

Zitaten-Statistiken

Die International Mathematical Union (IMU) hat in Kooperation mit dem „International Council of Industrial and Applied Mathematics (ICIAM)“ und dem „Institute of Mathematical Statistics (IMS)“ einen Bericht mit dem Titel *Citation Statistics* herausgegeben, für den das „Joint Committee on Quantitative Assessment of Research“, bestehend aus Robert Adler, John Ewing (Chair) und Peter Taylor verantwortlich zeichnet. Wir drucken im Folgenden zunächst das „Executive Summary“ dieses Berichts ab und geben anschließend einen Überblick über einige der wichtigsten Argumente und Ergebnisse des Berichts. Die darin wiedergegebenen Tabellen und Grafiken sind dem Bericht entnommen, wir danken den Autoren für die Genehmigung des Abdrucks des Executive Summary und dieser Tabellen und Grafiken. Soweit wir den Bericht in Übersetzung zitieren, handelt es sich nicht um eine autorisierte Übersetzung.¹

Executive Summary

This is a report about the use and misuse of citation data in the assessment of scientific research. The idea that research assessment must be done using “simple and objective” methods is increasingly prevalent today. The “simple and objective” methods are broadly interpreted as *bibliometrics*, that is, citation data and the statistics derived from them. There is a belief that citation statistics are inherently more accurate because they substitute simple numbers for complex judgments, and hence overcome the possible subjectivity of peer review. But this belief is unfounded.

- Relying on statistics is not more accurate when the statistics are improperly used. Indeed, statistics can mislead when they are misapplied or misunderstood. Much of modern bibliometrics seems to rely on experience and intuition about the interpretation and validity of citation statistics.
- While numbers appear to be “objective”, their objectivity can be illusory. The meaning of a citation can be even more subjective than peer review. Because this subjectivity is

less obvious for citations, those who use citation data are less likely to understand their limitations.

- The sole reliance on citation data provides at best an incomplete and often shallow understanding of research – an understanding that is valid only when reinforced by other judgments. *Numbers are not inherently superior to sound judgments.*

Using citation data to assess research ultimately means using citation-based statistics to rank things – journals, papers, people, programs, and disciplines. The statistical tools used to rank these things are often misunderstood and misused.

- For journals, the impact Factor is most often used for ranking. This is a simple average derived from the distribution of citations for a collection of articles in the journal. The average captures only a small amount of information about that distribution, and it is a rather crude statistic. In addition, there are many confounding factors when judging journals by citations, and any comparison of journals requires caution when using impact Factors. Using the impact Factor alone to judge a journal is like using weight alone to judge a person's health.
- For papers, instead of relying on the actual count of citations to compare individual papers, people frequently substitute the impact factor of the journals in which the papers appear. They believe that higher impact Factors must mean higher citation counts. But this is often *not* the case! This is a pervasive misuse of statistics that needs to be challenged whenever and wherever it occurs.
- For individual scientists, complete citation records can be difficult to compare. As a consequence, there have been attempts to find simple statistics that capture the full complexity of a scientist's citation record with a single number. The most notable of these is the h-index, which seems to be gaining in popularity. But even a casual inspection of the h-index and its variants shows that these are naïve attempts to understand complicated citation records. While they capture a small amount of information about the dis-

¹ Der vollständige Bericht ist im Internet unter der folgenden Adresse zugänglich: www.mathunion.org/publications/report/citationstatistics/

tribution of a scientist's citations, they lose crucial information that is essential for the assessment of research.

The validity of statistics such as the impact Factor and h-index is neither well understood nor well studied. The connection of these statistics with research quality is sometimes established on the basis of "experience." The justification for relying on them is that they are "readily available". The few studies of these statistics that were done focused narrowly on showing a correlation with some other measure of quality rather than on determining how one can best derive useful information from citation data.

We do not dismiss citation statistics as a tool for assessing the quality of research – citation data and statistics can provide some valuable information. We recognize that assessment must be practical, and for this reason easily-derived citation statistics almost surely will be part of the process. But citation data provide only a limited and incomplete view of research quality, and the statistics derived from citation data are sometimes poorly understood and misused. Research is too important to measure its value with only a single coarse tool.

