏付けられていない」、「研究履歴を理解する上で、(*h-index* の算出時に) 捨て去ってしまった情報が必要である」等として *h-indices* を批判している<sup>2,3)</sup>。

### 2.3. Eigenfactor, Article Influence

引用に重み付けを行うこと一すなわち、『Nature』のような雑誌からの引用と無名の雑誌からの引用を等価値として扱わないことは新しい考え方ではない。Eigenfactor.org の web サイトではそのランク付けの手法を「30年にわたり続

いてきた研究を拡張したもの」としており<sup>9)</sup>、また Google の PageRank は web ページのランキングにこのような手法を導入したものである<sup>10)</sup>。Eigenfactor.org では pagerank 同様の重み付けをトムソン・ロイター社の JCR のデータに基づいて行った場合の学術雑誌のランキングを提供していた。これが2009年に新たに JCR で提供されることになった Eigenfactor (EF) と Article Influence (AI) である。

EF における引用の重み付けの最も基本的な考え方は「質の高い雑誌からの1件の引用は、重要でない雑誌からの多数の引用より価値がある」というものであり<sup>11)</sup>、これは「多くの雑誌から引用されている雑誌からの引用は価値がある」、「少ない雑誌しか引用していない雑誌からの引用は価値がある」という2側面を含むものである。EF の詳細な計算方法については IF や *h-index* のようには単純でないためここでは省くが（詳細は Eigenfactor.org の web サイト<sup>9)</sup> を参照。中西印刷(株)の web サイトに邦訳も掲載されている<sup>12)</sup>）、EF の考案者である Bergstrom によれば、EF とは研究者がランダムに選んだある雑誌を手に取り、その雑誌に引用されている中からランダムに選んだ別の雑誌をまた手に取り…という行動を繰り返して行った場合に、当該雑誌を訪れる頻度をパーセンテージで示したものであり、パーセンテージであるため全雑誌の EF を合計すると100になるという<sup>13)</sup>（実際に筆者が JCR Science Edition/Social Sciences Edition 2007 で確認したところ、全雑誌の EF の合計は102をやや超える程度となった）。この考えに基づけば、他の多くの雑誌に引用されている引用元からの引用に価値があるのは、引用元の雑誌が読まれる可能性が高いためそこから引用された当該雑誌も読まれる可能性が高くなるからであり、多くの雑誌を引用している引用元からの引用の価値が低いのは、当該雑誌のほかにも引用されている雑誌が多いため当該雑誌が選ばれる可能性が低いからである、と説明できる。つまり EF とは「ランダムに引用を辿り続ける」という研究者を仮定した場合に、その雑誌がどれだけ読まれうるかを示す指標であると言える（なお、EF の算出時には自誌引用は計

算から除外されている)。また、Bergstrom はこのような重み付けを行うことにより、平均して引用数の多い分野では個々の引用の価値が引用数の少ない分野より下がる、あるいはレビュー論文からの引用は価値が下がるといった効果が生まれ、分野や雑誌の種別を超えた比較が可能になるともしている<sup>11)</sup>。これは分野ごとの引用慣行によって大きく値が異なる IF 等とは異なる特徴と言えるが、その信頼性については第 3 章で議論する。

EF は雑誌全体がどれだけ読まれるかの指標であるため、掲載論文数の多い雑誌では値が高くなる。そこで IF の場合と同様に雑誌の規模を超えた比較を可能にしたのが AI である。ただし IF が総被引用数を掲載論文数で単純に割っているのに対し、AI は EF を掲載論文数ではなく、当該雑誌の掲載論文数が全対象雑誌の論文数中で占める割合（パーセンテージ）で割ったものであり、掲載論文数に対し平均的な EF を得ていればその値は 1 となる<sup>11)</sup>。これによって IF 同様に雑誌の規模を考慮に入れた比較が可能になるが、実際には 1 つの雑誌内でも掲載論文の被引用状況は異なるため、AI によってその雑誌に掲載された論文の質を評価することはできないのもまた IF と同様であるとされている<sup>14)</sup>。

なお、現在 Eigenfactor.org, および JCR 2007 年版で提供されているのは過去 5 年間に期間を区切った EF および AI である（2007 年に出版された雑誌が、2002～2006 年の雑誌論文を引用した回数に基づき計算されている）。5 年 IF と同様に 5 年より古い論文への引用が評価に含まれていないことに注意が必要である。

EF への批判として、『Citation Statistics』では *h-index* の場合と同様に「分析に基づいた正当化もしておらず、評価するのは困難」と述べた上で、「いっそう複雑な計算に基づいているため、その背後にある（しばしば隠された）仮定を識別するのは、ほとんどの人にとって容易ではない」とも指摘している<sup>2,3)</sup>。具体例としては自誌引用の計算からの除外について触れており、『Mathematical Reviews』引用データベースのデータを一見しただけでも、引用の約 3 分の 1 を捨て去ることになる」と批判している。

