2018年度第1回J-STAGEセミナー
平成30年9月14日(金)
科学技術振興機構 東京本部別館1階ホール

ジャーナル・研究評価の指標について

国立情報学研究所
孫　媛

アウトライン

1. 研究評価とは
2. ビブリオメトリックス指標
   - 各指標の概要
   - 指標を使用する際に配慮すべき点，指標としての限界点
3. オルトメトリックス指標
   - オルトメトリックス指標の概要
   - 指標を使用する際に配慮すべき点，指標としての限界点
4. おわりに

1. 研究評価とは

研究評価とは

- 研究活動に関わる何らかの意思決定を行うために，評価対象の価値を判断する作業
- 評価対象
  - 国／大学・研究機関／研究者個人
  - 研究プロジェクト／研究に関わる政策
  - 論文／ジャーナル

何のための研究評価

- 研究評価の本来の目的
  - 優れた研究の探索
  - 研究の学問的質の維持
- 1940年代NIH；1950年代NSFが設立
- 1980年代，90年代に加速したトレンド
  - 科学政策と連動：研究優先度の定義・研究資金の配分
  - 典型例：
    - 大学ランキング（機関）
    - 研究費配分（プロジェクト，個人，政策）
    - 昇進・給与（個人）

研究評価手法の変遷

- 1970年代までのスタンダードな手法：

  ピアレビュー（peer review）

  - 評価対象と研究分野を同じくする専門家が判断を下す同業者評価
  - 科学コミュニティにおいて様々なレベルで行われる
    - 論文の査読，研究者への学術賞の授与，資金配分を行うプロジェクトの選定，人事採用，昇進…

研究評価手法の変遷

- 1980年代以降，客観的評価指標への要望が高まる
  - 研究活動の拡大と国の研究予算増加の鈍化
  - 評価による選択的資金配分
    - プロジェクト間，分野間比較の必要性
    - 専門内でのピアレビューには限界
  - 説明責任(accountability)の強調
    - 研究活動への公的資金投入の正当性が問われる
    - 投入資金に見合う効果が得られているか（パフォーマンス指標）
  - ピアレビューの問題点
    - 評価者の主観性・能力による意識的・無意識的バイアス
    - 多大の時間・労力・費用を要する
- ビブリオメトリックス指標による研究評価が増え，ビブリオメトリックスが評価ツールとして定着していく

研究評価手法の変遷

- 21 世紀に入り，情報の電子化，モビリティに大きな変化が起き，研究評価にも影響を与えている
- 研究成果は，従来の学術雑誌のほかOA ジャーナルに発表されることが増え，ResearchGate，Academia.edu，blog，Twitter，Facebook 等のソーシャルネットワークサービス（SNS），オンラインツール・環境を通じてアクセスされるようになっている
- 従来は得られなかった情報が手に入り，研究者コミュニティだけでなく一般社会からのアクセスも増えている
- 新しい研究評価の方法の探索が始まっている

2. ビブリオメトリックス指標

ビブリオメトリックスとは

- 図書などの情報伝達媒体(communication media)に対する数学と統計的方法の適用（ Alan Pritchard ， 1969）
  - 文献データの統計分析手法だが，科学コミュニケーションの解明・測定にも用いられ，Scientometricsと同義語で使われることが多い（科学文献データの統計分析は，ビブリオメトリックスという言葉の出現以前から）
- 研究対象は情報伝達媒体
  - 学術雑誌論文，図書，モノグラフ，レポート，学位論文，特許，ウェブデータ等。最近では，ソーシャルメディアデータ等も
- 分析データは書誌情報
  - 雑誌，論文，著者，著者の所属機関，引用文献，出版年，主題分野分類等

研究評価へのビブリオメトリックス指標の利用

- 論文数
  - 研究の生産性，活発さを表す
  - グレードの高い雑誌への掲載は，（ある程度）「質」も保証
- 被引用数
  - 研究の質（インパクト・影響度）を表す
- 基本指標（論文数，被引用数）をもとに多様な指標：
  - 雑誌評価指標：JIF; Citescore; SNIP; SJR; h5-index; h5-median
  - h-index
  - …
- 評価対象
  - 国／雑誌／大学・機関／個人

