Towards objectivity in research evaluation using bibliometric indicators - a protocol for incorporating complexity

Retzer, Vroni [a,*] & Jurasinski, Gerald [b]

[a] Biogeographie, Universität Bayreuth, 95440 Bayreuth, Germany

[b] Landschaftsökologie und Standortkunde; MLR; Universität Rostock; Justus-von-Liebig-Weg 6; 18059 Rostock

[*] Corresponding author. Tel.: +49 921 552259; fax: +49 921 552315;

E-mail address: Vroni.Retzer@uni-bayreuth.de (V. Retzer)

Abstract

Publications are thought to be an integrative indicator best suited to measure the multifaceted nature of scientific performance. Therefore, indicators based on the publication record (citation analysis) are the primary tool for rapid evaluation of scientific performance. Nevertheless, it has to be questioned whether the indicators really do measure what they are intended to measure because people adjust to the indicator value system by optimizing their indicator rather than their performance. Thus, no matter how sophisticated an indicator may be, it will never be proof against manipulation. A literature review identifies the most critical problems of citation analysis: database-related problems, inflated citation records, bias in citation rates and crediting of multi-author papers. We present a step-by-step protocol to address these problems. By applying this protocol, reviewers can avoid most of the pitfalls associated with the pure numbers of indicators and achieve a fast but fair evaluation of a scientist's performance. We as ecologists should accept complexity not only in our research but also in our research evaluation and should encourage scientists of other disciplines to do so as well.

Keywords

authorship; contributorship; citation analysis; citation index; peer review; publication bias

Zusammenfassung

Publikationen gelten als guter integrativer Gradmesser für die Beurteilung wissenschaftlicher Leistung. Deswegen werden Indikatoren, die auf der Publikationstätigkeit beruhen, zunehmend für die Evaluation eingesetzt. Dabei ist es jedoch fraglich ob der Indikator wirklich misst, was er messen soll, da sich Menschen an das System anpassen und eher ihren Indikatorwert optimieren als die Leistung für die der Indikator stehen soll. Kein Indikator ist per se immun gegen Manipulationen, wie hoch entwickelt er auch sein mag. Eine Literaturübersicht identifiziert die größten Problemfelder der Zitationsanalyse: Datenbankineffizienz, Aufblähen von Publikationstätigkeit, systematische Verzerrungen bei Referenzen, sowie Schwierigkeiten den Beitrag Einzelner zu Artikeln mit vielen Autor/innen zu bewerten. Wir präsentieren ein detailliertes Prüfschema, dessen systematische Anwendung die häufigsten Schwierigkeiten im Zusammenhang mit bibliometrischen Indikatoren minimiert und dadurch eine zügige und faire Bewertung wissenschaftlicher Leistung erleichtert. Schließlich sollten wir Ökolog/innen Komplexität nicht nur in unserer Forschung akzeptieren sondern uns und unsere Kolleg/innen aus anderen Fachbereichen ermutigen, Komplexität auch in der Evaluation von Forschungstätigkeit zu akzeptieren und in die Bewertung einzubeziehen.

Introduction - the development of bibliometric indicators

Communicating one’s results to others is central for the advancement of science. Lasting impact can mainly be achieved through publications in books or journals, which stimulate others working in the same field. In turn, the publication record of a scientist is representative of his/her scientific performance (Kostoff, 1998). Therefore evaluation of the scientific performance of individual researchers is mainly based on the analysis of the publication record (Schoonbaert & Roelants, 1996).

Ever since publishing records have been used as indicators for scientific performance, this indicator system has been subjected to adaptations. To address generalization problems of indicators on rather course levels of analysis, more and more detailed indicators were introduced – from indicators based on the level of journals to the level of a paper and finally to the single author (Fig. 1).

One of the most common indicators - the 'impact factor' - was introduced in 1955 to evaluate the impact of a particular paper on the literature and thinking of the period. Later it was applied to identify influential journals (see Garfield, 2006 for a review). The journal impact factor is calculated as the number of citations in a given year to items published in a journal within the two previous years, divided by the number of papers published in the journal during the same two year period (Garfield, 2006). Although citation rates of individual papers are positively correlated with the impact factor of the journal in which they are published, they also show considerable variability - especially in high ranking journals (Leimu & Koricheva, 2005). Thus, the number of citations a paper receives is regarded as a better indicator of the paper’s scientific influence than the journal impact factor (Kurmis, 2003). To acknowledge the impact of individual scientists, several indices have been proposed recently that combine the number of papers by a certain author with the citations they received (h-index by Hirsch, 2005; g-index by Egghe, 2006). These indices operate on the single paper level; but even here inconsistencies may arise from self-citations and multi-authored papers (see Fig. 1 and below).