## 2.4. 利用に基づく指標

これまでに挙げた指標はすべて論文の引用関係に基づくものである。しかし近年では学術雑誌の電子化の進展により、雑誌や論文単位での利用記録が容易に保存・活用できるようになったことを受け、利用記録に基づく雑誌の評価への関心が高まってきている。

Darmoni らは雑誌の平均利用数で当該誌の利用数を割った値を Reading Factor (RF) と定義し、フランスの Rouhen 大学病院における 1998 年の電子ジャーナル利用データにこれを適用した結果について報告している<sup>15)</sup>。その結果、RF は IF と異なる傾向を示し、RF が IF とは異なる計量的な情報を提供するものになりうるとしている（ただし 1997 年のデータに基づいており、電子ジャーナルの利用が浸透していない時期のものであることには注意が要る）。

Bollen らは雑誌の利用に基づく指標として、電子ジャーナル利用データに IF と同様の方法を適用した Usage Impact Factor (UIF) を定義している<sup>16)</sup>。例えば 2007 年のある雑誌の UIF は、2007 年における 2005～2006 年に出版された当該雑誌論文への利用を、2005～2006 年の掲載論文数で割ったものであると定義される。これをカリフォルニア州立大学の 2003～2005 年の利用データに適用した結果、UIF と IF は全体では有意な弱い負の相関があり、分野を区切ってみてもほとんどの分野で有意な相関がないか、弱い負の相関があったとされている。ただし大学院に所属する学生・教員数が多い分野では正の相関があったことなどから、大学院学生・教員が多ければ IF と UIF は近づくのではないかとしており、IF と UIF の集計対象となるサンプル（IF は全世界の学術的な文献の著者、UIF はカリフォルニア州立大学のみ、学部に属する学生や教員も含む利用者）の違いに言及している。

上記の 2 例はいずれもローカルな（ある大学等の）利用記録に基づくものであり、Bollen が指摘しているように IF 等と違いサンプルとなるコミュニティが限られている。また、ローカルな利用記録であるので自機関が購読していない雑誌には適用できず、図書館員等にとってはキャンセル

の参考にはできても購読の参考にはできないものである。そこで Shepherd はある期間における雑誌の利用全体を、ある期間にオンラインで出版された論文の総数で割った値を Usage Factor (UF) と定義し、その需要の有無と問題点について研究者、図書館員、出版社等に尋ねている<sup>17)</sup>。この中では研究者の70%がUFを歓迎するとしたこと、もしUFが利用できれば、図書館員は新規雑誌購入選択時には利用者からのフィードバックに次いで2番目に、契約更新・キャンセル時には利用者のフィードバック、実際の利用に次いで3番目に重要な指標となると答えたこと等が示されている。

このように利用に基づく指標への期待は大きい一方で、利用データには「論文をどの程度詳しく読んだのかわからない」、「偶然に、あるいは一時的な好奇心でアクセスすることが起こりやすい」と言った問題点も指摘されており、これは引用データの偏りや信頼性の問題以上に深刻であるとの指摘もある<sup>18)</sup>。現在利用データに基づく指標はまだ一般化していないが、今後そうした指標が日常的に使えるようになった場合には上記の点に留意する必要がある。

## 2.5. 複数指標の比較

以上のように近年では様々な研究評価指標が考案・提唱されている。これらの中で異なる指標を用いた場合に雑誌や研究者の評価がどのように変化するのか、というのは誰もが興味を持つところである。IFとRFの比較についてはすでに紹介したが、その他にも同じ雑誌や研究者に複数の指標を適用し、結果を比較した研究はいくつか存在する。

Leydesdorff は同一の雑誌群における総被引用数等 (JCR に基づく場合と Scopus に基づく場合の双方) と IF や *h*-index, PageRank, 媒介中心性<sup>#1)</sup> や近接中心性<sup>#2)</sup> などのその他のネットワーク評価指標について因子分析を行い、評価指標にはサイズの因子とインパクトの因子があり、PageRank は主としてサイズに基づく指標であ

ること、*h*-index はインパクトとサイズを結びつける指標となること等を示した<sup>19)</sup>。一方で Bornmann らはほぼ同時期に研究者の *h*-index, *g*-index 等の *h*-indices と総被引用数、総発表論文数等の指標について因子分析を行い、*h*-index や *g*-index は総発表論文数と同じく生産性に関する指標であり、*A*-index や *R*-index が総被引用数と同じくインパクトに関する指標であるとしている<sup>20)</sup>。このような異なる結果になるのは分析の対象にした指標群が異なるためと考えられる。