基本指標１：論文数

- 研究の生産性，活発さを表す指標として用いられる。最大の長所は，単純でわかりやすいこと
- 論文数だけで「質」を判断するのはおかしいという認識から，引用指標の併記が求められることも
  - 学術雑誌，特に「質」が高いとされる雑誌に掲載されるまでに，何人ものピアレビューを経ているはずであり，論文数はある程度「質」を保証するという考え方もある。
- 「論文」：原著，レビュー，…？（∉図書，会議録…）
  - Editorials, Discussions, Meeting abstracts, Book reviews, News items, Letters or Commentaries
- 「論文数」： WoS, Scopus, GoogleScholar…
- とくに国・機関評価の場合，以下の指標の利用も
  - 研究者1人あたりの論文数
  - 高引用論文数；良い雑誌に掲載の論文数
  - 加重；重み付けなし；（共著の場合のカウント法）

基本指標2：被引用数

- 論文が発表されてから現在までに他の論文から引用された回数
- 当該研究領域における被引用論文の貢献を示す
- 被引用数は論文の質に関するほとんど唯一の客観的データと考えられている

- 性質：研究が発表されてから引用されるまでタイムラグがある
  引用統計を取る期間の問題

- 国・機関評価の場合，以下の指標の利用も
  - 研究者1人あたりの被引用数
  - 1論文あたりの被引用数

- 研究評価のための指標が，ビブリオメトリックスの領域でいろいろと開発・提案されているが，そのほとんどは論文数と被引用数をもとに構成されている

インパクトファクター

- JIF (Journal Impact Factor)は，特定の1年間において，その雑誌に掲載された論文が、平均的に何回引用されているか。
  雑誌の重要度を示す指標
- ISI社（現クラリベイト・アナリティクス社）の創設者E. Garfield氏により考案されたもので，同社の引用データ評価・分析ツールJournal Citation Reports (JCR)にて，公表される
- わかりやすく便利な指標として，研究者個人，研究機関の評価指標に転用される
  - ➢IFの提唱者自身も，IFは雑誌の重要度を示す指標であり，研究者の業績評価の指標として利用すべきではないと繰り返し注意を喚起

インパクトファクター算出方法

雑誌Aの2017 年IF＝

$$\frac{\text{2015 年と2016 年に雑誌Aに掲載された論文が 2017 年に全採録対象誌において引用された回数}}{\text{2015 年と2016 年に雑誌Aに掲載された論文数}}$$

例：Scientometrics 2017年 $\mathrm{JIF} = \dfrac{\text{Citations in 2017 to items published in } \mathbf{2015\ (794) + 2016\ (725)}}{\text{Number of citable items in } \mathbf{2015\ (348) + 2016\ (351)}} = \dfrac{1{,}519}{699}$

「論文数」利用時の留意点

① 代表的なデータベース（Web of Science, Scopus）の収録対象は大半が英文誌であるため，カウント対象は英語で発表された論文に大きく偏る。英語以外の言語で発表されることが多い人文社会科学や応用研究は，優れた研究であっても正当に評価されない恐れがある

② 何を「論文」とするか。論文のタイプ

③ 分野によって雑誌論文よりも図書やカンファレンス等が重視されることがある

④ 共著論文では著者へのクレジットの与え方が一意ではない
- 共著者全員に1 編を与える整数カウント（full counting）
- 著者数に応じて案分する分数カウント（fractional counting）
  - 整数カウントだと研究評価の「水増し」や“artificial collaborations”（貢献しない著者への“土産”）を助長する可能性があり，近年は個人業績評価で第1著者とcorresponding author だけをカウントすることが増えている
  - しかし，分野の慣習や研究者間の関係によっては，これが最適だとはかぎらず，整数カウントは共同研究を過小評価するという指摘もある。一方，案分カウントの場合，その案分法は自明ではない

「被引用数」利用時の留意点

- 被引用数を研究評価に用いる場合には，さらに以下のような問題点も指摘されている

① 研究が発表されてから引用されるまでタイムラグがあり，新しい論文に不利

② 研究分野により，引用慣行や引用パターンが異なる（例えば，生物・医学分野では1 論文あたりの引用数は多いが引用期間は短い；数学では引用数は少ないが引用期間は長い）

③ 論文タイプ（原著よりレビューの方が引用されやすい）や自己引用のように，研究の質と関係しない要因によって，被引用数が増えることがある

- ➢引用統計をさまざまな属性（出版年，分野，ドキュメントタイプ等）で正規化
  各種の正規化評価指標が提案されている
  - Mean normalized citation score（MNCS）
  - Normalized mean citation rate （NMCR）
  - Relative citation rate excluding self-citations（RCRX）…