However, bibliometric analysis is not as objective and unbiased as it may seem (e.g. Schoonbaert & Roelants, 1996; Wallin, 2005). Bibliometric indicators are increasingly applied by personnel not trained in citation analysis (e.g. recruitment committees or funding agencies). Therefore, a review of a scientist’s performance based on citation analysis should always be accompanied by a critical evaluation of the analysis itself. To counteract the naive use of such indicators for the evaluation of individual researchers we review the literature, derive a general model on the suitability of indicators in general, and apply it to scientific evaluation using bibliometric indicators. Finally we develop a protocol that addresses the central problems to minimize bias in citation analysis.

The general dilemma of indicators and its application to scientist evaluation with publication indicators

It seems a general sociological phenomenon that the suitability of any indicator decreases with its application (see e.g. the debate on appropriate indicators for unemployment rates (Jones & Riddell, 1999); Fig. 2). After its introduction evaluated individuals adjust to its use and the indicator degrades and tends to no longer measure what it was intended to measure. This seems to be an unsolvable dilemma independent of discipline (compare e.g. monitoring of nature conservation results) and indicator. It can be described as a three-phase process:

(1) Indicator selection (Fig. 3A): In order to simplify the evaluation of some quality (e.g., scientific performance), an indicator (or a set of a few indicators) is identified, which is closely related to the quality in question, but much more easily measured (e.g., number of peer-reviewed publications).

(2) Application phase (Fig. 3B): Resources (e.g., funding, jobs) are awarded according to the quality as measured by the applied indicator(s).

(3) Adjustment phase (Fig. 3C): Beneficiaries (e.g. scientists) adjust to the system by optimizing the indicator rather than their performance itself (Frey, 2006). Thus the correlation between the scientist's performance and the value of the selected indicator becomes worse. After a while, it may even become questionable whether the indicator still measures what had originally been the focal quality. Therefore, rules are installed which aim at preventing misuse to maintain the reliability of the indicator. This is followed by further adaptation, stricter rules etc.

Indicators derived from the publication record of a scientist have long been used to evaluate scientific performance (e.g. Garfield, 1955). Today these indicators are widely applied: high scores in publication indicators directly increase the likelihood of receiving funding, getting jobs, earning better salaries (Hilmer & Hilmer, 2005) or receiving financial rewards (Fuyuno & Cyranoski, 2006). Due to the increasing competition for sparse funding and career opportunities - which is fought out by means of publications - frequent publishing in high ranking journals becomes more and more important. Contemporary science - at least in the natural sciences - clearly is in the adjustment phase. Typical adjustment strategies of scientist are "honorary" authorships, publishing of the "least publishable unit" (Huth, 1986; Brice & Bligh, 2005) and self-citation (Leimu & Koricheva, 2005). Editors adapt as well, e.g. by boosting the impact factor (Gowrishankar & Divakar, 1999; Krauss, 2007). Attempts to maintain the reliability of the indicator include quantification of authorship (Shapiro, Wenger & Shapiro, 1994; Tscharntke, Hochberg, Rand, Resh & Krauss, 2007), authorship guidelines (DFG, 1998; Weltzin, Belote, Williams, Keller & Engel, 2006) or novel ways to identify duplicate publication and plagiarism (Errami & Garner, 2008).

Review of potential pitfalls and problems of publication indicators and how to address them: A standardized protocol for dealing with complexity

A number of factors render it difficult to compare citation patterns objectively. These can be sorted into four categories associated with the different steps in citation analysis: (1) technical problems restrict the usefulness of databases, (2) the publication record may be boosted fraudulently, (3) citation rates are biased, and (4) fair crediting of multi-author papers is difficult.

(1) The use of electronic databases to assess the publication record of an individual is convenient, but not without problems. Available databases are largely biased towards journals published in English speaking countries (e.g., MacRoberts & MacRoberts, 1996; Kurmis, 2003). Therefore, it is difficult to separate language effects from regional effects. Furthermore, coverage of journals differs substantially between different disciplines (Seglen, 1997). When comparing candidates from different regional backgrounds or different research fields, these database biases have to be taken into account (Table 1).