Bollen らは IF や *h*-index, 引用データに基づく PageRank 等の引用に基づく指標に加え、自身らが構築した利用に基づくネットワークから利用データに基づく PageRank や UIF 等の指標を算出し、これら引用に基づくものと利用に基づくものを含む39の雑誌の評価指標間の関係について分析している<sup>21)</sup>。その結果、媒介中心性などでは利用に基づいた場合と引用に基づいた場合で強い相関があるのに対し、IF は引用に基づく他の指標との間であっても相関が低い場合があるとして、研究のインパクトを測る指標としての IF に疑問を投げかけている。

Davis は2006年の医学分野におけるIF、総被引用数とEFの関係について比較し、EFと総被引用数が強く相関していること等を明らかにしている<sup>22)</sup>。そこから Davis は医学分野において重み付けのメカニズムによって測られる「名声」は総被引用数等で測られる「人気」と極めて似通ったものであると結論付けている。しかしこれは医学分野に限った分析に基づくものであり、他の分野でも同様の結果になるとは限らない。また Davis は EF と総被引用数、IF の相関については見ているが AI との関係については見ていない。実際には IF と同様に掲載論文数を考慮に入れた指標は AI であり、これと IF の比較も併せて行うべきであると考えられる。

そこで以下では Davis の手法にならい、5年IF、EF、AI等のJCR新規導入指標と従来からある総被引用数、IFとの関係について、7つの

注1) 引用等による文献のネットワーク内において、すべての文献の組み合わせについて両者を最短距離で結んだ場合に当該文献を経由するものがどれだけあるか、を表す指標。

注2) 文献のネットワーク内で当該文献が他のすべての文献とどの程度の近さにあるかを表す指標。

分野を例に見てみることにする。

### 3. JCR に新たに導入された指標の分野間比較

JCR に新たに導入された指標と IF 等の従来の指標間の関係と新指標の意義を考察するために、1) IF と 5 年 IF の比較, 2) 総被引用数と IF, 5 年 IF, EF, AI の相関関係の分析, 3) 5 年 IF と AI の比較を行った。分析対象は JCR Science Edition 2007 収録誌のうち化学 (7つの細分を含む), 工学 (14の細分を含む), 数学 (3つの細分を含む), 医学 (3つの細分を含む), 薬学, 物理学 (8つの細分を含む) および Social Sciences Edition 2007 より図書館情報学の7分野で IF, 5年 IF, EF, AI をすべて持つ雑誌である (各分野の分析対象誌数は表1参照)。

#### 3.1. IF と 5 年 IF

表1は各分野における IF, 5年 IF の平均値と IF に対する 5年 IF の比率を示したもので、表2は同じく IF, 5年 IF の中央値と IF に対する 5年 IF の比率を示したものである。平均値と併せて中央値を示したのは、被引用数や IF 等の分布は偏っており、平均値は一部の IF 等の高い雑誌に

よる影響を受けやすいと考えられるからである。

上記の理由から平均値と中央値を併せて示したが、IF と 5 年 IF の比較においては多くの分野で両者の傾向は似通っていた。化学, 工学, 数学, 薬学, 図書館情報学では平均値・中央値ともに 5 年 IF の方が高い。医学では平均値が 5 年 IF の方が低く、物理学では中央値が 5 年 IF の方が低い。多くの分野で IF と 5 年 IF の差は平均値・中央値とも 1 割未満とわずかであるが、数学と工学, 図書館情報学では比較的差が大きく、特に図書館情報学では 5 年 IF の方が平均値・中央値とも 3 割前後高くなっている。

分析対象とした分野の中で、物理学は引用半減期が 5~8 年、医学は 6~7 年 (JCR 2007 年版による) と短く、いずれも出版後 2 年以内に引用のピークが来る。そのため IF と 5 年 IF の差が小さいか、IF の方が高くなる。これらの分野にとっては IF の期間の区切りが実態に即したものになっていると言える。

一方、数学は引用半減期が 9 年以上と長いために 5 年 IF の方が高くなると考えられる。工学は引用半減期が細分によって 5~10 年と幅が広い

表1 各分野の IF, 5 年 IF の平均値とその比

	a. IF 平均値	b. 5 年 IF 平均値	b/a
化学 (N=423)	2.065	2.256	1.093
工学 (N=690)	0.899	1.091	1.213
数学 (N=351)	0.788	0.943	1.197
医学 (N=172)	2.965	2.908	0.981
薬学 (N=188)	2.754	2.879	1.046
物理学 (N=281)	2.131	2.225	1.044
図書館情報学 (N=52)	0.939	1.264	1.346

