

Bibliometrische Verfahren zur Bewertung von Forschungsleistung

Hintergründe, Aussagekraft und Grenzen

Von Werner Marx und Lutz Bornmann

Zusammenfassung: Das klassische Gutachterverfahren zur Bewertung von Forschungsleistung wird zunehmend durch Zitationsanalysen zur Messung der Wirkung von Publikationen ergänzt und unterstützt. Die quantitativen Verfahren der Wirkungsmessung sind jedoch nur dann aussagekräftig, wenn sie die jeweils geeigneten Indikatoren heranziehen. Der Trend hin zu einer Amateur-Bibliometrie stellt eine ernste Gefahr für die Forschungsevaluierung dar. Der Wunsch der Wissenschaftspolitik nach Rankings und Zahlen wird durch scheinbar einfach zu bedienende Werkzeuge erfüllt. Wir plädieren in diesem Artikel für den Einsatz einer professionellen Bibliometrie, die in den Kontext des Forschungsgebiets Bibliometrie eingebettet ist. Danach ist z.B. der Journal Impact Factor (JIF) zur Messung der Wirkung der Arbeiten eines Forschers ungeeignet, da er keinen Rückschluss auf die Wirkung einer bestimmten in einer Zeitschrift publizierten Arbeit erlaubt. Der h-Index misst Produktivität und Wirkung in Form von Zitierungen in einer einzigen Zahl und verknüpft beide Größen willkürlich. Perzentile (Prozentränge) sind dagegen für die Bewertung von Publikationen und Personen sehr gut geeignet. Die Arbeiten eines Forschers und deren Wirkung, gemessen anhand von Perzentilen, kann als sogenannter Beam-Plot übersichtlich dargestellt werden. Mit Beam-Plots kann man auf einen Blick feststellen, in welchen Jahren ein Forscher wie produktiv war und wie häufig seine Arbeiten im Vergleich zu ähnlichen Publikationen zitiert wurden.

1. Einleitung

Die Ergebnisse wissenschaftlicher Aktivität werden heute zumeist als Artikel in den entsprechenden Fachzeitschriften veröffentlicht. Je nach Fachdisziplin spielen aber auch Bücher als Publikationsmedium eine wichtige Rolle (wie z.B. in der Soziologie). Auf diese Weise werden die Forschungsergebnisse den Fachkollegen vorgestellt und können von diesen diskutiert, kritisiert oder als Ausgangsbasis für deren eigene Forschungsarbeiten herangezogen werden. Die Bezugnahme auf vorangegangene Arbeiten erfolgt typischerweise in der Form von Fußnoten bzw. Verweisen (auch Referenzen, Zitate, Zitationen oder Zitierungen genannt), welche in alphabetischer oder nummerierter Auflistung den Publikationen im Anhang beifügt werden. Damit kann der Leser die Bezugnahme auf vorangegangene Arbeiten nachvollziehen und Ursprung und Entwicklung von Methoden, Theorien usw. nachverfolgen. Die Zitierungen sind die Fäden im Netzwerk der wissenschaftlichen Arbeiten, welche diese inhaltlich verbinden. Für die Einschätzung der Bedeutung von Zitierungen ist es wichtig festzuhalten, dass diese Daten von den forschenden und publizierenden Autoren selbst und nicht etwa von Datenbankherstellern oder Statistikern stammen (Bornmann / Marx 2014 a). Wird ein Manuskript bei einer Zeitschrift eingereicht, unterliegen auch die Zitierungen einer Kontrolle seitens der Redaktion und der Gutachter (hinsichtlich ihrer Vollständigkeit, ihrer Richtigkeit und ihres Formats).

Zitationsanalysen sind erst dadurch möglich geworden, dass beim Aufbau und der Entwicklung von Literaturdatenbanken nicht nur Publikationen, sondern auch deren Zitierungen untereinander integriert und zur Auswertung bereitgestellt wurden (Garfield 1979). Vorreiter war das Institute for Scientific Information (ISI, heute Thomson Reuters) in Philadelphia (USA), das unter der Leitung von Eugene Garfield im Jahr 1980 zunächst den Science Citation Index (SCI) und später weitere Zitierrindexe (wie z.B. den Social Science Citation Index, SSCI) in elektronischer Form zur Verfügung stellte. Diese werden seit 1997 als Web

of Science (WoS) angeboten und dienen seitdem als Datengrundlage für eine ganze Reihe von anderen Datenbanken (wie z.B. InCites von Thomson Reuters) zur Bewertung der Publikationsleistung (Output) und der Wirkung (Impact) von Forschern, Institutionen und Ländern. Inzwischen sind die Datenbanken Scopus von Elsevier¹ und Google Scholar² ernstzunehmende Konkurrenten des WoS geworden, die je nach Absicht und Fachgebiet alternativ oder zusätzlich bei einer Zitationsanalyse herangezogen werden können. Diese Datenbanken weisen eine Reihe von spezifischen Vor- und Nachteilen auf, die sich insbesondere auf die Funktionalität und auf Umfang und Zeitraum der Erfassung der Fachliteratur in den diversen Disziplinen beziehen.

Zitierindexe wurden ursprünglich als Hilfsmittel zur Auffindung von Fachinformation und nicht zur Bewertung von Forschungsleistung erfunden. Arbeiten, die von anderen Arbeiten zitiert werden, behandeln zumeist gleiche oder ähnliche Themen und können somit ohne die problematische Suche über Schlagworte als inhaltlich verwandte Arbeiten identifiziert werden (Garfield 1979). Inzwischen steht jedoch bei der Nutzung die Heranziehung von Zitierungsdaten zur bibliometrischen Bewertung von Forschungsleistung im Vordergrund (Moed 2005; Abbott et al. 2010). Diese Entwicklung resultiert auch aus Schwächen des klassischen Gutachterwesens (dem sogenannten Peer Review Verfahren), durch das beispielsweise die Qualifikation von Stellenbewerbern oder die Forschungsleistung von Universitäten beurteilt wird. Das Peer Review Verfahren basiert auf der Heranziehung von erfahrenen Fachkollegen (Peers), die als „Wächter der Wissenschaft“ die Qualität der Forschung gewährleisten sollen.

Gutachter im Peer Review Verfahren unterliegen der Gefahr, die Bedeutung von Forschung nicht zu erkennen oder sachfremde Einflüsse in die Bewertung von Forschern und Institutionen einfließen zu lassen. Wenn es um die Einschätzung von Fachkollegen geht, sind Wissenschaftler nicht immer objektiv – und manchmal überfordert. Gutachtergremien setzen sich häufig aus nur wenigen Experten zusammen, was die Gefahr der Subjektivität erhöht. Deshalb wurden und werden zusätzlich objektive und nachvollziehbare Kriterien der Bewertung verlangt. Eine auf Zitierungen basierende Methode bietet zwei Vorteile: Sie beruht (1) auf einem größeren Blickwinkel, indem sie das Votum vieler Fachkollegen (statt weniger Gutachter) heranzieht, und (2) die Ergebnisse sind überprüfbar. Entwickelt und verfeinert wird diese Methode in einer Fachdisziplin, die sich weltweit unter dem Namen Bibliometrie mit inzwischen eigenen Lehrstühlen, Tagungen und Zeitschriften etabliert hat (Abbott 2010). Als Königsweg der Forschungsevaluation gilt das Informed Peer Review, bei dem Gutachter mit Hilfe von bibliometrischen Berichten über die Güte von Forschung urteilen.