Another problem is the considerable time lag between acceptance of a paper and its incorporation in searchable databases. Younger researchers are especially affected because a higher share of their total papers may be stuck in this queue, and thus can hardly be found and cited. Other problems are associated with attributing papers to the actual author(s). This is problematic for people with surnames that are very common (homonyms; see Wooding, Wilcox-Jay, Lewison & Grant, 2005), that contain characters which are not found in the English alphabet (e.g., ä, å, ø, ß), or - even worse - that originally were spelt in non-Latin alphabet. Such technical problems can be addressed by carefully cross-checking the reference list provided by the evaluated persons with database results to minimize misunderstandings (Table 1). Another technical problem is the exact matching of cited papers and citing articles (Van Raan, 2005). All these factors result in a percentage of mismatches of up to 25 % (Seglen, 1997).

(2) Publishing the same results twice is considered as a scientific misconduct. Therefore it is a well known strategy to minimize the novelty content per paper to maximize the number of papers from a given study. This is the so-called "least publishable unit" or "salami tactic" (Huth, 1986; Brice & Bligh, 2005). Figures regarding "honorary" or "gift" authorship are harder to come by, but Flanagin et al. (1998) found that 19 % of the medical papers evaluated showed evidence of honorary authorship. Although such practices are banned by the policy of many journals (see e.g. ethical guidelines of Elsevier: www.elsevier.com/wps/find/intro.cws_home/ethical_guidelines for BAAE) and funding agencies (e.g.DFG, 1998) they nevertheless are rather common (Eastwood, Derish, Leash & Ordway, 1996).

The increase in multi-authored scientific papers has long been debated (Zuckerman, 1968; Weltzin et al., 2006). Obviously multi-authorship reflects increasing research complexity which leads to intensified collaboration and democratization of reporting: Postgraduates and research assistants now more often receive appropriate credit (Manten, 1977). Additionally, the increase in multi-authored papers is interpreted as a signal that scientists are trying to boost their publication records (Brice & Bligh, 2005). As long as the credit for a particular article is given equally to all authors, the sequence of authors matters only a little. However, this frequently does not appropriately reflect the contributions of the authors involved (e.g. Shapiro et al., 1994; Weltzin et al., 2006; Tscharntke et al., 2007). To address these issues, at least a number of selected papers of any evaluated candidate or beneficiary should be examined in detail (Kaltenborn & Kuhn, 2003) to check for self-citation, methodological and conceptional breadth and novelty of content (Table 1).

(3) Citation analysis is based on the assumption that the influence and quality of a paper is reflected in the number of articles citing it. This approach neglects the fact that citing - and not-citing as well - is "a complex social-psychological behavior" (MacRoberts & MacRoberts, 1996). Citation rates are biased by the field of research (Batista, Campiteli, Kinouchi & Martinez, 2006), towards authors with English as their native language (Leimu & Koricheva, 2005), or even by the alphabetical position of the surname (Tregenza, 1997), and - at least in some disciplines - by gender towards being male (Trimble, 1993). Additionally, citation practices reflect the geographic region of author and citer (Wong & Kokko, 2005) as well as group membership and friendship (White, Wellman & Nazer, 2004). Researchers who are cooperating in large projects and/or working in a field that is already influential are over-represented (Glänzel, 2002; Kretschmer, 2004). Zuckerman (1968) found that an above-average proportion of co-authors and students of Nobel laureates received a Nobel Prize themselves. However, it is difficult to tell whether this is due to the high quality of science conducted in such groups or rather attributable to the prestige associated with certain affiliations (Leimu & Koricheva, 2005). Therefore the factors gender, affiliation and regional/language background of a candidate should be discussed in parallel to the citation record when evaluating scientific performance (Table 1).

Another difficult point is the deviation from scientific routine. Very progressive work may not be detected by publication indicators because it is neither easy to publish (as it challenges old paradigms) nor heavily cited because the new area of research is advanced by a small number of people only (Kuhn, 1976). Therefore, citation analysis is not suited to differentiate between below-average and brilliant science. Papers with controversial results have to be evaluated even more carefully (Table 1).

Space restriction by journals limits the number of citations per article. Often citations are chosen so that one reference covers many aspects but will be just one item in the list of references. This is frequently encouraged by editors suggesting to shorten manuscripts considerably. However, during the creative process of designing a study, analyzing it and writing a paper, numerous publications have influenced the author(s) - many more than can be cited (Seglen, 1998). As a result citations are to a certain extent arbitrary. All this may be reflected by the fact that an astonishingly large proportion of citations do not clearly support the statement made (Todd, Yeo, Li & Ladle, 2007). In conclusion, citation rates are heavily affected by factors other than the scientific utility (Leimu & Koricheva, 2005), which clearly shows the limits of citation analysis (Table 1).