We hope those involved in assessment will read both the commentary and the details of this report in order to understand not only the limitations of citation statistics but also how better to use them. If we set high standards for the conduct of science, surely we should set equally high standards for assessing its quality.

Joint IMU/ICIAM/IMS-Committee on
Quantitative Assessment of Research
Robert Adler, *Technion – Israel Institute of Technology*
John Ewing (Chair), *American Mathematical Society*
Peter Taylor, *University of Melbourne*

Citation statistics – Argumente und Ergebnisse

Statistiken über Häufigkeit und Gewicht von Zitaten eines wissenschaftlichen Artikels in anderen wissenschaftlichen Artikeln werden in Politik und Verwaltung in zunehmendem Maße als Entscheidungshilfe, mitunter auch als wesentliches oder gar alleiniges Entscheidungskriterium, herangezogen, und zwar bei Entscheidungen über die Vergabe von Haushalts- und Fördermitteln, über Abonnements wissenschaftlicher Zeitschriften durch Bibliotheken und über Einstellung und Bezahlung von wissenschaftlichem Personal.

Es wird dabei häufig angenommen, dass sich mit statistischen Methoden Kennziffern ableiten lassen, die eine vergleichende Beurteilung

der Qualität von wissenschaftlichen Zeitschriften, von Forschungsabteilungen oder von einzelnen WissenschaftlerInnen ermöglichen.

Dem Bericht folgend geben wir zunächst einen Überblick über einige der Verfahren zur Ermittlung solcher Kennziffern und diskutieren dann die folgenden Fragen:

- Was wird durch diese Kennziffern gemessen?
- Wie zuverlässig sind diese Messungen?
- Welche Aspekte von Qualität werden durch diese Messungen erfasst?

Wir beschränken uns dabei aus Platzgründen auf einige Teilaspekte des Berichts und empfehlen interessierten LeserInnen die Lektüre des vollständigen Berichts.

I Kennziffern und deren Ermittlung

Der Bericht stellt einige der am häufigsten benutzten Kennziffern vor.

(a) *Der Impact Factor.* Der Impact Factor einer Zeitschrift wird von Thomson Scientific, dem Verleger der *Journal Citation Reports* (früher ISI) so ermittelt: Sind in Zeitschrift Z in den beiden Jahren $x-1, x-2$ insgesamt a Artikel erschienen und lassen sich im Jahr x in allen von Thomson Scientific erfassten Zeitschriften insgesamt b Zitate von in Z in den beiden Jahren $x-1, x-2$ erschienen Artikeln nachweisen, so hat Z (im Jahr x) den Impact factor $\frac{a}{b}$. Er misst also die durchschnittliche Anzahl von Zitaten von in Z in den beiden Vorjahren erschienen Artikeln, wobei zur Berechnung in der Mathematik und ihren Randgebieten eine Basis von ca. 400 von Thomson Scientific ausgewählten Zeitschriften dient (das ist weniger als die Hälfte der von Math Reviews bzw. Zentralblatt erfassten Zeitschriften). Die Auswahl erfolgt wiederum auf Grund von Zitierhäufigkeiten. Neuerdings werden kompliziertere Algorithmen zur Ermittlung des Impact Factors vorgeschlagen, etwa solche, die den Gebrauch eines page rank-Algorithmus involvieren (page rank-Algorithmen werden z. B. von Google zur Sortierung der Suchergebnisse verwendet).

(b) *Der h-Index einer Person.* Der h-Index, vorgeschlagen von J. E. Hirsch, einer Person ist das Maximum aller n , für die diese Person n Artikel veröffentlicht hat, die jeweils wenigstens n -mal zitiert wurden.

(c) *Der m-Index einer Person.* Der m-Index einer Person, ebenfalls vorgeschlagen von Hirsch, ist der Quotient aus dem h-Index und der Anzahl der Jahre seit der ersten Publikation dieser Person.

(d) *Der g-Index einer Person.* Der g-Index, vorgeschlagen von Leo Egge, einer Person ist das

Maximum aller n , für die diese Person n Artikel veröffentlicht hat, für die die Gesamtsumme aller Zitate $\geq n^2$ ist.