- 技術的な問題
  - 論文数・被引用数を正確に求めること自体，膨大な手間を要する作業
  - いわゆる「名寄せ」の問題：著者の記述通りの情報がデータベースに収録されることが多いが，
    - 引用文献の表記法，雑誌名の省略法の違い，書誌情報記載の誤りなどにより，同一の論文が別のものとしてカウントされることがある
    - 姓と名のイニシャルの英文表記だけでは，著者を同定できないことがある
    - 所属機関名英文表記のゆれ・略名，組織改組などによる機関名の変化

インパクトファクター利用時の留意点

- IF は被引用数に基づく指標であり，被引用数について先述した問題点は，そのままIF の問題点，限界でもある

① 専門誌よりもScience やNature のような総合誌の方が高い値が出やすい

② レビュー論文の比率が高い雑誌で高い値が出やすい

③ IF の算出には過去2 年間の論文への引用を用いるため，分野の特徴や研究のテーマによる有利・不利が生じる

④ IF はWeb of Science 採録雑誌の範囲内で計算されることにも注意が必要である。収録対象外となる各国の学会誌や多くの論文が電子出版に投稿される計算機科学や工学分野では，実態よりも低い評価がされる

⑤ 論文の被引用数の分布は，一般的に右に大きく歪んでいる（対数正規分布）。被引用回数が抜きん出て高い少数の論文が存在するとき平均は大きく影響されるため，平均を用いるのは適切ではないとの指摘も多い

他の雑誌評価指標

- Clarivate Analystics Metrics
  - (JIF), 5-year-JIF
  - Immediacy Index, Cited Half-Life, and Citation Distribution
- Scopus Metrics
  - **CiteScore**
  - SCImago Journal Rank (SJR); Source Normalized Impact per Paper (SNIP)
- Google Scholar Metrics
  - **h-5 Index, h-5 Median**
- Eigenfactor score, Normalized Eigenfactor
- Journal Usage Factors(JUF)
- …

CiteScore (Scopus)

2016年12月8日，エルゼビア社は330分野22,200以上の学術雑誌の引用頻度と影響度を示した新評価指標“CiteScore”を発表

Scopus収録期間: 1869 年から現在まで
出版社: Springer Nature
ISSN: 0028-0836 E-ISSN: 1476-4687
分野: Multidisciplinary

すべての文献を表示 文献アラートを設定 ジャーナルホームページ ROIS Find Full Text ROIS Find Full Text

SJR 2017
17.875

SNIP 2017
8.524

CiteScore CiteScoreランクとトレンド Scopus収録期間

CiteScore 2017 算出に使用したデータの日付: 30 April, 2018

14.59 = 被引用数 2017 / 文献数 2014 - 2016* = 被引用数 114,639 回 > / 文献数 7,860 件 >

*CiteScoreはすべての文献タイプを含みます CiteScore算出方法 > CiteScore FAQ >

CiteScoreランク

分野	ランク	パーセンタイル
Multidisciplinary └ Multidisciplinary	#2/87	98

CiteScoreトレンド > CiteScoreをサイトに追加

CiteScoreTracker 2018 最終更新日 09 August, 2018 毎月更新

8.61 = 被引用数 2018 / 文献数 2015 - 2017 = 現在までの被引用数 67,727 > / 現在までの文献数 7,868 >

CiteScore 2017 の算出方法

CiteScore 2017 は、2014、2015、2016 年に出版された文献が 2017 年に引用された回数を、2014、2015、2016 年に出版された文献の数で割ったものです。

CiteScoreTracker 2018 は、同じ算出方法を使用し、最新の{0}年のデータに基づいて被引用数をトラッキングします。

頻度

3年の出版ウィンドウ

文献タイプ

すべての文献タイプ (research articles、review articles、conference proceedings、editorials errata、letters、notes、short surveys) が CiteScoreの算出に含まれます。Articles in Press はScopusに収録されていますが、CiteScoreの算出には含まれません。