2. Wirkung vs. Qualität

Die Aussagekraft von Zitierungen als Messlatte zur Bewertung von Forschungsleistung ist ein ausgiebig diskutiertes Thema geworden. Eine Debatte um die Frage, inwieweit Wirkung (Impact) und Qualität korrelieren, wird seit der Erfindung der Zitierindexe geführt. Aus der Anzahl der Zitierungen ergibt sich nicht ohne weiteres ein Maß für die Qualität bzw. den Wert einer Arbeit. Mit Qualität können ganz unterschiedliche Aspekte gemeint sein: Eleganz, Originalität, Wichtigkeit, Richtigkeit, aber auch Popularität oder gar Verwertbarkeit. Die Zuschreibung dieser Aspekte ist aber subjektiver Art und wesentlich durch normative Vorstellungen geprägt. Qualität ist nicht eindeutig definiert und zuschreibbar; und dementsprechend ist sie auch nicht ohne weiteres messbar. Zitierungen dokumentieren zunächst

1 Abrufbar unter: <http://www.elsevier.com/online-tools/scopus>, letztes Abrufdatum: 9.6.2015.

2 Abrufbar unter: <http://scholar.google.de/>, letztes Abrufdatum: 9.6.2015.

lediglich die Wirkung in Form der Beachtung durch die Fachkollegen und sind keine Werturteile (im Sinne eines „besser“ oder „schlechter“). Einige Untersuchungen deuten jedoch darauf hin, dass Wirkung (gemessen anhand von Zitierungen) und Bedeutung bzw. Wert (gemessen anhand der Einschätzung von Fachkollegen) korrelieren (Diekmann / Näf / Schubiger 2012). Indem Zitierungszahlen die Wirkung als einen wichtigen Teilaspekt von Qualität quantifizieren, können sie durchaus als Stellvertreter-(Proxy-)Daten für die Bewertung von Qualität aufgefasst und eingesetzt werden (Moed 2005).

Die bibliometrische Bewertung von Forschung setzt hinreichend große Ensembles von zu untersuchenden Arbeiten voraus: Auf der Makroebene (ganze Länder) sind die Ergebnisse bibliometrischer Analysen statistisch am besten abgesichert, und auch auf der Mesoebene (Universitäten, Forschungsinstitute, Forschungsgruppen) ist die Aussagekraft der Methode hoch. Die Bewertung auf der Mikroebene (Personen, einzelne Arbeiten) erfordert dagegen eine vorsichtige Interpretation der Ergebnisse. Personenbezogene Rankings sind vielfach problematisch bis irreführend. Im Falle einzelner Arbeiten ist die Anzahl der Zitierungen durchaus nicht immer proportional zur Bedeutung der Arbeit (eine starke Wirkung ist eine allenfalls notwendige, aber keinesfalls eine hinreichende Voraussetzung für Qualität). Manche Arbeiten gelten in Fachkreisen als Pionierarbeiten, doch sie wurden aus diversen Gründen verhältnismäßig selten zitiert. So wurde beispielsweise eine Arbeit von Karl Jansky, welche 1933 in *Nature* erschienen ist und das Forschungsfeld Radioastronomie begründet hat, bisher lediglich 20 Mal zitiert (Jansky 1933). Die Aussagekraft von Peer Review verhält sich eher umgekehrt als die der Bibliometrie: Sie ist am größten auf der Mikroebene (sofern die Gutachter einzelne Arbeiten wirklich lesen) und am geringsten auf der Makroebene. Ein Gutachter kann zwar einzelne Arbeiten bewerten, aber kaum behaupten, eine Institution oder gar ein Land sei „gut“ oder „schlecht“ (ohne dafür Indikatoren heranzuziehen).

3. Journal Impact Factor

Die bibliometrische Forschung hat inzwischen eine Vielzahl zitierungsbasierter Indikatoren vorgeschlagen, die je nach Bedarf und Verfügbarkeit der Daten zur Forschungsbewertung herangezogen werden. Bibliometrisch Unkundige setzen häufig die Verwendung von Zitierungsdaten mit dem Einsatz des sogenannten Impact Factor gleich. Mit dem Impact Factor ist in der Regel der Journal Impact Factor (JIF) gemeint, der den durchschnittlichen Impact der Publikationen in einer Zeitschrift innerhalb eines Jahres angibt. Es handelt sich dabei um einen bibliometrischen Indikator, der ursprünglich zur Einstufung von Zeitschriften als Ganzes erfunden wurde, der seit Jahren aber auch zur Bewertung der einzelnen Arbeiten von Forschern, Institutionen usw. herangezogen wird (Garfield 2006). Dabei geht man von der Prämisse aus, dass eine bestimmte Arbeit, die in einer Zeitschrift mit einem vergleichsweise hohen JIF erschienen ist, eine entsprechend hohe Wirkung erzielen wird bzw. bereits erzielt hat. Der JIF ist inzwischen von einem zunächst wenig bekannten Hilfsmittel für Bibliothekare und Datenbankhersteller zum wohl bekanntesten bibliometrischen Indikator in der Forschungsevaluation geworden.

Die alljährlich aktualisierten fachspezifischen Ranglisten, die auf der Grundlage des JIF erzeugt werden, berücksichtigen derzeit rund 8500 Zeitschriften aus Naturwissenschaft, Medizin und Technik sowie rund 3000 Zeitschriften aus den Geistes- und Sozialwissenschaften. In die Ranglisten werden nicht alle gegenwärtig erscheinenden Zeitschriften einbezogen, sondern lediglich eine kleine Auswahl (die sogenannten Kernzeitschriften). Die Ranglisten werden von Thomson Reuters in den Journal Citation Reports (JCR) publiziert.³ Auswahlkriterien für die Einbeziehung von Zeitschriften in die JCR sind vor allem die inter-

3 Abrufbar unter: <http://scientific.thomson.com/products/jcr/>, letztes Abrufdatum: 9.6.2015.

ationale Reputation der Editoren, die internationale Relevanz der fachlichen Ausrichtung einer Zeitschrift, ihr regelmäßiges Erscheinen und ein funktionierendes Qualitätssicherungsverfahren. Eine nicht berücksichtigte Zeitschrift muss jedoch keine schlechte Zeitschrift sein.

Zur Bestimmung der JIFs werden die Publikationen einer Zeitschrift jeweils aus einem Zeitfenster von zwei Jahren herangezogen und deren Zitierungen innerhalb des darauf folgenden Jahres ermittelt. Die Anzahl der Zitierungen wird dann durch die Anzahl der zitierbaren Publikationen (das sind Articles, Reviews und Short Communications) aus den beiden vorangegangenen Jahren dividiert. Etwas vereinfacht ausgedrückt kann man sagen, dass der JIF einer Zeitschrift die mittlere Anzahl der Zitierungen innerhalb eines Jahres von Veröffentlichungen der vorangegangenen zwei Jahre nach deren Publikation angibt. Zusätzlich werden von Thomson Reuters auch JIFs auf der Basis eines fünfjährigen (statt des klassischen zweijährigen) Publikationszeitraums ermittelt. Hier geht es also um die mittlere Anzahl der Zitierungen innerhalb eines Jahres von Veröffentlichungen aus den vorangegangenen fünf Jahren. Damit soll der verhältnismäßig langsamen Rezeption der Arbeiten in manchen Disziplinen (wie z.B. der Soziologie) Rechnung getragen werden, bei denen der kurze Zitationszeitraum zu einer kaum zuverlässigen Impact-Messung führt. Publikationen aus dynamischen Gebieten (wie z.B. einigen Biowissenschaften) erhalten einen verhältnismäßig hohen Anteil ihrer Zitierungen bereits in den ersten Jahren nach der Publikation und sind durch die JIF-Formel begünstigt.