(4) Under the assumption of different contributions per author, multi-authorship generally presents a two stage problem: First, it has to be determined who qualifies for authorship. Second, a sequence has to be found, that fairly acknowledges each contribution. Regarding the first point, several publishing institutions have developed criteria concerning the qualification for authorship (see Weltzin et al., 2006). However, these are not applied consistently (Leash, 1997). One reason among many others are subordination dependencies: 32% of researchers on Post-Doc positions would include people as authors - who according to their own definition do not qualify for authorship - if they believe it benefits their career (Eastwood et al., 1996). Regarding the second point, the meaning of author sequence varies between scientific communities, groups, and journals. Only some decades ago an alphabetical order of author names was not unusual (Zuckerman, 1968). Today it is generally accepted, that the sequence reflects the authors' contributions. For instance, in clinical research (e.g. Zuckerman, 1968; Drenth, 1998) and increasingly in the biological sciences (Tscharntke et al., 2007), the last author is regarded as the senior author and ranks second after the first author: The first is thought to have done most of the writing; the last is thought to bear most of the responsibility. However, even within one discipline contributions of authors vary between scientific groups and single papers (Shapiro et al., 1994).

Thus, without knowing the actual contribution of each author, evaluation committees and research funding agencies can only guess. This leads to considerable uncertainty among reviewers and authors alike (Laurance, 2006). In a real case example it has been shown that simple differences in the perception of authorship order can change the ranking of scientists (Moulopoulos, Sideris & Georgilis, 1983). Thus, several authors have proposed standardized ranking schemes to match the ranking by the authors with the perception of readers/referees (e.g. Hunt, 1991; Verhagen, Collins & Scott, 2003; Tscharntke et al., 2007). Alternatively, Moulopoulos et al. (1983) suggest the inclusion of a footnote or box in which the contributions of all authors are explained (see also Huth, 1986; Shapiro et al., 1994; Rennie, Yank & Emanuel, 1997). Consequently, Shapiro et al. (1994) and Rennie et al. (1997) recommend to substitute the term authorship with contributorship. If authors are forced to explain their contribution, misuse and social conflicts among researchers (Klein & Moser-Veillon, 1999) might be prevented by communication rather than sanctions (Rennie, Flanagin & Yank, 2000). Nevertheless, even improved authorship assignment schemes (e.g. by Hunt, 1991; Tscharntke et al., 2007) or detailed descriptions (Hueston & Mainous, 1998) may just induce a shift from "honorary" authorship to "honorary" contribution. If available, additional information on authorship contribution should be taken into account (Fig. 3). Otherwise the authorship ranking traditions of different research fields and nations should be given a second thought (Table 1).

Finally, it has to be mentioned that bias isn't restricted to citation but also exists before publication: The typical single-blind peer review process itself is biased e.g. regarding gender (Budden et al., 2008), study outcome (Koricheva, 2003) or study organism (Bonnet, Shine & Lourdais, 2002), and is not easily reproducible (Peters & Ceci, 1982; Hojat, Gonnella & Caelleigh, 2003). Furthermore, non peer reviewed articles are not included in citation analysis.

The possibilities to address the problems identified above are summarized in a protocol that should be considered in addition to bibliometric indicators to achieve a more objective evaluation of scientific excellence (Table 1). As citation analysis is often applied by scientists who are not trained in citation analysis, comparable results can only be obtained by standardized analysis.

Objectivity in research evaluation

All these facts "*cast doubt on the validity of using citation counts as an objective and unbiased tool for academic evaluation*" (Leimu & Koricheva, 2005). Thus we have to go back to the question of how scientific potential and achievement can be measured. It has been shown that selecting another or finding a new indicator cannot be the solution because it will share the same fate as the previous one (Fig. 2). Furthermore, numerous reviews on the topic agree that scientific performance can hardly be assessed by a single indicator (Schoonbaert & Roelants, 1996; MacRoberts & MacRoberts, 1996; Kostoff, 1998; Phelan, 1999; Golder, 2000; Bloch & Walter, 2001; Kurmis, 2003; Kaltenborn & Kuhn, 2003; Cartwright & McGhee, 2005; Wallin, 2005; Ha, Tan & Soo, 2006). Citation analysis can supplement, but never substitute a thorough peer review, which assesses also a number of other abilities and activities such as fund raising, communication with other scientists or teaching performance. "*This confrontation with the content of the science, which demands time and care, is the essential core of peer review for which there is no alternative*" (DFG, 1998).