表2 各分野の IF, 5 年 IF の中央値とその比

	c. IF 中央値	d. 5 年 IF 中央値	d/c
化学 (N=423)	1.389	1.489	1.072
工学 (N=690)	0.669	0.852	1.274
数学 (N=377)	0.625	0.733	1.173
医学 (N=172)	1.466	1.471	1.004
薬学 (N=188)	2.140	2.212	1.034
物理学 (N=281)	1.306	1.256	0.962
図書館情報学 (N=52)	0.611	0.786	1.287

が、多くの細分化した分野で引用のピークは出版後2~3年であり、4~5年たっても2年目と同様かそれ以上に引用が続く分野も多い。図書館情報学は引用半減期は6.7年と特に長いわけではないが、引用のピークが出版の3~4年後に訪れる(2007年の論文からの引用が最も多いのは2004年、ついで2003年に出版された論文)。そのため

分析対象の中でも5年IFの方がIFより特になくなってきている。このように、5年IFの方が引用行動の実態を的確にとらえている分野にとっては、これを新しいJCRで容易に参照できるようになったことの意義は大きいと言えるだろう。

### 3.2. 指標間の相関関係

表3は各分野における総被引用数、IF、5年

表3 各分野における指標間のスピアマンの順位相関係数\*

化学 (N=423)

	総被引用数	IF	5年IF	EF	AI
総被引用数		0.695	0.664	0.954	0.619
IF	0.695		0.978	0.775	0.949
5年IF	0.664	0.978		0.749	0.979
EF	0.954	0.775	0.749		0.718
AI	0.619	0.949	0.979	0.718	

工学 (N=690)

	総被引用数	IF	5年IF	EF	AI
総被引用数		0.799	0.809	0.957	0.758
IF	0.799		0.965	0.799	0.872
5年IF	0.809	0.965		0.820	0.931
EF	0.957	0.799	0.820		0.820
AI	0.758	0.872	0.931	0.820	

数学 (N=351)

	総被引用数	IF	5年IF	EF	AI
総被引用数		0.521	0.528	0.873	0.490
IF	0.521		0.944	0.542	0.718
5年IF	0.528	0.944		0.547	0.771
EF	0.873	0.542	0.547		0.626
AI	0.490	0.718	0.771	0.626	

医学 (N=172)

	総被引用数	IF	5年IF	EF	AI
総被引用数		0.706	0.713	0.943	0.686
IF	0.706		0.982	0.806	0.954
5年IF	0.713	0.982		0.819	0.980
EF	0.943	0.806	0.819		0.810
AI	0.686	0.954	0.980	0.810	

薬学 (N=188)

	総被引用数	IF	5年IF	EF	AI
総被引用数		0.606	0.613	0.946	0.566
IF	0.606		0.979	0.742	0.962
5年IF	0.613	0.979		0.744	0.978
EF	0.946	0.742	0.744		0.716
AI	0.566	0.962	0.978	0.716	

物理学 (N=281)

	総被引用数	IF	5年IF	EF	AI
総被引用数		0.595	0.559	0.953	0.438
IF	0.595		0.963	0.661	0.886
5年IF	0.559	0.963		0.635	0.938
EF	0.953	0.661	0.635		0.546
AI	0.438	0.886	0.938	0.546	

図書館情報学 (N=52)

	総被引用数	IF	5年IF	EF	AI
総被引用数		0.846	0.875	0.925	0.855
IF	0.846		0.931	0.806	0.889
5年IF	0.875	0.931		0.798	0.948
EF	0.925	0.806	0.798		0.857
AI	0.855	0.889	0.948	0.857	

\* 順位データに対する相関関係の指標。一般に「相関係数」と呼ばれるピアソンの相関係数はデータが正規分布することを前提とするため、正規分布しない(低い値にデータが偏る)ことが知られる被引用数やそれに基づく指標には適用することができない。これに対しスピアマンの順位相関係数はデータが正規分布することを前提としないため、本稿ではこれを分析に用いた。なお、医学分野でEFと他の指標の関係を見たDavisもスピアマンの順位相関係数を使用している<sup>22)</sup>。

IF, EF, AI の関係について、指標間のスピアマンの順位相関係数を示したものである（すべての指標間で  $p < 0.001$  で有意な相関が存在）。

表から、多くの分野で総被引用数と EF, IF と 5年 IF, 5年 IF と AI の間で特に強い正の相関（相関係数  $> 0.9$ ）があることがわかる。IF と 5年 IF は全分野で強い相関関係にあり、前節の分析で 5年 IF の方が 2年 IF より特に高くなった図書館情報学分野においても相関係数は 0.93 以上である。IF による雑誌ランキングと 5年 IF によるそれとは（分野を区切れば）似通っていると考えられる。