Der Unterschied in der Rezeption von Publikationen in der Soziologie im Vergleich zu den Biowissenschaften wird in Abbildung 1 dargestellt. Diese Abbildung zeigt die Zeitentwicklung der Zitierungen der 6059 Arbeiten des Jahrgangs 2000 in der WoS-Fachkategorie *Sociology* im Vergleich zu derjenigen in der Fachkategorie *Biochemistry & Molecular Biology* (letztere bezieht sich nur auf Arbeiten, die auch der Fachkategorie *Biology* zugeordnet wurden: 8895 Arbeiten). Die Zitierungen der Arbeiten in der Soziologie erreichen erst nach zehn Jahren ihr Maximum, während diejenigen der Arbeiten in der Biochemie/Biologie bereits nach drei Jahren abklingen.

Abbildung 1: Zeitentwicklung der Zitierungen der 6059 Arbeiten des Jahrgangs 2000 in der WoS-Fachkategorie *Sociology* im Vergleich zu derjenigen in der Fachkategorie *Biochemistry & Molecular Biology* (letztere bezieht sich nur auf Arbeiten, die auch der Fachkategorie *Biology* zugeordnet wurden: 8895 Arbeiten)

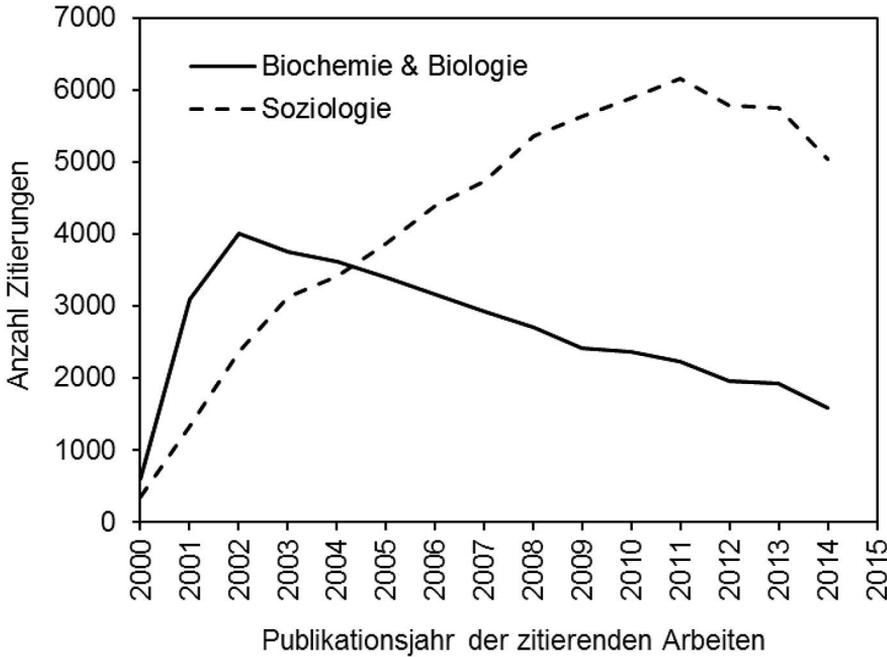


Tabelle 1 zeigt die JIFs einer führenden Zeitschrift der Soziologie (*American Journal of Sociology*, *Am. J. Soc.*) im Vergleich zu denjenigen einer führenden Zeitschrift der Biochemie/Biologie (*Cell*) über die Jahre 2007 bis 2012, jeweils auf Basis eines zweijährigen und eines fünfjährigen Publikationszeitfensters. Das *American Journal of Sociology* erreicht mit einem Zeitfenster von 5 Jahren deutliche höhere JIFs, während die JIFs von *Cell* nahezu unabhängig vom Zeitfenster sind.

Tabelle 1: JIFs einer führenden Zeitschrift der Soziologie (American Journal of Sociology) im Vergleich zu denjenigen einer führenden Zeitschrift der Biochemie / Biologie (Cell) über die Jahre 2007 bis 2012 bzw. 2013, jeweils auf Basis eines zweijährigen und eines fünfjährigen Publikationszeitfensters

JIF-Jahrgang	<i>Cell</i>		<i>Am. J. Soc.</i>	
	JIF-2Jahre	JIF-5Jahre	JIF-2Jahre	JIF-5Jahre
2007	29,887	28,779	3,338	5,113
2008	31,253	30,149	2,808	5,046
2009	31,152	32,628	3,476	5,411
2010	32,406	34,931	3,358	5,113
2011	32,403	34,774	3,169	4,561
2012	31,957	34,366	3,414	5,239

Wie die Beispiele zur Soziologie und Biochemie zeigen, zeichnen sich die verschiedenen Disziplinen durch ganz unterschiedliche JIFs und Zitieraten aus. Dementsprechend können auch nur Zeitschriften, Forscher und Institutionen miteinander verglichen werden, die zur gleichen Fachkategorie gehören. Für die Unterschiede zwischen den Disziplinen ist nicht nur das fachspezifische Rezeptionsverhalten, sondern auch die unterschiedliche Abdeckung der fach-spezifischen Literatur in den Datenbanken verantwortlich (siehe auch weiter unten). Während beispielsweise die Literatur aus den Biowissenschaften nahezu vollständig vom WoS erfasst wird, fehlt ein großer Teil der Literatur aus den Sozial- und Geisteswissenschaften (Archambault / Gagné 2004; Archambault et al. 2006). Die Wissenschaftler dieser Disziplinen publizieren nur einen kleinen Anteil ihrer Arbeiten in den von Datenbanken erfassten Kernzeitschriften. Viele ihrer Publikationen erscheinen in Büchern oder in Zeitschriften, die wegen ihrer engen fachlichen Ausrichtung oder ihrer lediglich nationalen Bedeutung nicht berücksichtigt werden. Vergleichsweise niedrige mittlere Zitieraten sind die Folge.

Die Tatsache, dass eine bestimmte Arbeit in einer Zeitschrift mit hohem oder niedrigem JIF erschienen ist, sagt wenig über den Impact dieser speziellen Arbeit aus. Selbst in Zeitschriften mit hohem JIF hat die Masse der Arbeiten keinen ungewöhnlich hohen Impact. Die unsymmetrische bzw. schiefe Verteilung der Zitierungen auf die Publikationen (Pareto-Verteilung) bedeutet, dass nur ein kleiner Anteil hochzitatierter Arbeiten einen großen Anteil aller Zitierungen erhält (Seglen 1992). Die 6059 Arbeiten des Jahrgangs 2000 in der WoS-Fachkategorie *Sociology* beispielsweise erhielten bis Ende 2014 insgesamt 63199 Zitierungen. Davon entfallen alleine 6163 (rund 10%) auf die Top-10 der meistzitierten Arbeiten (rund 1/10 Promille aller Arbeiten). Der arithmetische Mittelwert (und damit der JIF einer Zeitschrift) wird also nur von relativ wenigen Arbeiten bestimmt und erlaubt deshalb keinen Rückschluss auf den Impact einer bestimmten in dieser Zeitschrift publizierten Arbeit (Adler et al. 2009).