Of course, most of the points discussed above are known to experienced reviewers (but not necessarily followed). But, as Lawrence (2003) stresses, a lot of knowledge on ethical scientific behaviour is transferred informally. Thus senior scientists have to take the lead as role models. Generally, science tries to reduce complexity to answer general questions. But maybe we as ecologists have to accept complexity not only in our research but also in research evaluation. By applying the developed protocol, reviewers can avoid most pitfalls associated with the pure numbers of indicators and achieve a sufficiently straightforward but comparable and fair evaluation of a scientist's performance.

Acknowledgments

We thank David Tait for revising the English. The comments of three anonymous reviewers on a former version of the manuscript helped improve the manuscript considerably.

Tables and Figures

Table 1: A protocol to incorporate more complexity in research evaluation: Problems of citation analysis and how they can be addressed.

Identified problem	How to address it
Database-related problems	Technical problems: Consider biases in databases regarding journals covered, time until incorporation (papers in press!), mismatches, and spelling of names. Carefully cross-check the reference list provided by the persons in question with the results from the database to minimize misunderstandings.
	Databases show a regional and language bias: When comparing candidates from different regional backgrounds, be aware of database bias towards English language and US/UK journals.
	Database coverage differs between research fields: When comparing candidates from different scientific backgrounds, be aware that the database might not cover the different fields equally.
Boosted citation record	Self-citation: Check whether the publication record is flawed by unnecessary self-citation.
	Least publishable unit and repeated publications: Check whether the different papers really offer novel results. Use information on highly similar and duplicate papers such as Déjà vu database (http://discovery.swmed.edu/dejavu/) where available.
Bias in citation rates	Citations are not (only) for merit and not all sources are cited: These general problems cannot be dealt with. Remember the limitations of citation analysis.
	Citation patterns are influenced by group membership and "citation clubs": Take into account the scientific background of a person including group performance, institutional background (e.g. research institution or university) and his/her status within the group.
	Citation rates are biased by regional and language background: Consider the different regional backgrounds, which includes different publication and citing traditions, e.g., regarding positioning in multi-authored papers.
	Carefully check the content of controversial papers.
Crediting multi-author papers	Take into account contributorship / authorship ranking information where available. Otherwise consider that different author ranking traditions may exist in different nationalities and research fields.

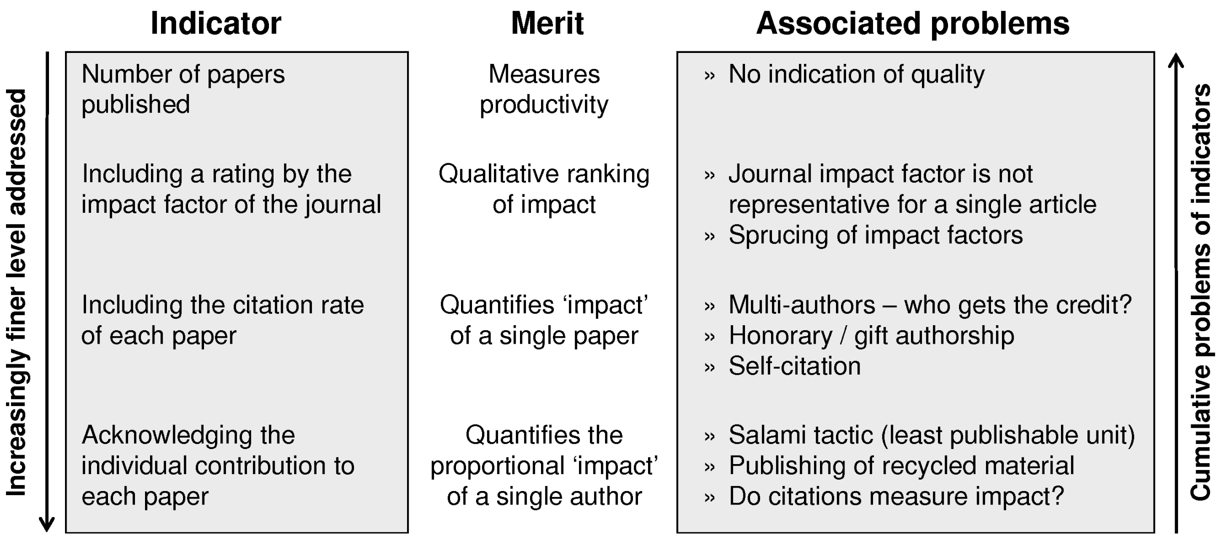

Figure 1. Conceptual sketch of the development of indicators measuring scientific performance, their merits and associated problems. Indicators for measuring scientific impact have been constantly developed to quantify the impact of papers on ever finer scales (left panel). By doing so the merits are more efficiently attributed to individuals (middle panel). Each of these steps has solved some associated problems, but some general problems cannot be tackled (right panel).