2 Der Gebrauch der Kennziffern

Der Bericht gibt Beispiele für den Gebrauch des Impact Factors, die auf Berichte von Mathematikern aus verschiedenen (im Bericht nicht genannten) Ländern zurückgehen:

- Meine Universität hat eine Klassifikation von Zeitschriften eingeführt, die auf den *Science Citation Index Core Journals* basiert. Die Zeitschriften werden nach der Größe des Impact factors in drei Gruppen eingeteilt. Die erste Gruppe besteht aus 30 Zeitschriften, darunter keine mathematische. Die zweite Gruppe besteht aus 667 Zeitschriften, darunter 21 mathematische. Publikation in einer Zeitschrift der ersten Gruppe verdreifacht die Forschungsgelder, Publikation in einer Zeitschrift der zweiten Gruppe verdoppelt sie. Ferner erhält man für Publikation in einer von Thomson Scientific im Index geführten Zeitschrift 10 Punkte, für Publikation in einer Zeitschrift aus Thomsons *core list* besonders wichtiger Zeitschriften erhält man 15 Punkte. Für Beförderungen (promotions) ist eine bestimmte Mindestpunktzahl erforderlich.
- In meinem Land werden Lehrende in Dauerstellen an Hochschulen alle sechs Jahre evaluiert. Einer der wichtigsten Faktoren ist dabei ein ranking von fünf Publikationen. Eine Publikation im obersten Drittel der (nach Impact Factor geordneten) Liste von Thomson Scientific bringt drei Punkte, eine Publikation im zweiten Drittel bringt zwei Punkte, eine Publikation im untersten Drittel bringt einen Punkt.
- In meinem Department wird zur Evaluation der Lehrenden die mit dem Impact Factor der jeweiligen Zeitschrift gewichtete Anzahl der zu einer Arbeit mit nur einem Autor äquivalenten Arbeiten herangezogen. Beförderungen und Einstellungen basieren teilweise auf dieser Zahl.

Thomson Scientific schlägt die Benutzung des Impact Factors zudem als Instrument für die Entscheidung vor, welche Zeitschriften eine Bibliothek bezieht, zusammen mit Kosten und Verbreitungszahlen der Zeitschriften. Der Bericht enthält hierzu keine Beispiele.

3 Kritikpunkte

Der Bericht setzt sich kritisch mit der Frage auseinander, ob Kennziffern, wie sie oben beschrieben wurden, grundsätzlich ein geeignetes Instrument zur Beurteilung der Qualität von

wissenschaftlichen Zeitschriften oder wissenschaftlich tätigen Personen sein können.

Generell wird von Kennziffern erwartet, dass sie das subjektive (und daher fehlerbehaftete und insbesondere für Vorurteile anfällige) System der peer reviews, also der gegenseitigen Begutachtung, durch objektive Messgrößen ersetzen. Der Bericht kritisiert, dass der Ermittlung und Verwendung dieser Messgrößen kein klares Modell zu Grunde liegt:

Die wichtigste Kritik am Impact Factor ist, dass man seine Bedeutung nicht gut versteht. Benutzt man den Impact Factor, um zwei Zeitschriften zu vergleichen, so gibt es kein *a priori*-Modell, das definiert, was es heißt, „besser“ zu sein. Das einzige Modell ist vom Impact Factor selbst abgeleitet – je größer der Impact factor ist, desto besser ist die Zeitschrift. Im klassischen statistischen Paradigma definiert man ein Modell, formuliert eine Hypothese und findet dann eine Statistik, die es erlaubt, in Abhängigkeit von ihrem Wert die Hypothese zu akzeptieren oder zu verwerfen. Information (und möglicherweise ein Modell) aus den Daten selbst herzuleiten, ist ein legitimer Ansatz in der statistischen Analyse, aber in diesem Fall ist nicht klar, welche Information hergeleitet wurde. Wie misst der Impact Factor Qualität? Ist er die beste Statistik, um Qualität zu messen? Was genau misst er? (Die unten folgende Diskussion der Bedeutung von Zitaten ist hier relevant.) Man weiß bemerkenswert wenig über ein Modell für Qualität von Zeitschriften oder darüber, wie sie mit dem Impact Factor zusammenhängt.