Google Scholar Metrics

- 雑誌の言語や分野ごとにh-indexを提供
- h5-index:
  - 過去5年分の当該雑誌の掲載論文数とそれらの被引用数に基づいて算出される雑誌h-index
- h5-median:
  - h5-indexの集計に用いられた論文の被引用数の中央値

h-index

- 研究者の生涯業績を示す指標として，物理学者Hirschが2005年に提案（Hirsch index）
- ある研究者が発表した論文のうち，少なくともh回以上引用された論文がh編あるとき，その研究者のh指数はh
  - 被引用数10以上の論文を10編持つ研究者のh指数は10
  - 被引用数が10以上の論文が9編ならばh指数は9
- 高いh指数を得るためには，高い質の（引用度の高い）論文を数多く発表しなければならない。
  - 量と質の両立が求められる

h-indexの長所

- 論文数を稼ぐために，完成度の低い論文を発表したり，内容を複数の論文に分けて発表する場合があると言われるが，h指数による評価を導入すれば，このような風潮に歯止めがかかることが期待できる
- 論文数の少ない研究者にとって厳しい評価法に見えるが，h指数は生涯かけて減ることがなく，質の高い論文を発表し続ければ，いずれ確実に高い評価が得られる。研究者の研究意欲を長期的に刺激し続けるのではないか
- 研究者が発表した各論文の被引用回数がわかっていれば直ちに求まること，頑健性があること（質の低い論文の数を重ねても，質が高くても生産性がきわめて低いときには，h指数の値にあまり影響しない）

h-indexの短所

- 若い研究者には不利
- 一定以上の実績を積んだ研究者は高い評価を保持できる
- トップレベルの研究者の評価に向くが一般研究者を識別する力は弱い
- 業績の変化に敏感ではないことも場合によっては短所になりうる
- 研究分野によって引用慣行が大きく異なるため，分野を超えてh指数を比較することは難しい。
  - たとえば，Hirsch(2005)は，h指数の高い研究者の例として，h=191の生命科学者S.H.Snyder，h=132の物理学者Edward Witten，h=64の心理学者Endel Tulvingを挙げているが，彼らの業績をh指数で比べてもあまり意味がない
  - さらに参考までに，2005年の時点でノーベル物理学賞受賞者のh指数平均は41（中央値は35），ワシントン大学（セントルイス）の心理学教授12人のh指数平均は28.1であるという

Categories ▾ English ▾

	Publication	h5-index	h5-median
1.	Nature	362	542
2.	The New England Journal of Medicine	358	602
3.	Science	345	497
4.	The Lancet	278	417
5.	Chemical Society reviews	256	366
6.	Cell	244	366
7.	Nature Communications	240	318
8.	Chemical Reviews	239	373
9.	Journal of the American Chemical Society	236	309
10.	Advanced Materials	235	336
11.	Proceedings of the National Academy of Sciences	226	291

Japanese ▾

	Publication	h5-index	h5-median
1.	情報処理学会論文誌	14	18
2.	電気学会論文誌Ｄ（産業応用部門誌）	12	14
3.	土木学会論文集 B1 (水工学)	10	12
4.	電気学会論文誌 B (電力・エネルギー部門誌)	10	12
5.	人工知能学会論文誌	9	14
6.	日本公衆衛生雑誌	9	11
7.	胃と腸	9	11
8.	情報管理	9	10
9.	日本消化器内視鏡学会雑誌	8	13
10.	火山	8	12
11.	理学療法科学	8	10
12.	教育心理学研究	8	9
13.	日本教育工学会論文誌	8	9

ビブリオメトリックス指標を用いた研究評価（まとめ）

- 被引用数利用のメリット
  - 論文の質（インパクト）を測る上で現時点では最も優れている
  - 研究者自らの意思に基づく
- デメリット，注意点
  - 分野間の違いが大きい
    - 引用の習慣，引用年数，主な成果とするもの
  - 引用のタイムラグ問題
    - 論文が出版されてから引用されるまで時間を要する

    ⇒すぐに評価できない⇒IFを使ってしまう
  - 引用分布の歪み（性質）
    - 平均値を用いる問題
  - 引用種類の違い(自己引用・批判的等)　・収録範囲の問題

3. オルトメトリックス指標

オルトメトリックスの背景

- 情報の電子化，モダリティに大きな変化
- ジャーナルの電子化・OA化による多次元リンク
- SNS利用による研究の利用・アクセス方法の変化
  - ➢研究者コミュニティだけでなく，一般社会からも
- アカウンタビリティ視点の重視
  - 公的資金を用いた研究は国民のために役立っているのか
  - 科学的インパクト以外に，経済・社会・教育的なインパクト
- ピアレビューの限界と出版後評価へ
  - 出版論文数の増加・研究分野の細分化・早期出版の必要性