総被引用数と EF も強い相関関係にある。両者の相関関係については前述の医学分野において分析を行った Davis も指摘しており<sup>22)</sup>、本稿でも医学分野における結果は Davis の分析とほぼ同様となっている（Davis の分析対象は医学分野の中でも MEDICINE, GENERAL & INTERNAL のみを対象とし、期間も 2006 年を対象としているが、結果はスピアマンの順位相関係数 0.95 と本稿における結果と類似している）。EF と総被引用数は対応関係にある（いずれも掲載論文数を考慮しない、雑誌の影響度を表す指標である）ことを考えればこれは当然と言える。ただし

数学分野では比較的相関が弱い。

5年 IF と AI も対応関係にあり、いずれも過去 5 年間の引用を対象とし、雑誌掲載論文数を考慮に入れた指標である。しかし他の対応関係にある指標同士の比較の場合と異なり、5年 IF と AI の関係には分野によって大きな差がある。図 2 は 7 分野のうち 5年 IF と AI が強い相関関係にある医学、化学、薬学と比較的相関の弱い工学、数学の 5 つの分野の雑誌ごとの 5年 IF と AI の分布について、横軸に 5年 IF、縦軸に AI を取り示したものである。また、図 3 は同じく 5 つの分野の 5年 IF と AI の分布について、比較のために 5年 IF, AI ともに 10 以下の雑誌に限定して示したものである。

医学、化学、薬学の 3 分野では 5年 IF と AI の相関係数は約 0.98 であり、非常に強い正の相関が存在する。5年 IF と AI による雑誌の分布は一直線上に並び、同じ 5年 IF を取る雑誌は AI の値もほとんど変わらない。これらの分野では自誌引用の除外や引用に対する重み付けは雑誌のランキングに重大な影響を及ぼさないとと言える。