Wir haben es hier also mit einer Art von gerichtetem Zusammenhang zu tun: Der Impact von einzelnen Arbeiten verhilft einer Zeitschrift zu einem mehr oder weniger hohen JIF, die Publikation einer Arbeit in einer Zeitschrift mit hohem JIF bedeutet jedoch nicht, dass diese auch automatisch hoch zitiert ist. Deshalb ist es nicht verwunderlich, dass die zwischen den Zitierungen der Arbeiten eines einzelnen Forschers und den JIFs der Zeitschriften dieser Arbeiten untersuchte Korrelation häufig gering ausfällt (Seglen 1997). JIFs sind deshalb zur Messung des Impacts von Arbeiten, die ein Forscher veröffentlicht hat, ungeeignet. Dazu sollte man die Zitierungen jeder einzelnen Publikation recherchieren, auch wenn die Ermittlung und Interpretation dieser Daten mit höherem Aufwand verbunden ist. Der Umweg der Impact-Bestimmung über die JIFs macht wenig Sinn, auch wenn in einzelnen Disziplinen (wie z.B. der Medizin) und Ländern (wie z.B. in skandinavischen Ländern, der Volksrepublik China und Indien) die Forschungsbewertung danach erfolgt und sogar die Mittelvergabe häufig danach bemessen wird. Diese fragwürdige Verfahrensweise wird inzwischen nicht nur von Bibliometrikern, sondern auch von den betroffenen Wissenschaftlern kritisiert. Die Kritik resultierte in Initiativen wie z.B. DORA⁴ und in Publikationen, wie dem Report der International Mathematical Union (IMU) (Adler et al. 2009).

4. h-Index

Neben dem JIF ist der h-Index (auch Hirsch Index oder h Zahl genannt) zu einem weithin bekannten bibliometrischen Indikator geworden, der insbesondere zur Bewertung von Forschern herangezogen wird (Hirsch 2005). Dieser Index hat den Vorteil, dass er über die übli-

4 Abrufbar unter: <http://am.ascb.org/dora/>, letztes Abrufdatum: 9.6.2015.

chen Zitierindexe, wie z.B. den WoS, schnell und einfach ermittelt werden kann. Dazu erfasst man die zu bewertenden Arbeiten eines Forschers und sortiert sie nach der Häufigkeit ihrer Zitierungen. Der h-Index ist nun diejenige Zahl, bei der die Rangnummer mit der Anzahl der Zitierungen übereinstimmt. Ein h-Index von 40 bedeutet beispielsweise, dass der betreffende Forscher 40 Arbeiten verfasst hat, die jeweils mindestens 40 Mal zitiert wurden.

Der h-Index misst den Output (Anzahl der Arbeiten) und den Impact (Anzahl der Zitierungen) in einer einzigen Maßzahl. Dieser Indikator ist inzwischen eine vielgenutzte Messlatte für die Abschätzung des Lebenswerkes eines Forschers bzw. dessen „Gewicht“ unter seinen Fachkollegen. Die Anzahl der Zitierungen pro Arbeit (und damit der h-Index) ist jedoch stark fachabhängig, was Personenvergleiche auf der Basis des h-Index beeinträchtigt. Außerdem ist der h-Index vom Schaffensalter abhängig: Je länger ein Forscher bereits wissenschaftlich aktiv ist, desto mehr Publikationen und Zitierungen kann man für sie/ihn erwarten. Im Hinblick auf das unterschiedliche Schaffensalter ist vorgeschlagen worden, den h-Index durch die Anzahl der wissenschaftlich aktiven Jahre (z.B. vom Jahr der Dissertation bis zur Gegenwart) zu teilen. Man erhält dann den sogenannten m-Quotient, der immerhin den Vergleich unterschiedlich alter Forscher innerhalb der gleichen Fachdisziplin erlaubt. Ein h-Index von 40 basierend auf einem Schaffensalter von 20 Jahren entspräche z.B. einem m-Quotienten von 2,0. Den m-Quotienten hatte bereits Hirsch (2005) vorgeschlagen.

Die leichte Ermittlung des h-Index ist allerdings nur gegeben, wenn die Autorennamen der publizierenden Wissenschaftler eindeutig sind oder durch Kombination mit Adressen und/oder Ortsnamen oder Fachkategorien präzise eingegrenzt werden können. Nicht selten muss ein erheblicher (Recherche)aufwand betrieben werden, um „saubere“ Publikationslisten als Datengrundlage zu erhalten. Viele gelegentliche Benutzer des WoS scheitern jedoch daran oder sind sich des Problems nicht bewusst.

Auch wenn es häufig als ein Vorteil angesehen wird, kann man es auch als einen Nachteil ansehen, dass der h-Index den Publikationsumfang und die Wirkung in Form von Zitierungen in einer einzigen Zahl und nicht getrennt voneinander misst. Es gibt eigentlich keinen Grund, diese beiden Maßzahlen in einer Zahl zusammenzufassen (zumal es sich nur um zwei Maßzahlen handelt). Ein anderes Problem des h-Index besteht darin, dass Output und Impact in der Formel für die Bestimmung des h-Index (also Anzahl Papers mit mindestens h Zitierungen) willkürlich miteinander verknüpft sind und der Indikator durchaus auch anders definiert werden könnte (wie z.B. Anzahl Papers mit mindestens 2 h Zitierungen). Diese und andere Probleme des h-Index haben dazu geführt, dass der Indikator vielfach modifiziert wurde – es gibt inzwischen mehr als 40 Varianten (Bornmann / Marx 2011).

Gegenüber dem h-Index wird vielfach der Einwand vorgebracht, dass man komplexe Systeme grundsätzlich nicht mit einer einzelnen Maßzahl abbilden kann. So würde z.B. ein Arzt die Gesundheit eines Patienten niemals nur nach einem einzelnen Kriterium, wie dem Körpergewicht, bemessen. Auch wenn es in der Medizin leicht bestimmbare Indikatoren gibt (wie z.B. die Höhe der Körpertemperatur), die für sich alleine genommen bereits eine hohe Aussagekraft haben, werden diese Indikatoren immer im Zusammenhang mit anderen Informationen interpretiert. Wie in der Medizin reicht auch in der Forschungsevaluierung ein einzelner bibliometrischer Indikator nicht aus, um eine zuverlässige und umfassende Beurteilung von Forschungsleistung vorzunehmen.