Insbesondere spricht der Bericht die Frage an, was mögliche Motive für ein Zitat in einer wissenschaftlichen Arbeit sind:

Es gibt eine reichhaltige Literatur zur Frage der wirklichen Bedeutung von Zitaten, die nahelegt, dass Zitierungen komplizierter sind, als diese vagen Aussagen glauben machen. Zum Beispiel schreiben Martin und Irvine 1983 in ihrem Aufsatz über Bewertung von Forschung: „All diesen Problemen mit dem Gebrauch von Zitaten als Maß von Qualität liegt unsere Unkenntnis der Gründe, aus denen Autoren bestimmte Arbeiten zitieren und andere nicht, zu Grunde ... Einfache Zitatenanalyse nimmt ein hochgradig rationales Modell des Angebens von Referenzen an, in dem man annimmt, dass Zitate in erster Linie die wissenschaftliche Wertschätzung früherer Arbeiten von hoher Qualität oder großer Wichtigkeit ausdrücken, und dass potentielle Zitierer eine bestimmte Arbeit alle mit der gleichen Wahrscheinlichkeit zitieren werden.“

Cozzens behauptet in ihrer 1989 erschienenen Arbeit [Cozzens 1989] über die Bedeutung von Zitaten, dass Zitate das Resultat von zwei verschiedenen Systemen sind, die dem wissenschaftlichen Publizieren zu Grunde liegen. Das eine ist ein Belohnungs-System (reward system), das andere ein rhetorisches. Zitate des erstgenannten Systems haben die Bedeutung,

The Ostrowski Prize 2009

Call for proposals

The aim of the Ostrowski Foundation is to promote the science of mathematics by periodically awarding an international prize for the best performances in the field of pure mathematics and of the theoretical foundations of numerical mathematics. As a rule, the prize is awarded every two years to the scientist, or group of scientists, who, during the preceding five years, has achieved the highest scientific accomplishments in these fields. It is awarded independently of politics, race, religion, domicile, nationality or age. The prize in 2007 amounted to 100 000 Swiss francs.

The Foundation awards at the same time a scholarship for a talented young mathematician, whose name is to be suggested by the current prizewin-

ners. The scholarship will enable the winner to spend a year for further education (as a postdoctoral fellow) at a university of his or her own choice.

The previous prizewinners are, in chronological order: L. de Branges, J. Bourgain, M. Ratner and M. Laczkovich, A. Wiles, Y. Nesterenko and G. Pisier, A. Beilinson and H. Hofer, H. Iwaniec, P. Sarnak and R. Taylor, P. Seymour, B. Green and T. Tao, O. Schramm.

The jury invites proposals for candidates for the Ostrowski Prize 2009. The proposals including a short justification should be sent to David. Masser@unibas.ch before 1st December 2008.

die meistens mit Zitaten assoziiert wird – die Anerkennung, dass die zitierende Arbeit der zitierten Arbeit gegenüber intellektuell in der Schuld steht. Zitate des zweiten Systems haben dagegen eine gänzlich andere Bedeutung – eine Referenz zu einer früheren Arbeit, die ein Ergebnis erklärt, das vielleicht gar nicht von diesem Autor stammt. Solche rhetorischen Zitate sind einfach nur eine Art, die wissenschaftliche Konversation zu pflegen, nicht aber intellektuelle Schuld festzustellen. Natürlich kann ein Zitat in manchen Fällen auch beide Bedeutungen haben.

...

Cozzens beobachtet, dass die meisten Zitate rhetorisch sind. Die Erfahrung der meisten Mathematiker bestätigt das. (Zum Beispiel sind 30% der über drei Millionen Zitate in der Datenbank der Math Reviews Referenzen zu Büchern und nicht zu Originalartikeln in wissenschaftlichen Zeitschriften.) Warum ist das wichtig? Weil beim rhetorischen Zitieren, im Gegensatz zu Belohnungs (reward)-Zitaten, die dazu tendieren, sich auf einflussreiche Forschungsartikel zu beziehen, die Auswahl der zu zitierenden Arbeiten von mehreren Faktoren abhängt – dem Prestige des zitierten Autors (der „halo“-Effekt), den Beziehungen zwischen zitierendem und zitiertem Autor, der Zugänglichkeit der Zeitschrift (haben Open-Access-Zeitschriften eine höhere Zitierwahrscheinlichkeit?), der Tatsache, dass es bequem ist, mehrere Resultate aus einer einzigen Arbeit zu zitieren, und so weiter. Wenige dieser Faktoren hängen direkt mit der Qualität der zitierten Arbeit zusammen.