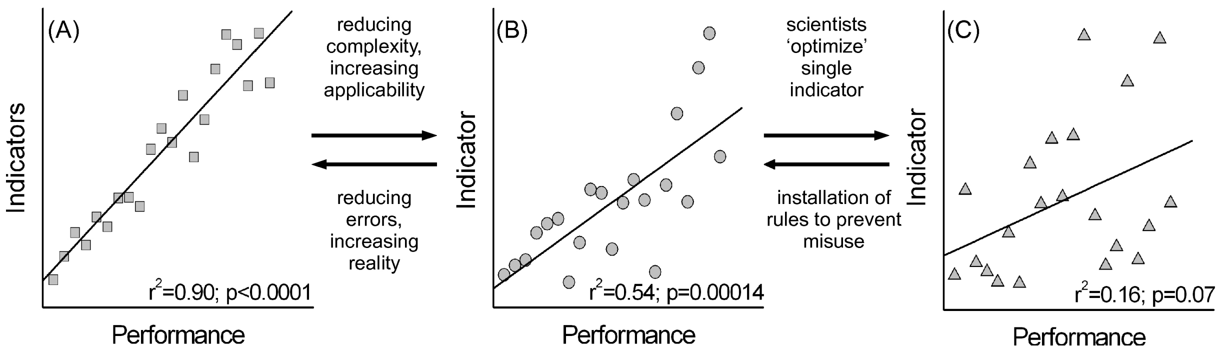

Figure 2. Indicator degradation - conceptual diagram of the three phases of measuring scientific performance of different researchers using indicators. Scientific performance is always the same, only the indicator(s) measuring it differ. (A) Phase 1 – indicator selection: Performance can be well predicted using a set of multiple indicators (relative scale). To reduce complexity and increase applicability, the full set of available indicators is reduced to a few – in the most extreme cases to a single indicator. (B) Phase 2 – application: The selected indicator can be applied rather successfully, although the correlation is lower than that of a set of multiple indicators. (C) Phase 3 – adjustment phase: As soon as a certain indicator is frequently used, beneficiaries are tempted to improve the indicator value rather than their performance. In turn, rules have to be established that prevent such 'optimization strategy'. Otherwise the correlation between indicator and performance diminishes further until it may no longer be significant. Data are artificial; statistics are based on these artificial data.

Relative contributorship (vr/gj)	
Initial conception	50%/50%
Design of the study	80%/20%
Provision of resources	00%/00%
Data collection	70%/30%
Analysis and interpretation	70%/30%
Writing and revising draft(s)	50%/50%

Figure 3. Example of acknowledging contributorship: proportional contribution (%) of Vroni Retzer (vr) and Gerald Jurasinski (gj) to the different stages of producing this paper.

References

Batista, P.D., Campiteli, M.G., Kinouchi, O., & Martinez, A.S. (2006). Is it possible to compare researchers with different scientific interests? Scientometrics, 68/1, 179-189.

Bloch, S. & Walter, G. (2001). The Impact Factor: Time for change. Australian and New Zealand Journal of Psychiatry, 35/5, 563-568.

Bonnet, X., Shine, R., & Lourdais, O. (2002). Taxonomic chauvinism. Trends in Ecology & Evolution, 17/1, 1-3.

Brice, J. & Bligh, J. (2005). Author misconduct: not just the editors' responsibility. Medical Education, 39/1, 83-89.

Budden, A.E., Tregenza, T., Aarssen, L.W., Koricheva, J., Leimu, R., & Lortie, C.J. (2008). Double-blind review favours increased representation of female authors. Trends in Ecology & Evolution, 23/1, 4-6.

Cartwright, V.A. & McGhee, C.N.J. (2005). Ophthalmology and vision science research – Part 1: Understanding and using journal impact factors and citation indices. Journal of Cataract and Refractive Surgery, 31/10, 1999-2007.

DFG 1998: Recommendations of the Commission on Professional Self Regulation in Science. http://www.dfg.de/antragstellung/gwp/, last access: 27.8.2007.