5. Zitierreten

Ohne geeignete Messlatte ist die reine Anzahl der Zitierungen einer Arbeit wenig aussagekräftig. Die Zahl der Zitate ist nicht nur von der Wertschätzung durch die Fachkollegen abhängig, sondern auch von einer ganzen Reihe anderer Faktoren, die mit der fachlichen

Bedeutung einer Arbeit wenig oder nichts zu tun haben (Bornmann / Daniel 2008; Bornmann et al. 2012 a). Bibliometriker wissen seit Langem, dass die Fachdisziplinen in der Wissenschaft sehr unterschiedliche mittlere Zitierraten aufweisen, die bei fachübergreifenden Vergleichen berücksichtigt werden müssen (van Raan 2000). Wie Abbildung 2 exemplarisch für das Publikationsjahr 2000 zeigt, sind die mittleren Zitierraten (Zitierungen pro Arbeit) in den großen Fachdisziplinen (z.B. in den Naturwissenschaften im Vergleich zu den Geistes- und Sozialwissenschaften) sehr unterschiedlich (siehe die grauen Balken in der Abbildung). Während beispielsweise eine Arbeit in den Geisteswissenschaften im Mittel 1,7 Mal zitiert wurde, beträgt die mittlere Zitierrate in den Naturwissenschaften 27. Wenn man die mittleren Zitierraten verschiedener naturwissenschaftlicher Fächer (z.B. in der Chemie oder Physik) untereinander vergleicht, findet man ebenfalls eine Bandbreite von einem Faktor zehn.⁵ Die Zitierraten zwischen den Sub-Disziplinen innerhalb der Fächer variieren ähnlich stark (Neuhaus et al. 2005). Unterschiedliche Zitierraten kann man demnach auf jeder fachlichen Aggregationsebene feststellen.

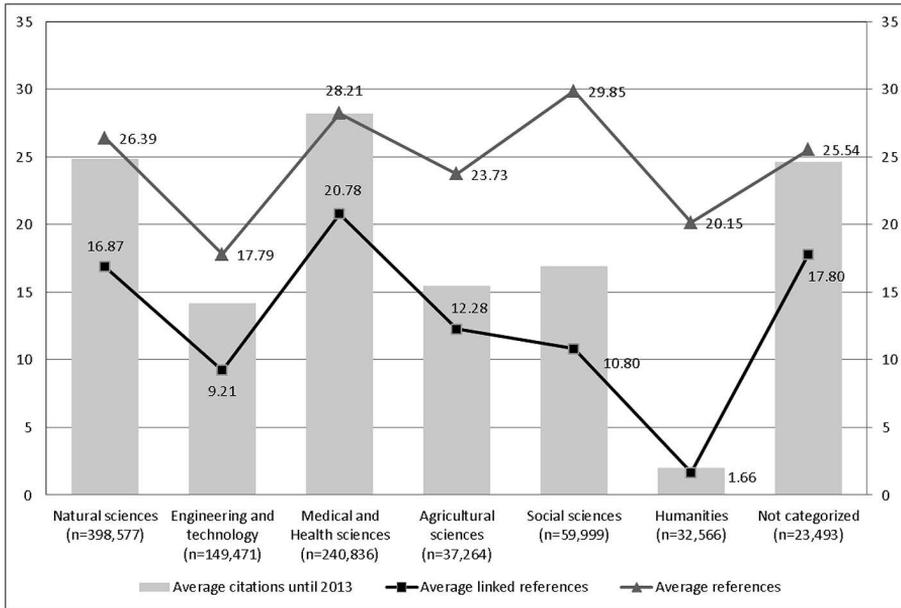
Als Grund für die fachlich deutlich variierende mittlere Anzahl von Zitierungen pro Arbeit wird vielfach die unterschiedliche Zitiergewohnheit in den Disziplinen genannt (Garfield 1979). Die Disziplin-spezifischen Zitierraten würden dementsprechend aus der unterschiedlichen Anzahl von Literaturverweisen (Referenzen) pro Arbeit – und damit der unterschiedlichen Wahrscheinlichkeit, dass eine bestimmte Arbeit zitiert wird – resultieren. In diesem Zusammenhang muss man jedoch zwischen Referenzen differenzieren, welche sich auf Arbeiten beziehen, die in den vom WoS erfassten Kernzeitschriften erschienen sind, und solchen, die sich auf andere Verweise beziehen (das können neben Literaturstellen auch Internet-Links oder Patente sein). Erstere Referenzen beziehen sich auf die sogenannten aktiven bzw. verlinkten Referenzen, letztere sind inaktiv bzw. nicht verlinkt. Die aktiven Referenzen sind, wie die zitierenden Arbeiten, als Datenbank-Dokumente abgespeichert und als solche suchbar.

Abbildung 2 zeigt (neben der mittleren Anzahl der Zitierungen pro Arbeit) die mittlere Anzahl der Referenzen pro Arbeit und die mittlere Anzahl der verlinkten Referenzen pro Arbeit (also denjenigen Anteil der Referenzen, der in Form von Datenbank-Dokumenten vom WoS erfasst wird) in verschiedenen Disziplinen. Die mittlere Anzahl der Referenzen pro Arbeit ist bei allen Disziplinen ähnlich und liegt zwischen rund 20 und 30 Referenzen. Im Mittel werden also in den wissenschaftlichen Disziplinen ähnlich viele Referenzen im Anhang der Arbeiten aufgeführt. Dagegen variiert die mittlere Anzahl der verlinkten Referenzen erheblich: Sie liegt zwischen 1,7 und 20,6 verlinkter Referenzen pro Arbeit. Der Unterschied zwischen der mittleren Anzahl der Referenzen insgesamt und der mittleren Anzahl der verlinkten Referenzen ist im Falle der Geisteswissenschaften am größten; er ist aber auch in den Agrarwissenschaften und den Sozialwissenschaften verhältnismäßig deutlich ausgeprägt.

Wie Abbildung 2 weiter zeigt, korreliert zwar die mittlere Anzahl der Zitierungen pro Arbeit mit der mittleren Anzahl der verlinkten Referenzen; sie hängt aber kaum mit der mittleren Anzahl der Referenzen pro Arbeit insgesamt zusammen. Mit anderen Worten: Die mittlere Zitierrate wird weniger vom Umfang der Referenzen insgesamt bestimmt, sondern vielmehr davon, in welchem Umfang die als Referenzen zitierten Arbeiten durch den WoS erfasst werden. Da die Erfassung in den Geisteswissenschaften besonders niedrig ausfällt, ist auch die mittlere Zitierrate entsprechend niedrig.

5 Abrufbar unter: <http://thomsonreuters.com/essential-science-indicators/>, letztes Abrufdatum: 9.6.2015.

Abbildung 2: Mittlere Anzahl Zitierungen, mittlere Anzahl Referenzen und mittlere Anzahl aktiver (verlinkter) Referenzen von Arbeiten aus unterschiedlichen Disziplinen, die im Jahr 2000 publiziert wurden (Datenquelle: eine in-house Datenbank der Max-Planck Digital Library, die auf Daten aus dem WoS beruht)



Die niedrigen Zitierraten in den Sozial- und Geisteswissenschaften resultieren also keinesfalls aus der Angewohnheit, im Mittel wenige Referenzen in Arbeiten aufzuführen. Ursache ist vielmehr der besonders niedrige Anteil von verlinkten Referenzen, also solchen zitierten Publikationen, die in den Kernzeitschriften des WoS erschienen sind. Die nicht verlinkten Publikationen sind beispielsweise die für die Sozial- und Geisteswissenschaften besonders wichtigen Bücher und Buchkapitel, die von den klassischen Zitierindizes in der Regel nicht oder nur unzureichend erfasst werden. Sozial- und Geisteswissenschaftler zitieren jedoch nicht nur Bücher, sondern vielfach auch Arbeiten, die in Zeitschriften erschienen sind, welche nicht zu den Kernzeitschriften des WoS gehören. Die beschriebenen Unterschiede zwischen der mittleren Anzahl der verlinkten und aller Referenzen (besonders ausgeprägt in den Sozial- und Geisteswissenschaften) deuten darauf hin, dass man im Zusammenhang mit der Fachabhängigkeit von Zitierungen weniger die fach-spezifischen Zitierkulturen bzw. Zitiergewohnheiten der Forscher, als vielmehr den unterschiedlichen Abdeckungs- bzw. Erfassungsgrad der entsprechenden Zitierdatenbank thematisieren sollte.