Selbst wenn Zitate Belohnungs-Zitate sind, können sie eine Vielzahl von Motiven widerspiegeln, unter anderem (Originaltext) „currency, negative credit, operational Information, persuasiveness, positive credit, reader alert, and social consensus“ [Brooks 1996]. In den meisten Fällen erfolgen Zitate aus mehr als einem dieser Motive. Manche wichtigen Ergeb-

nisse können dem Verdeckungs-Effekt (obliteration effect) zum Opfer fallen, indem sie sofort in die Arbeiten anderer integriert werden, die dann als Basis weiterer Zitate dienen. Andere Zitate sind keine Belohnungen für hervorragende Forschung sondern Warnungen vor fehlerhaften Resultaten oder Denkfehlern. Der Bericht selbst liefert etliche Beispiele solcher Warnungs-Zitate.

Der Bericht setzt sich dann mit dem Argument auseinander, dass die Frage der Bedeutung von Zitaten außer Acht gelassen werden könne, da Zitaten-Statistiken stark mit den Ergebnissen anderer Methoden zur Messung von Forschungsqualität korreliert seien. Er zitiert dazu den „Evidence report 2007“ der Firma Evidence Ltd, die die britische Regierung bei der Forschungsevaluation berät:

Evidence has argued that bibliometric techniques can create indicators of research quality that are congruent with researcher perception. [Evidence Report 2007, p. 9]

und antwortet:

Die Schlussfolgerung ist anscheinend, dass auf Zitaten basierende Statistiken, unabhängig von deren genauer Bedeutung, andere Bewertungsmethoden ersetzen sollten, da sie oft mit ihnen übereinstimmen.

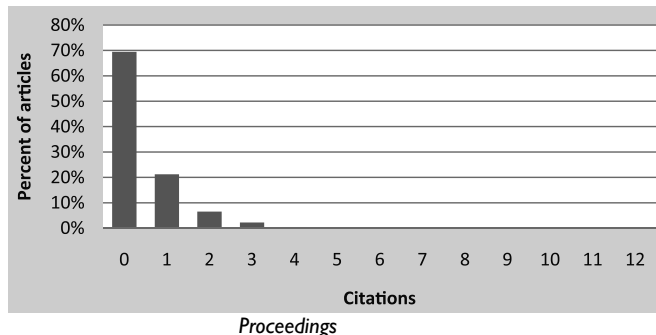
Auch abgesehen von der Kreisschlüssigkeit dieses Arguments ist die Fehlerhaftigkeit einer derartigen Begründung leicht zu sehen.

Der Bericht macht aber auch darauf aufmerksam, dass der unvorsichtige Gebrauch der Kennziffern selbst unter der Annahme, dass sie grundsätzlich zur Messung von Qualität geeignet seien, leicht zu Fehlschlüssen führen kann:

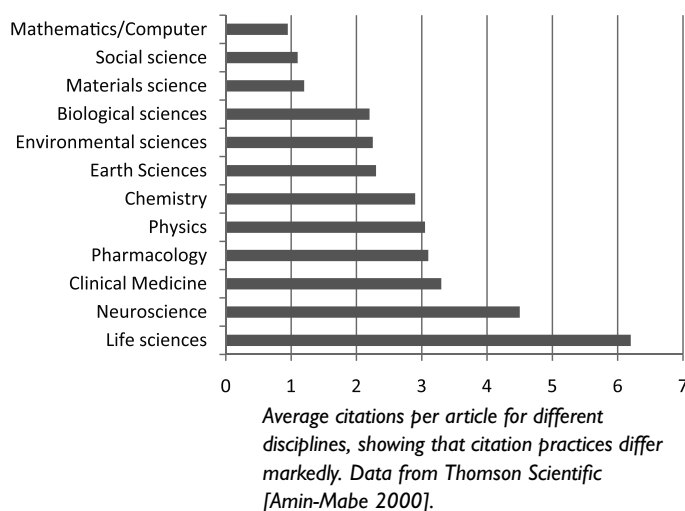
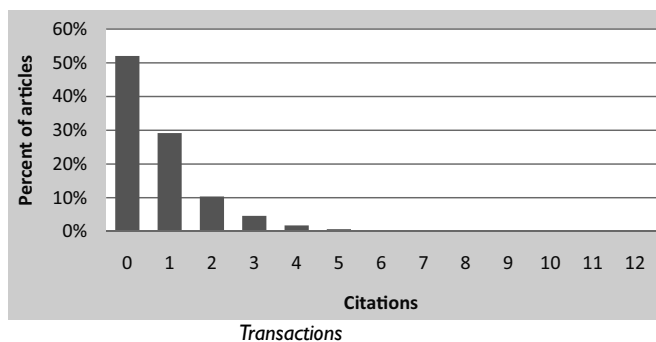
Es ist seit langem bekannt, dass die Verteilung der Zitierhäufigkeiten für eine einzelne Arbeit in einer Zeitschrift hochgradig schief (skewed)

ist und ein Potenzgesetz approximiert. Das hat Folgen, die an einem Beispiel präzisiert werden können:

Im folgenden Diagramm sieht man die Verteilung für Arbeiten in den *Proceedings of the American Mathematical Society* im Zeitraum 2000–2004. Die *Proceedings* publizieren kurze Arbeiten, normalerweise unter zehn Seiten Umfang. Im genannten Zeitraum publizierten sie 2381 Arbeiten (ungefähr 15 000 Seiten). Betrachtet man die Zeitschriften von 2005 in der Datenbank der Math Reviews so erhält man einen Impact Factor von 0.434, also durchschnittlich 0,434 Zitate pro Artikel.



Die *Transactions of the AMS* veröffentlichen längere Artikel, die in der Regel substantieller sind, sowohl von der Anlage her als auch vom Inhalt. Im gleichen Zeitraum publizierten die *Transactions* 1165 Artikel (mehr als 25 000 Seiten) mit Zitierhäufigkeiten zwischen 0 und 12. Die durchschnittliche Anzahl von Zitaten je Artikel war 0,846, also etwa doppelt so hoch wie bei den *Proceedings*.



Wir betrachten jetzt zwei Mathematiker, von denen der eine eine Arbeit in den *Proceedings* und der andere eine Arbeit in den *Transactions* publiziert. Benutzt man einige der Methoden von Institutionen, die wir oben erwähnt haben, so wird der zweite als dem ersten überlegen beurteilt, da er eine Arbeit in einer Zeitschrift mit einem höheren (sogar doppelt so hohen) Impact Factor publiziert. Ist das eine gültige Bewertung? Sind Arbeiten in den *Transactions of the AMS* doppelt so gut wie solche in den *Proceedings*?

Wenn wir aussagen wollen, dass eine einzelne Arbeit in den *Transactions* besser (im Sinne von Zitaten) ist als eine einzelne Arbeit in den *Proceedings*, so dürfen wir keine Frage über Durchschnitte stellen, sondern eine über Wahrscheinlichkeiten: Was ist die Wahrscheinlichkeit dafür, dass eine zufällig ausgewählte Arbeit in den *Proceedings* wenigstens genauso viele Zitate hat, wie eine zufällig ausgewählte Arbeit aus den *Transactions*?

Das ist eine elementare Rechnung, und die Antwort ist 62%. Das heißt, dass wir in 62% der Fälle falsch liegen, und dass eine zufällig ausgewählte Arbeit aus den *Proceedings* genauso gut wie (oder besser als) eine zufällig ausgewählte Arbeit aus den *Transactions* sein wird – obwohl der Impact Factor der *Proceedings* nur die Hälfte dessen der *Transactions* ist! Die meisten Leute finden das überraschend, aber es ist eine Konsequenz der hochgradig schiefen Verteilung und des engen Zeitfensters, das zur Berechnung des Impact Factor benutzt wird (letzteres ist der Grund dafür, dass ein hoher Prozentsatz aller Arbeiten überhaupt keine Zitate aufweist). Das zeigt den Wert von exaktem statistischem Denken im Gegensatz zu intuitiver Beobachtung.