⇒新しい研究評価方法の探索

オルトメトリックス（Altmetrics）

- Altmetrics：alternative とmetrics からの造語
- 学術論文などの研究成果物の影響度を
  - ① ソーシャルメディアの反応を中心に，定量的に測定する
  - ② その測定手法を用いて新しい研究の影響度を測定・評価し，被引用数を代替・補完する

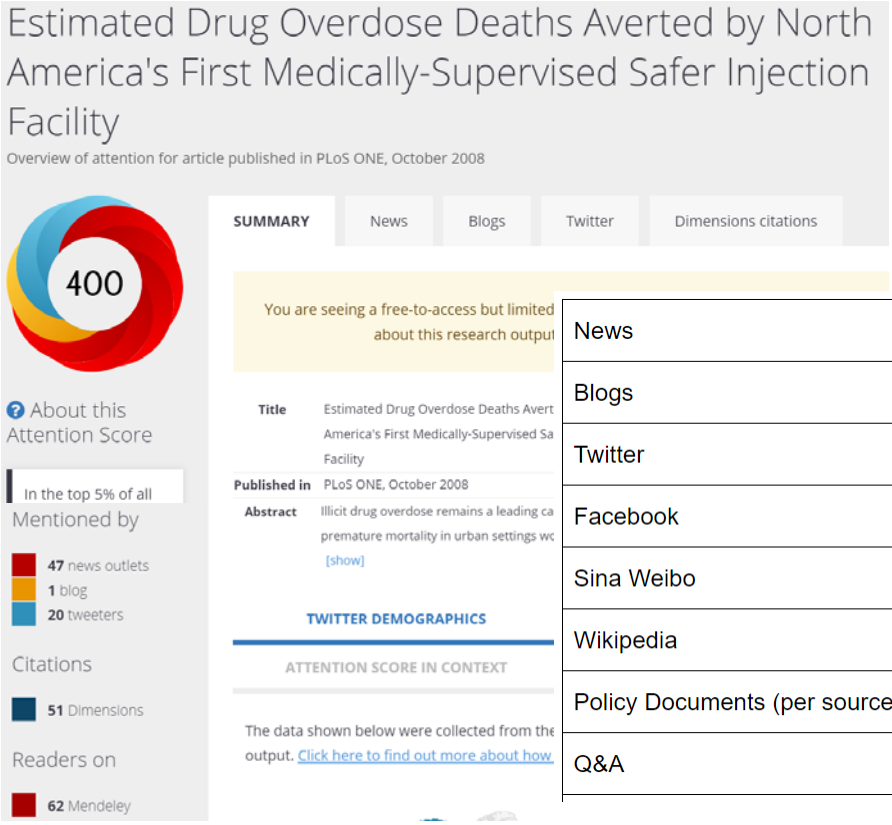
オルトメトリックス指標とは

- オルトメトリックス指標は，ソーシャルメディア上での論文の閲覧や引用，言及などの「イベント（event）」に基づいて構成される。具体的には
  - MendeleyやCiteULike 等文献管理ツールへの保存数
  - ダウンロード数
  - ウェブリンクやブックマークされた回数
  - ブログやTwitter・Facebook 等で取り上げられた数
  - F1000 での推薦数
  - …
- 特徴：研究者コミュニティ内だけではなく，より広範な社会一般への影響度を追跡できること

オルトメトリックスのデータソースの分類

	一般の利用者	学術分野の利用者
推薦	ウェブ上の主流メディア	Faculty of 1000
引用	ウィキペディア	学術誌からの引用
会話	Twitter，Facebook，ブログ	研究者によるブログ，コメント，ツイートなど
参照	ソーシャルブックマーク	ソーシャルリファレンスマネージャー
読み	ページの閲覧	PDFのダウンロード

（出典：Priem, 2014）



Altmetric.com社の
オルトメトリックドーナッツと呼ばれる論文単位で様々なソーシャルメディア上で論文や研究データがどれだけ注目を集めたかを測定するAltmetric Attention Score指標

News	8
Blogs	5
Twitter	1
Facebook	0.25
Sina Weibo	1
Wikipedia	3
Policy Documents (per source)	3
Q&A	0.25

F1000/Publons/Pubpeer	1
YouTube	0.25
Reddit/Pinterest	0.25
LinkedIn	0.5
Open Syllabus	1
Google+	1
Patents	3