一方、工学では相関係数は約 0.93 であり、強い正の相関は存在するものの 5年 IF と AI によ

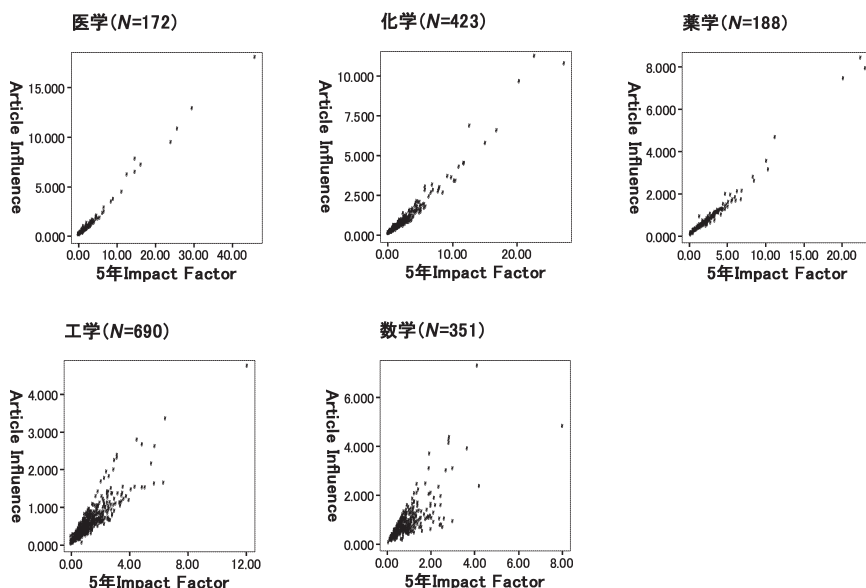


図 2 5 分野の 5 年 IF と AI の分布（すべての雑誌）



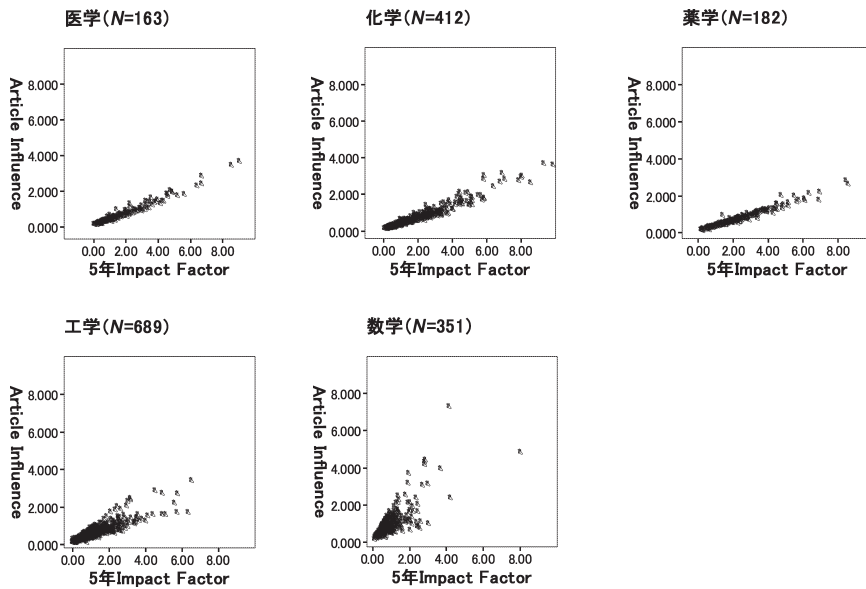


図3 5分野の5年IFとAIの分布 (5年IF, AIともに10以下の雑誌)

る雑誌の分布は直線と言うには幅の広い分布を示す。5年IFが高い雑誌はAIも高い傾向があるが、同じ5年IFであってもAIの値が大きく異なる場合もある。

さらに数学では5年IFとAIの相関係数は約0.77と他の分野に比べ低く、相関関係はあるものの5年IFは高いがAIは低い、あるいはその逆という雑誌が多い。例えば数学分野の中でもJCRにおける「MATHEMATICS, INTERDISCIPLINARY APPLICATIONS」に属する雑誌のうち、『STRUCTURAL EQUATION MODELING-A MULTIDISCIPLINARY JOURNAL』誌と『ECONOMETRICA』誌(数学分野の散布図の真ん中付近の2誌)は5年IFでは前者が4.222, 後者が4.152と近い値を取るが、AIで見ると前者が2.283であるのに対し後者は7.208と大きな差がある。これは同じく学際的な数学分野の雑誌であっても、前者は主に心理学や教育心理学等の雑誌に、後者は経済学の雑誌に引用されており、引用元となる分野の傾向の違いがAIの値に大きく影響したと考えられる。

このように医学や化学、薬学では5年IFで見てもAIで見ても雑誌のランキングはあまり変わ

らず、自誌引用の除外や引用への重み付けは雑誌のランキングに大きく影響しない。一方、工学や数学分野においては5年IFとAIでは雑誌のランキングが大きく変動し、自誌引用の除外と引用の重み付けが大きく影響する分野であると言える。後者のような分野では5年IFとは別にAIによる評価を見ることも大きな意義を持つと考えられる。

### 3.3. 5年IFとAI

表4は各分野における5年IFとAIの平均値、表5は同じく中央値を示したものである。

前節の分析で医学や化学、薬学では分野内における雑誌のランキングは5年IFとAIで大きく差がないこと、工学や数学では変化が大きいことがわかったが、分野を超えた比較では5年IFとAIでは大きく異なる結果を示す。5年IFでは平均値、中央値ともに最下位であった数学はAIで見ると平均値で工学、図書館情報学、化学を抜き4位に、中央値では分析対象中最高となった。AIはEFに基づいて算出されており、EFでは「よく引用される雑誌からの引用は重みが大きい」という重み付けのほかに「あまり引用をしない雑誌からの引用は重みが大きい」という重み付けも行っている。そのためEFの考案者らは分野によっ

表4 各分野の5年IFとAIの平均値

	5年IF平均値	AI平均値
化学 (N=423)	2.256	0.734
工学 (N=690)	1.091	0.450
数学 (N=351)	0.943	0.810
医学 (N=172)	2.908	1.047
薬学 (N=188)	2.879	0.834
物理学 (N=281)	2.225	1.044
図書館情報学 (N=52)	1.264	0.459

表5 各分野の5年IFとAIの中央値

	5年IF中央値	AI中央値
化学 (N=423)	1.489	0.438
工学 (N=690)	0.852	0.351
数学 (N=377)	0.733	0.627
医学 (N=172)	1.471	0.458
薬学 (N=188)	2.212	0.586
物理学 (N=281)	1.256	0.584
図書館情報学 (N=52)	0.786	0.277

て引用慣行が異なる（数学はあまり引用をせず、医学等ではよく引用をする）ことを考慮に入れた、分野間での比較が可能になる指標であるとしているが、本稿の結果から少なくとも引用数自体がそれほど多くない分野にとっては有利となる指標であると言える。ただし分野を超えて雑誌を比較する必要があるのか、という疑問は残る。