Dieses Verhalten ist typisch für Zeitschriften, und die Auswahl dieser beiden Zeitschriften ist nicht speziell. (Das *Journal of the AMS* etwa hat im gleichen Zeitraum einen Impact Factor von 2,63 – sechsmal so hoch wie der der *Proceedings*. Dennoch ist ein zufällig gewählter Artikel in den *Proceedings* in 32% der Fälle wenigstens so gut (im Sinne von Zitaten) wie ein Artikel im *Journal*.)

Während es falsch wäre, zu sagen, dass der Impact Factor keine Information über einzelne Arbeiten in einer Zeitschrift gibt, ist diese Information daher überraschend vage und kann dramatisch irreführend sein.

Weitere wichtige derartige „technische“ Kritikpunkte des Berichts sind die folgenden:

- Der Impact Factor variiert erheblich zwischen verschiedenen Fächern ... Wenn in einem Fach viele Zitate außerhalb des zweijährigen Zeitfensters erfolgen, werden seine Zeitschriften einen sehr viel niedrigeren Impact Factor haben. Auf der anderen Seite ist ein Teil des Unterschiedes einfach, dass die Zitierkulturen verschiedener Fächer unterschiedlich sind, und dass Wissenschaftler Arbeiten mit unterschiedlicher Häufigkeit und aus unterschiedlichen Gründen zitieren ...

Es folgt, dass man Zeitschriften aus verschiedenen Fächern nicht in sinnvoller Weise mit Hilfe des Impact Factors vergleichen kann.

- Der Impact Factor kann von Jahr zu Jahr erheblich schwanken, und die Schwankung tendiert dazu, für kleinere Zeitschriften größer zu sein [Amin-Mabe 2000]. Für Zeitschriften, die weniger als 50 Artikel publizieren war zum Beispiel die durchschnittliche Änderung im Impact Factor zwischen 2002 und 2003 nahezu 50 %. Das ist natürlich zu erwarten, da die Stichprobengröße für kleine Zeitschriften klein ist. Andererseits vergleicht man oft Zeitschriften in einem gegebenen Jahr, ohne die größere Schwankung für kleine Zeitschriften in Rechnung zu stellen.
- Zeitschriften, die Artikel in anderen Sprachen als Englisch publizieren, werden wahrscheinlich seltener zitiert, da ein großer Teil der community sie nicht lesen kann (oder nicht liest). Auch der Typ der Zeitschrift und nicht alleine ihre Qualität kann den Impact Factor beeinflussen: Zeitschriften, die Übersichtsartikel publizieren zum Beispiel bekommen oft sehr viel mehr Zitate als Zeitschriften, die das nicht tun, und können deshalb höhere (mitunter wesentlich höhere) Impact Factors haben. [Amin-Mabe 2000]

Wir lassen es bei diesen Ausschnitten aus dem Bericht bewenden und fordern zum Abschluss noch einmal dazu auf, den gesamten Bericht im Internet zu lesen und die Argumente gegen das schematische Anwenden von Zitierstatistiken im Kollegenkreis und gegenüber Verwaltungen und Politik zu verbreiten.

Referenzen

Amin, M.; Mabe, M. 2000. Impact factor: use and abuse. *Perspectives in Publishing*, No. 1, October, pp. 1–6.

Brooks, Terrence. 1986. Evidence of complex citer motivations. *Journal of the American Society for Information Science*, Vol 37, No. 1, pp. 34–36, 1986.

Cozzens, Susan E. 1989. What do citations count? The rhetoric-first model. *Scientometrics*, Vol 15, Nos 5–6, (1989), pp. 437–447.

Evidence Report. 2007. The use of bibliometrics to measure research quality in the UK higher education system. (A report produced for the Research Policy Committee of Universities, UK, by Evidence Ltd., a company specializing in research performance analysis and interpretation. Evidence Ltd. has „strategic alliance“ with Thomson Scientific.)

Übersetzung: Rainer Schulze-Pillot