6. Normierung

In der Liste der Einflüsse, die Zitierungszahlen als Bemessungsgrundlage verzerren können, steht neben dem unterschiedlichen Alter der Publikationen die Fachabhängigkeit der Zitierungen an vorderster Stelle. Deshalb kann beispielsweise ein Mathematiker nicht ohne weiteres mit Forschern aus anderen Fachgebieten verglichen werden. Wie wir oben bereits ausgeführt haben, unterliegen die wohl bekanntesten und am weitesten verbreiteten bibliometrischen Indikatoren, der JIF und der h-Index, dieser Verzerrung. Mit Hilfe dieser Indikatoren können deshalb Zeitschriften bzw. Forscher lediglich unter der Voraussetzung von ver-

gleichbarem Publikationszeitraum bzw. Schaffensalter und gleichem Fachgebiet direkt und ohne Normierung miteinander verglichen werden.

Da bei Personenvergleichen (aber auch bei Institutionen- und Ländervergleichen) gleiches akademisches Alter der Forscher und gleiches Fachgebiet der Publikationen zumeist nicht gegeben sind, benötigt man für einen fairen Vergleich alters- und fachnormierte Daten. Ein normierter Citation Impact (Normalized Citation Score, NCS) kann als Quotient aus (1) der beobachteten Anzahl der Zitierungen einer Arbeit und (2) der mittleren Anzahl der Zitierungen aller Arbeiten aus dem gleichen Fachgebiet und Publikationsjahr (dem Erwartungswert bzw. Referenzset) gemessen werden. Der normierte Citation Impact gibt darüber Auskunft, welche Wirkung eine bestimmte Arbeit im Vergleich zu ähnlichen Arbeiten aus dem gleichen Fachgebiet und Publikationsjahr erzielt hat. Die Zitierungen einer Arbeit, die beispielsweise im Jahr 2000 im *American Journal of Sociology* erschienen ist, werden mit den gemittelten Zitierungen aller *Articles* des Jahrgangs 2000, die in den Zeitschriften der WoS-Fachkategorie *Sociology* (dieser Kategorie hat Thomson Reuters die Zeitschrift zugeordnet) erschienen sind, in Relation gesetzt. Diese Quotienten werden für jede einzelne der zu bewertenden Arbeiten eines Forschers oder einer Forschungseinrichtung berechnet. Schließlich wird ein Mittelwert über die Quotienten gebildet, der das Verhältnis der beobachteten zur erwarteten Anzahl der Zitierungen aller zu bewertenden Arbeiten ausdrückt. Dieser Mittelwert ist der Mean Normalized Citation Score (MNCS) eines Forschers oder einer Einrichtung.

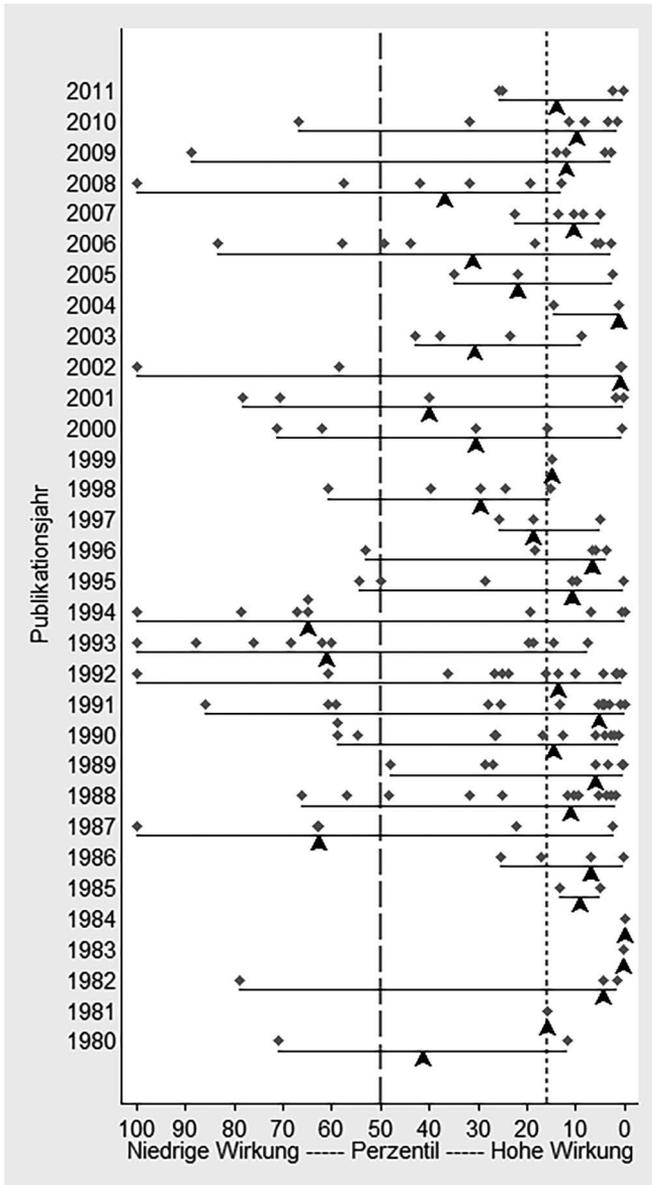
Auch wenn es sich beim MNCS inzwischen um einen Standardindikator in der Bibliometrie handelt, hat er jedoch auch Nachteile. Da der Indikator auf arithmetischen Mittelwerten von Zitierungen beruht und Zitierungen in der Regel schief verteilt sind (siehe oben), können einzelne Publikationen einen (zu) großen Einfluss auf den Wert des MNCS ausüben.

Perzentile (auch Prozentränge genannt) sind besser geeignet, die Zitierhäufigkeiten einzelner Arbeiten im Hinblick auf das Fachgebiet und das Publikationsjahr zu normieren. Das Perzentil einer bestimmten Arbeit gibt analog zum NCS darüber Auskunft, welche Wirkung diese Arbeit im Vergleich zu ähnlichen Arbeiten (im gleichen Fachgebiet und Publikationsjahr) erzielt hat. Im Gegensatz zum NCS sind Perzentile jedoch kaum von der Schiefe der Verteilung der Zitierungen auf die Arbeiten in einem Publikationsset beeinflusst. Das Perzentil bringt die relative Position einer Publikation innerhalb der Arbeiten im Referenzset zum Ausdruck: Die zu untersuchenden Arbeiten werden jeweils in den Wirkungsrahmen ähnlicher Arbeiten des gleichen Fachgebietes und Publikationsjahres innerhalb einer Skala zwischen 0 und 100 eingestuft und damit aussagekräftig bewertet (Bornmann / Marx 2013 a, 2013 b). Ein Perzentil von 10, das nach der Perzentil-Methode von Thomson Reuters (InCites), berechnet wurde, besagt beispielsweise, dass 90% der Publikationen im Referenzset weniger als die betreffende Publikation zitiert wurden.