以上のように、JCR新規導入指標と従来からある指標間、あるいは新規導入指標間の関係はその分野の引用の在り方によって異なることがわかった。これらの指標を用いる際には対象とする分野における引用の傾向を把握しておく必要がある。また、各指標がどのようなデータや考え方に基づいた、何を表す指標なのかにも注意しなければならない。EFは重み付けを行っている点で単なる総被引用数より研究評価にふさわしい指標のようにも捉えられるが、一方で前述のように自誌引用を計算から除外しており、そのことの持つ意味は正確にはわかっていない。注意すべきは著者の自著論文引用を除いているわけではなく、ある雑誌から当該雑誌への引用を除いている点である。例えば雑誌数が少なく自誌引用が多い分野で

あれば、これを無視することは引用の実態を正確には反映できていないことにもなりうる。EFに基づいて算出されるAIも重み付けを行った指標であり、前述のように数学分野等にとっては有利な指標ともなりうるが、一方で数学分野では分野内の雑誌ランキングもAIに基づく場合と5年IFに基づく場合で大きく異なる傾向があり、ここでも自誌引用の除外と引用への重み付けの意味について慎重に考える必要がある。また、工学や図書館情報学ではAIを用いた場合でも分野を超えた順位は変化していない（低いままである）。JCR収録分野内で影響力に差があるとも考えられるが、分野ごとの引用の構造、例えば分野の中心的な役割を果たす雑誌が存在するか否かや他分野からも引用される雑誌があるか<sup>23)</sup>といったことが、EFやAIの重み付けにどう影響するかを確認せずに Bergstrom が言うような分野を超えた比較に適用することは危険であるとも言える。各指標が何を表し、何に役立て何に使えないのかを考えながら利用することが重要である。

#### 4. おわりに

近年の学術研究評価指標の動向と、JCR新規導入指標と従来からある指標間の関係について見てきた。これらの指標を見比べてみてわかることは、いずれの指標も論文数や被引用数（あるいはダウンロード数）などの単純な数値をどう操作すれば公平かつ最大公約数的な賛同が得られる評価指標となりうるかを動機として考えられたものであるにもかかわらず、実際にはやはりそれぞれが測っているものは指標によって異なるということである。IFは2ないし5年間に掲載された平均的な論文への注目度の指標であり、*h-index*は生産性と論文の質を同時に測ろうとするものであり、EFはその雑誌が研究者にどれだけ読まれるかの指標であり、UFはより直接的にどれだけ読まれているかを評価する。それぞれ評価するのは異なり、単純に優劣を付けられるものではない。なぜならば、論文や研究者、雑誌の「質」に関しての絶対的な、正解となる評価指標は存在しないからである。紹介した中には指標間の関係を見ることで他と異なる傾向を持つ指標を「コンセ

ンサスを得られるものではない」とする研究もあるが、異なるものを測る指標についてコンセンサスを探ろうとすることには「目隠しをして象の身体の一部に触った人々がいたとして、鼻に触った人のグループや脚に触った人のグループ、しっぽに触った人の話を聞いてそれらの中間をとっても『コンセンサス』にはならない。それらはそれぞれ異なる身体のパーツを表現している」との批判もある<sup>24)</sup>。何か1つの「質を測るもの」, 「研究評価のコンセンサス」となる指標を求める行為は、評価者が自分が本当は何を測っているかがわからないまま評価を行う危険性をはらむ。目的や対象に応じて適切な指標を用い、時には複数用いて比較することが重要である。長さを測りたければ定規かメジャーを使えばよく、重さを量りたければね秤を、質量が測りたければ天秤を用いばいい。研究評価も同じことであり、そもそも単一の指標でなんでも測ろうとすることが無理なのである。

本稿で取り上げた以外にも数多くの評価指標が存在する。今後も研究評価への注目を反映し新たな指標が次々と出てくるものと考えられる。いずれの指標も研究者や雑誌の動向を明らかにできる便利な道具となりうるものである。道具を乱用・過信することなく、各々の特性を理解して人間が上手に使いこなすことを心がけねばならない。

本稿の執筆にあたり筑波大学大学院図書館情報メディア研究科の逸村裕教授、トムソン・ロイター社の宮入暢子氏より多大なるご協力をいただきました。深く感謝申し上げます。

## 参 考 文 献

- 1) Thomson Reuters. “トムソン・ロイター：Journal Citation Reports®の最新版を発表”. トムソン・ロイター サイエンス・インテリジェンス 特許 文献 引用文献 知的財産 医薬 原案 工業規格. (オンライン), 入手先 <[http://www.thomsonscientific.jp/news/press/pr\\_200901/350008.shtml](http://www.thomsonscientific.jp/news/press/pr_200901/350008.shtml)>, (参照 2009-03-19).
- 2) Ewing, J. et al. Citation Statistics. Joint IMU/ICIAM/IMS-Committee on Quantitative Assessment of Research, 2008, 26 p.
- 3) 日本数学会情報システム運用委員会. “IMUからの情報”. 日本数学会. (オンライン), 入手先

<<http://mathsoc.jp/IMU/>>, (参照 2009-03-19).