オルトメトリックス指標の特徴

- 即時性：論文公表直後から，注目度や影響度をリアルタイムに測定でき，今後の予測にすばやく役立てられる
- 社会性：学術領域への影響にかぎらず，広く社会一般への影響を測定できる
- 補完・代替性：引用で測りにくい雑誌や論文や，データセットやスライド，ブログ投稿といった非伝統的な研究成果にも適用でき，引用・被引用による分析を代替・補完する機能を持つ
- 論文レベルの指標：個別研究のインパクトを測定できる

オルトメトリックスの注意点と課題

- データの多様性
  - Twitter上の言及，F1000上の専門家の推薦，Mendeley上の読者のカウント，Facebook上のいいね，Blog投稿での引用など多種多様なものが含まれており，今後も種類が増えると予想される
  - これらプラットフォームはそれぞれ目的と機能が異なり，生成される指標の意味も異なる
  - 各々の指標の意味を理解するために，利用者の属性や利用目的・動機を明らかする研究が求められている

オルトメトリックスの注意点と課題

- 概念的枠組みの欠如
  - 研究の影響度と何らかの関係があって収集も容易であれば，どんなものでもオルトメトリックス指標になるのが現状
  - 新しいデータソースが現れれば新たな指標が追加され，反対にデータソースとなるオンラインツールが消えれば指標も消える
  - 一般社会への影響を捉えられることがオルトメトリックスの特徴の一つであるが，社会的影響度を定義する理論的な枠組みは未だ存在しない

オルトメトリックスの注意点と課題

- データの質の問題
  - 提供者毎にデータが不均一である
  - オルトメトリックスを測定するツール間で結果が一貫しない

  ⇒偏りのないデータ収集，新たな指標を用いる対象や目的など，オルトメトリックスの標準化への取組が米国情報標準化機構（NISO）によって行われている（www.niso.org/tompics/tl/altmetrics_initiative）
- 依存性の問題
  - オルトメトリックス指標は，データ収集方法，データアグリゲータ，とくにプロバイダーに依存

  ⇒オルトメトリックスに関連する活動の持続可能性・運用安定性も研究評価指標として用いる際に課題になる

オルトメトリックスの注意点と課題

- データの操作可能性
  - オルトメトリックス指標はソーシャルメディアに基づくため，一般受けしやすい記事に反応が偏る可能性や操作が容易であることも大きな問題
- 近年，引用統計に基づく伝統的な既存の指標とオルトメトリックス指標の相関研究やオルトメトリックスの内容分析研究，オルトメトリックスのデータソースとなる行動に対するいわゆる”citer motivation”または”citer behavior”として知られる creator feedback studies等，ビブリオメトリックスの研究手法を適用したオルトメトリックス研究が多く行われている
- オルトメトリックスが研究評価の指標としての地位を確立するためには，まだ多くの研究を重ねる必要がある

4. おわりに

- JIF等のビブリオメトリックス指標をジャーナル評価に用いるとき，さまざまな限界と問題点に留意
- ジャーナルの電子化・OA化により，論文の被引用数以外の多面的なインパクト測定が可能となった
- 現在多くの出版社が研究論文を中心とした研究成果にオルトメトリックスを採用。ジャーナルのトピックや内容に対する社会一般からの関心・注目度を即時に把握できる
- ただし，注目度(attention)≠研究の質(quality)
被引用数が不要になるわけではない。両者を用いる総合的視点が大事
- 現時点でオルトメトリックス指標の有効性を断定的に語るのは難しい
- 指標を用いるとき，専門家によるデータ・結果の詳細分析，冷静・客観的な判断が必要

参考文献

- 孫媛，研究評価のための指標：その現状と展望，情報の科学と技術. 2017, 67(4) p.179-184.
- 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, no.134, p.20–29.
- Karolinska Institutet. Bibliometric handbook for Karolinska Institutet, 2014, 39p. https://kib.ki.se/sites/default/files/bibliometric_handbook_2014.pdf
- Todeschini, R., Baccini, A. Handbook of bibliometric indicators: Quantitative tools for studying and evaluating research. 2016, 483p. Wiley-VCH
- IEEE2013. Appropriate Use of Bibliometric Indicators for the Assessment of Journals, Research Proposals, and Individuals.
- Priem, J. Altmetrics. In B. Cronin & C.R. Sugimoto (Eds.), Beyond bibliometrics: harnessing multidimensional indicators of scholarly impact. 2014, p.263–287. Cambridge, MA: MIT Press.