Die Arbeiten eines Forschers und deren Wirkung – gemessen anhand von Perzentilen – kann als sogenannter Beam-Plot übersichtlich dargestellt werden. Mit Beam-Plots kann man auf einen Blick feststellen, in welchen Jahren ein Forscher wie produktiv war und wie häufig seine Arbeiten im Vergleich zu den Publikationen im Referenzset zitiert wurden (Bornmann / Marx 2014 b, 2014 c; Marx / Bornmann 2013 b, 2013 c). Abbildung 3 zeigt exemplarisch einen solchen Beamplot für einen Forscher. Die Perzentile der einzelnen Arbeiten erscheinen als Rauten, deren Mittelwert (Median) für die einzelnen Publikationsjahre jeweils als Dreieck angezeigt wird. Die Bandbreite der Wirkung der Arbeiten in den einzelnen Publikationsjahren erstreckt sich von links (niedrige Wirkung) nach rechts (hohe Wirkung). Die vertikale grob gestrichelte Linie in der Mitte markiert ein Perzentil von 50 und entspricht einer durchschnittlichen Wirkung in den jeweiligen Fachgebieten und Publikationsjahren. Die vertikale fein gestrichelte Linie rechts gibt den Mittelwert (Median) über alle

Perzentile für den betreffenden Forscher an. Mit einem mittleren Perzentil von rund 15 ist die durchschnittliche Wirkung der Publikationen des Forschers von Abbildung 3 erheblich besser als die zu erwartende Wirkung. Da Perzentile Werte zwischen 0 und 100 annehmen können, beträgt der Erwartungswert 50.

Abbildung 3: Anzahl und Wirkung (gemessen anhand von Perzentilen) der Arbeiten eines Forschers, graphisch dargestellt als Beamplot



7. Exzellenzindikator

Statt auf die Verteilung der Perzentile kann man sich alternativ auf eine bestimmte Perzentilrangklasse konzentrieren: Arbeiten, die zu den Top 10% der meist-zitierten Arbeiten in ihrem Fachgebiet und Publikationsjahr gehören, werden vielfach als hoch zitierte (exzellente) Arbeiten angesehen (für diese Arbeiten liegt – nach InCites – ein Perzentil ≤ 10 vor). Den Anteil dieser Arbeiten kann man deshalb als ein Qualitätsmerkmal für einen Forscher oder eine Institution ansehen. So wird beispielsweise im aktuellen Leiden Ranking⁶ der Perzentil-basierte Indikator „Anteil der Publikationen einer Institution, die in ihrem Fachgebiet und Publikationsjahr zu den 10% der meist-zitierten Publikationen gehören“ als der wichtigste Indikator für die Messung der Wirkung von Publikationen angesehen (Waltman et al. 2012). Ein Vorteil dieses Indikators besteht darin, dass es einen Erwartungswert von 10% gibt: Auf der Basis einer Zufallsstichprobe aus der Gesamtheit aller Arbeiten im WoS würde man einen Anteil von 10% der Arbeiten erwarten, die zu den 10% der meist-zitierten Arbeiten in ihrer Fachkategorie und ihrem Publikationsjahr gehören.

8. Aussagekraft

Grundsätzlich kann nur die Forschung selbst den für eine inhaltliche Bewertung notwendigen Sachverstand aufbringen. Forschungsleistung sollte deshalb niemals ausschließlich auf der Basis quantitativer Indikatoren, sondern immer nur in Verbindung mit der Beurteilung durch Fachgutachter bewertet werden (Informed Peer Review). Die Bibliometrie ist also keinesfalls als Ersatz für Gutachter anzusehen. Die Frage sollte nicht sein, ob man bei der Forschungsevaluierung zitierungsbasierte Daten als Ersatz für Gutachten heranzieht, sondern wie man zuverlässige Daten erstellt und sinnvoll interpretiert. Die Kritik an der Bibliometrie sollte sich gegen eine unbedarfte bzw. unsachgemäße Durchführung und Interpretation der Ergebnisse richten. Der wachsende Trend hin zu einer Amateur-Bibliometrie stellt eine ernste Gefahr für die Forschungsevaluierung dar. Der Wunsch nach verführerischen Rankings und Zahlen wird durch einfach zu bedienende Werkzeuge erfüllt. Die Erstellung und Interpretation bibliometrischer Daten bedarf viel Erfahrung und Hintergrundinformation. Das Zählen von Zitaten ist wegen der Komplexität der Datenbanken und den diversen Fehlerquellen bei der praktischen Anwendung mit allerlei Fallstricken verbunden. Deshalb ist eine fachmännische und sorgfältige Verfahrensweise die entscheidende Voraussetzung für den sinnvollen Einsatz der Methode in der Forschungsbewertung.

Entscheidungssträger in der Leitung und Verwaltung von Forschungseinrichtungen nutzen bibliometrische Daten vermehrt für die Mittelzuweisung, für Stellenbesetzungen und für die wissenschaftliche Selbstdarstellung ihrer Einrichtung (Abbott et al. 2010). Die gegenwärtige Situation ist jedoch vielfach von der Meinung geprägt, dass das Zählen von Publikationen und deren Zitierungen eine einfache Angelegenheit sei und die Daten infolgedessen selbsterklärend seien. Die Stärken und Schwächen bibliometrischer Indikatoren werden zwar umfangreich in der bibliometrischen Fachliteratur abgehandelt, doch diese kann von den Entscheidungsträgern oft kaum beurteilt, ausgewertet und als Hintergrundinformation herangezogen werden.

9. Diskussion

Wie andere wissenschaftliche Methoden auch, stößt die Bibliometrie irgendwann an die Grenze einer sinnvollen Anwendbarkeit. Die ist spätestens dann erreicht, wenn die zu bewertenden Arbeiten nur sehr unvollständig in den Zitierindexen erfasst sind. Quantitative Indikatoren, wie der h-Index oder der JIF, werden vielfach als Grundlage für Vergleiche

6 Abrufbar unter: <http://www.leidenranking.com/>, letztes Abrufdatum: 9.6.2015.

bzw. umfassende Rankings von Personen, Institutionen und Zeitschriften herangezogen. Gerade Rankings bewirken nicht selten eine enorme Faszination (sowohl innerhalb, als auch außerhalb der Forschung) und unterliegen deshalb in besonderer Weise der Gefahr der Fehlinterpretation.

Wenn es um Geld und Ehre geht, sind auch Forscher nur Menschen und vergessen manchmal die Regeln ihrer Disziplin. Bibliometrische Daten werden dann gerne überinterpretiert (wenn es der positiven Selbstdarstellung nützt), rundweg abgelehnt (wenn sie nicht die Eigenwahrnehmung bestätigen) oder als Munition im Kampf gegen Konkurrenten missbraucht. Aus dieser Gefahr der Parteilichkeit erwächst für diejenigen, die mit der Erstellung bibliometrischer Daten beschäftigt sind, eine besondere Verantwortung. Wissenschaftler bzw. Entscheidungsträger sind als die Endnutzer solcher Daten aufgefordert, die durch die bibliometrische Forschung entwickelten Standards zu berücksichtigen (Marx / Bornmann 2014; Weingart 2005).