- 4) 逸村 裕ほか. インパクトファクター：研究評価と学術雑誌. 名古屋高等教育研究. 6, 2006, 131-144. 入手先 <<http://www.cshe.nagoya-u.ac.jp/publications/journal/no6/10.pdf>>, (参照 2009-03-19).
- 5) Hirsch, J. E. An index to quantify an individual's scientific research output. Proc Natl Acad Sci USA. 102(46), 2005, 16569-16572.
- 6) Egghe, L. Theory and practices of the *g*-index. Scientometrics. 69(1), 2006, 131-152.
- 7) Jin, B. et al. The *R*- and *AR*-indices: Complementing the *h*-index. Chin Sci Bull. 52(6), 2007, 855-863.
- 8) Bornmann, L. et al. Hirsch-type index values for organic chemistry journals: a comparison of new metrics with the Journal Impact Factor. European J Org Chem. 10, 2009, 1471-1476.
- 9) Bergstrom, C. “methods”. eigenfactor. org. (オンライン), 入手先 <<http://www.eigenfactor.org/methods.htm>>, (参照 2009-03-19).
- 10) Brin, S. et al. The anatomy of a large-scale hypertextual web search engine. WWW7/Computer Networks. 30(1-7), 1998, 107-117.
- 11) Bergstrom, C. Eigenfactor: Measuring the value and prestige of scholarly journals. C & RL News. 68(5), 2007. (オンライン), 入手先 <<http://www.ala.org/ala/mgrps/divs/acrl/publications/crlnews/backissues2007/may07/eigenfactor.cfm>>, (参照 2009-03-19).
- 12) “EigenFactor (アイゲンファクター) の数学的基礎”. 中西印刷株式会社公式サイト. (オンライン), 入手先 <[http://www.nacos.com/nakanishi/104\\_abouteigen.html](http://www.nacos.com/nakanishi/104_abouteigen.html)>, (参照 2009-03-19).
- 13) Bergstrom, C. et al. The Eigenfactor™ metrics. J Neurosci. 28(45), 2008, 11433-11434.
- 14) Bergstrom, C. et al. Assessing citations with the Eigenfactor™ metrics. Neurology. 71(23), 2008, 1850-1851.
- 15) Darmoni, S. J. et al. Reading factor: a new bibliometric criterion for managing digital libraries. J Med Libr Assoc. 90(3), 2002, 323-327.
- 16) Bollen, J. et al. Usage impact factor: The effects of sample characteristics on usage-based impact metrics. J Am Soc Inf Sci Technol. 59(1), 2008, 136-149.
- 17) Shepherd, P. T. Final report on the investigation into the feasibility of developing and implementing journal usage factors. 2007. (オンライン), 入手先 <<http://uksg.org/sites/uksg.org/files/FinalReportUsageFactorProject.pdf>>, (参照 2009-03-19).
- 18) 小野寺夏生. 電子ジャーナル閲覧データに基づく論文利用の研究：海外文献紹介. 薬学図書館. 52(3), 2007, 288-295.
- 19) Leydesdorff, L. How are new citation-based journal indicators adding to the bibliometric toolbox?. J Am Soc Inf Sci Technol. 2009. (オ

- ンライン), 入手先 <<http://www3.interscience.wiley.com/journal/121675950/abstract>>, (参照 2009-03-19).
- 20) Bornmann, L. et al. Do we need the  $h$  index and its variants besides standard bibliometric measures?. *J Am Soc Inf Sci Technol.* 2009. (オンライン), 入手先 <<http://www3.interscience.wiley.com/journal/121675951/abstract>>, (参照 2009-03-19).
- 21) Bollen, J. et al. A principal component analysis of 39 scientific impact measures. 2009. (オンライン), 入手先 <<http://arxiv.org/abs/0902.2183>>, (参照 2009-03-19).
- 22) Davis, P. M. Eigenfactor : Does the principle of repeated improvement result in better estimates than raw citation counts?. *J Am Soc Inf Sci Technol.* 59(13), 2008, 2186-2188.
- 23) 緑川信之ほか. 理工学諸分野の雑誌構造. *Library and Information Science.* 20, 1982, 63-80.
- 24) Davis, P. M. "Scientific impact measures compared". *The Scholarly Kitchen.* (オンライン), 入手先 <<http://scholarlykitchen.sspnet.org/2009/02/17/scientific-impact-measures-compared/>>, (参照 2009-03-19).

(原稿受付け: 2009.3.20)