Drenth, J.P.H. (1998). Multiple Authorship: The Contribution of Senior Authors. JAMA, 280/3, 219-221.

Eastwood, S., Derish, P., Leash, E., & Ordway, S. (1996). Ethical issues in biomedical research: perceptions and practices of postdoctoral research fellows responding to a survey.Science and engineering ethics, 2/1, 89-114.

Egghe, L. (2006). Theory and practise of the g-index. Scientometrics, 69/1, 131-152.

Errami, M. & Garner, H. (2008). A tale of two citations. Nature, 451/7177, 397-399.

Flanagin, A., Carey, L.A., Fontanarosa, P.B., Phillips, S.G., Pace, B.P., Lundberg, G.D., & Rennie, D. (1998). Prevalence of articles with honorary authors and ghost authors in peer-reviewed medical journals. Jama-Journal of the American Medical Association, 280/3, 222-224.

Frey, B.S. (2006). Evaluitis - eine neue Krankheit. Working Paper Series 293. Institute for Empirical Research in Economics, University Zürich.

Fuyuno, I. & Cyranoski, D. (2006). Cash for papers: putting a premium on publication. Nature, 441/7095, 792-792.

Garfield, E. (1955). Citation Indexes for Science - New Dimension in Documentation Through Association of Ideas. Science, 122/3159, 108-111.

Garfield, E. (2006). The history and meaning of the journal impact factor. Jama-Journal of the American Medical Association, 295/1, 90-93.

Glänzel, W. (2002). Coauthorship patterns and trends in the sciences (1980-1998): A bibliometric study with implications for database indexing and search strategies. Library Trends, 50/3, 461-473.

Golder, W. (2000). Who controls the controllers? Ten statements on the so-called impact factor. Onkologie, 23/1, 73-75.

Gowrishankar, J. & Divakar, P. (1999). Sprucing up one's impact factor. Nature, 401/6751, 321-322.

Ha, T.C., Tan, S.B., & Soo, K.C. (2006). The journal impact factor: Too much of an impact? Annals Academy of Medicine Singapore, 35/12, 911-916.

Hilmer, C.E. & Hilmer, M.J. (2005). How do journal quality, co-authorship, and author order affect agricultural economists' salaries? American Journal of Agricultural Economics, 87/2, 509-523.

Hirsch, J.E. (2005). An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences of the United States of America, 102/46, 16569-16572.

Hojat, M., Gonnella, J.S., & Caelleigh, A.S. (2003). Impartial judgment by the "gatekeepers" of science: Fallibility and accountability in the peer review process. Advances in Health Sciences Education, 8/1, 75-96.

Hueston, W.J. & Mainous, A.G. (1998). Authors vs contributors: Accuracy, accountability, and responsibility. Jama-Journal of the American Medical Association, 279/5, 356-356.

Hunt, R. (1991). Trying An Authorship Index. Nature, 352/6332, 187-187.

Huth, E.J. (1986). Irresponsible Authorship and Wasteful Publication. Annals of Internal Medicine, 104/2, 257-259.

Jones, S.R.G. & Riddell, W.C. (1999). The measurement of unemployment: An empirical approach. Econometrica, 67/1, 147-161.

Kaltenborn, K.F. & Kuhn, K. (2003). The journal impact factor as a parameter for the evaluation of researchers and research. Medizinische Klinik, 98/3, 153-169.

Klein, C.J. & Moser-Veillon, P.B. (1999). Authorship: Can you Claim a Byline? Journal of the American Dietetic Association, 99/1, 77-79.

Koricheva, J. (2003). Non-significant results in ecology: a burden or a blessing in disguise? Oikos, 102/2, 397-401.

Kostoff, R.N. (1998). The use and misuse of citation analysis in research evaluation - Comments on theories of citation? Scientometrics, 43/1, 27-43.

Krauss, J. (2007). Journal self-citation rates in ecological sciences. Scientometrics, 73/1, 79-83.

Kretschmer, H. (2004). Author productivity and geodesic distance in bibliographic co-authorship networks, and visibility on the Web. Scientometrics, 60/3, 409-420.

Kuhn, T.S. (1976). Die Struktur wissenschaftlicher Revolutionen. Frankfurt am Main: Suhrkamp.

Kurmis, A.P. (2003). Current concepts review - Understanding the limitations of the journal impact factor. Journal of Bone and Joint Surgery-American Volume, 85A/12, 2449-2454.