Literatur

- Abbott, A. / D. Cyranoski / N. Jones / B. Maher / O. Schiermeier / P. Van Noorden (2010): Do metrics matter?, in: *Nature* 465 / 7300, S. 860-862.
- Adler, R. / J. Ewing / P. Taylor / P.G. Hall (2009): A report from the International Mathematical Union (IMU) in cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS), in: *Statistical Science* 24 / 1, S. 1-28.
- Archambault, É. / É.V. Gagné (2004): *Science-Metrix: the use of bibliometrics in the Social Sciences and Humanities*, Montreal (Canada): Social Sciences and Humanities Research Council of Canada (SSHRC).
- Archambault, E. / E. Vignola-Gagne / G. Cote / V. Lariviere / Y. Gingras (2006): Benchmarking scientific output in the social sciences and humanities: The limits of existing databases, in: *Scientometrics* 68 / 3, S. 329-342.
- Bornmann, L. / H.D. Daniel (2008): What do citation counts measure? A review of studies on citing behavior, in: *Journal of Documentation* 64 / 1, S. 45-80.
- Bornmann, L. / H.D. Daniel (2009): Der h-Index – Messung von Forschungsleistungen, in: *BIOspektrum* 15, S. 336-337.
- Bornmann, L. / W. Marx (2011): The h index as a research performance indicator, in: *European Science Editing* 37 / 3, S. 77-80, online abrufbar unter: <http://www.lutz-bornmann.de/icons/viewpoints.pdf>, letztes Abrufdatum: 1.5.2015.
- Bornmann, L. / W. Marx / A.Y. Gasparyan / G.D. Kitas (2012 a): Diversity, value and limitations of the journal impact factor and alternative metrics, in: *Rheumatology International* 32, S. 1861-1867.
- Bornmann, L. / W. Marx (2013 a): Evaluating individual researchers' performance, in: *European Science Editing* 39 / 2, S. 39-40.
- Bornmann, L. / W. Marx (2013 b): Vorschläge für Standards zur Anwendung der Szientometrie bei der Evaluation von einzelnen Wissenschaftler(inne)n im Bereich der Naturwissenschaften, in: *Zeitschrift für Evaluation* 12 / 1, S. 103-127.
- Bornmann, L. / W. Marx (2014 a): The wisdom of citing scientists, in: *Journal of the Association for Information Science and Technology* 65 / 6, S. 1288-1292.
- Bornmann, L. / W. Marx (2014 b): Distributions instead of single numbers: Percentiles and beam plots for the assessment of single researchers, in: *Journal of the Association for Information Science and Technology* 65 / 1, S. 206-208.
- Bornmann, L. / W. Marx (2014 c): How to evaluate individual researchers working in the natural and life sciences meaningfully? A proposal of methods based on percentiles of citations, in: *Scientometrics* 98 / 1, S. 487-509.

- Bornmann, L. / H. Schier / W. Marx / H.D. Daniel (2012 b): What factors determine citation counts of publications in chemistry besides their quality?, in: *Journal of Informetrics* 6 / 1, S. 11-18.
- Diekmann, A. / M. Näf / M. Schubiger (2012): Die Rezeption (Thyssen-)preisgekrönter Artikel in der „Scientific Community“, in: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 64, S. 563-581.
- Dimitrov, J.D. / S.V. Kaveri / J. Bayry (2010): Metrics: journal's impact factor skewed by a single paper, in: *Nature* 466 / 7303, S. 179.
- Garfield, E. (1979): *Citation Indexing – Its Theory and Application in Science, Technology, and Humanities*, New York / NY.
- Garfield, E. (2006): The history and meaning of the Journal Impact Factor. *Journal of the American Medical Association* 295 / 1, S. 90-93.
- Hirsch J.E. (2005): An index to quantify an individual's scientific research output, in: *Proceedings of the National Academy of Sciences of the United States of America* 102 / 46, S. 6569-16572.
- Jansky, K. (1933): Radio waves from outside the solar system, in: *Nature* 132, S. 66-66.
- Marx, W. (2011): Bibliometrie in der Forschungsbewertung, in: *Forschung & Lehre* 18 / 11, S. 858-860, online abrufbar unter: <http://www.forschung-und-lehre.de/wordpress/?p=9147>, letztes Abrufdatum: 1.5.2015.
- Marx, W. / L. Bornmann (2012). Der Journal Impact Factor: Aussagekraft, Grenzen und Alternativen in der Forschungsevaluation, in: *Beiträge zur Hochschulforschung* 34 / 2, S. 50-66, online abrufbar unter: <http://www.ihf.bayern.de/>, letztes Abrufdatum: 1.5.2015.
- Marx, W. / L. Bornmann (2013 a): Journal Impact Factor: ‚the poor man's citation analysis‘ and alternative approaches, in: *European Science Editing* 39 / 2, S. 62-63.
- Marx, W. / L. Bornmann (2013 b): Wie gut ist Forschung wirklich? Perzentile zur Messung von Publikationsleistungen, in: *BIOspektrum* 19 / 3, S. 332-334.
- Marx, W. / L. Bornmann (2013 c): How good is research really? Measuring the citation impact of publications with percentiles to ensure correct assessments and fair comparisons, in: *EMBO Reports* 14 / 3, S. 226-230.
- Marx, W. / L. Bornmann (2014): On the problems of dealing with bibliometric data, in: *Journal of the Association for Information Science and Technology* 65 / 4, S. 866-867.
- Moed, H.F. (2005): *Citation analysis in research evaluation*, Dordrecht.
- Neuhaus, C. / W. Marx / H.D. Daniel (2009): The publication and citation impact profiles of *Angewandte Chemie* and the *Journal of the American Chemical Society* based on the sections of *Chemical Abstracts*: A case study on the limitations of the Journal Impact Factor, in: *Journal of the American Society for Information Science and Technology* 60 / 1, S. 176-183.
- Seglen, P.O. (1992): The skewness of science, in: *Journal of the American Society for Information Science* 43 / 9, S. 628-638.
- Seglen, P.O. (1997): Why the impact factor of journals should not be used for evaluating research, in: *British Medical Journal* 314 / 7079, S. 498-502.
- Vanclay, J.K. (2012): Impact Factor: outdated artefact or stepping-stone to journal certification?, in: *Scientometrics* 92 / 2, S. 211-238.
- Van Raan, A.F.J. (2000): The Pandora's box of citation analysis: Measuring scientific excellence – The last evil? *The Web of Knowledge – A Festschrift in honour of Eugene Garfield*, Medford / NJ, S. 301-319.
- Waltman, L. et al. (2012): The Leiden Ranking 2011 / 2012: data collection, indicators, and interpretation, in: *Journal of the American Society for Information Science and Technology* 63 / 12, S. 2419-2432.
- Weingart, P. (2005): Das Ritual der Evaluierung und die Verführung der Zahlen, in: Ders., *Die Wissenschaft der Öffentlichkeit: Essays zum Verhältnis von Wissenschaft, Medien und Öffentlichkeit*, Weilerswist, S. 102-122.

Dr. Werner Marx
Max-Planck-Institut für Festkörperforschung
Heisenbergstraße 1
70569 Stuttgart
w.marx@kf.mpg.de

Dr. Dr. habil. Lutz Bornmann
Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V.
Generalverwaltung
Hofgartenstraße 8
80539 München
bornmann@gv.mpg.de