Laurance, W.F. (2006). Second thoughts on who goes where in author lists. Nature, 442/7098, 26-26.

Lawrence, P.A. (2003). The politics of publication - Authors, reviewers and editors must act to protect the quality of research. Nature, 422/6929, 259-261.

Leash, E. (1997). Is it time for a new approach to authorship? Journal of Dental Research, 76/3, 724-727.

Leimu, R. & Koricheva, J. (2005). What determines the citation frequency of ecological papers? Trends in Ecology & Evolution, 20/1, 28-32.

MacRoberts, M.H. & MacRoberts, B.R. (1996). Problems of citation analysis. Scientometrics, 36/3, 435-444.

Manten, A.A. (1977). Multiple Authorship in Animal-Science. Applied Animal Ethology, 3/4, 299-304.

Moulopoulos, S.D., Sideris, D.A., & Georgilis, K.A. (1983). For debate . . . Individual contributions to multiauthor papers. British Medical Journal (Clinical research ed.), 287/6405, 1608-1610.

Peters, D.P. & Ceci, S.J. (1982). Peer-Review Practices of Psychological Journals - the Fate of Accepted, Published Articles, Submitted Again. Behavioral and Brain Sciences, 5/2, 187-195.

Phelan, T.J. (1999). A compendium of issues for citation analysis. Scientometrics, 45/1, 117-136.

Rennie, D., Yank, V., & Emanuel, L. (1997). When authorship fails. A proposal to make contributors accountable. JAMA, 278/7, 579-585.

Rennie, D., Flanagin, A., & Yank, V. (2000). The Contributions of Authors. JAMA, 284/1, 89-91.

Schoonbaert, D. & Roelants, G. (1996). Citation analysis for measuring the value of scientific publications: Quality assessment tool or comedy of errors? Tropical Medicine & International Health, 1/6, 739-752.

Seglen, P.O. (1997). Citations and journal impact factors: questionable indicators of research quality. Allergy, 52/11, 1050-1056.

Seglen, P.O. (1998). Citation rates and journal impact factors are not suitable for evaluation of research. Acta Orthopaedica Scandinavica, 69/3, 224-229.

Shapiro, D.W., Wenger, N.S., & Shapiro, M.F. (1994). The contributions of authors to multiauthored biomedical research papers. JAMA, 271/6, 438-442.

Todd, P.A., Yeo, D.C.J., Li, D.Q., & Ladle, R.J. (2007). Citing practices in ecology: can webelieve our own words? Oikos, 116/9, 1599-1601.

Tregenza, T. (1997). Darwin a better name than Wallace? Nature, 385/6616, 480-480.

Trimble, V. (1993). Patterns in Citations of Papers by American Astronomers. Quarterly Journal of the Royal Astronomical Society, 34/2, 235-250.

Tscharntke, T., Hochberg, M.E., Rand, T.A., Resh, V.H., & Krauss, J. (2007). Author sequence and credit for contributions in multiauthored publications. Plos Biology, 5/1, 13-14.

Van Raan, A.F.J. (2005). Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. Scientometrics, 62/1, 133-143.

Verhagen, J.V., Collins, S.C., & Scott, T.R. (2003). QUAD system offers fair shares to all authors - Showing who did what could solve the problem of authorship abuse and disputes. Nature, 426/6967, 602-602.

Wallin, J.A. (2005). Bibliometric methods: Pitfalls and possibilities. Basic & Clinical Pharmacology & Toxicology, 97/5, 261-275.

Weltzin, J.F., Belote, R.T., Williams, L.T., Keller, J.K., & Engel, E.C. (2006). Authorship in ecology: attribution, accountability, and responsibility. Frontiers in Ecology and the Environment, 4/8, 435-441.

White, H.D., Wellman, B., & Nazer, N. (2004). Does citation reflect social structure? Longitudinal evidence from the "Globenet" interdisciplinary research group. Journal of the American Society for Information Science and Technology, 55/2, 111-126.

Wong, B.B.M. & Kokko, H. (2005). Is science as global as we think? Trends in Ecology & Evolution, 20/9, 475-476.

Wooding, S., Wilcox-Jay, K., Lewison, G., & Grant, J. (2005). Co-author inclusion: A novel recursive algorithmic method for dealing with homonyms in bibliometric analysis. Scientometrics, 66/1, 11-21.

Zuckerman, H.A. (1968). Patterns of Name Ordering Among Authors of Scientific Papers: A Study of Social Symbolism and Its Ambiguity. The American Journal of Sociology, 74/3, 276-291.