Reprinted from Chapter 10 of Garfield E. "Citation Indexing -- Its Theory and Application in Science, Technology, and Humanities" p.1-274, Publisher John Wiley & Sons, Inc. 1979. Reprinted by ISI Press, Philadelphia, USA. 1983

277

REPRINT

Scientometrics, Vol. 1, No. 4 (1979) 359–375

IS CITATION ANALYSIS A LEGITIMATE EVALUATION TOOL? *

E. GARFIELD

Institute for Scientific Information® (ISI®), 325 Chestnut Street, Philadelphia, PA 19106 (USA)

(Received December 27, 1978)

A comprehensive discussion on the use of citation analysis to rate scientific performance and the controversy surrounding it. The general adverse criticism that citation counts include an excessive number of negative citations (citations to incorrect results worthy of attack), self-citations (citations to the works of the citing authors), and citations to methodological papers is analyzed. Included are a discussion of measurement problems such as counting citations for multiauthored papers, distinguishing between more than one person with the same last name (homographs), and what it is that citation analysis actually measures. It is concluded that as the scientific enterprise becomes larger and more complex, and its role in society more critical, it will become more difficult, expensive and necessary to evaluate and identify the largest contributors. When properly used, citation analysis can introduce a useful measure of objectivity into the evaluation process at relatively low financial cost.

Introduction

The use of citation analysis to produce measures, or indicators, of scientific performance has generated a considerable amount of discussion.[1-10] Not surprisingly, the discussion grows particularly intense when the subject is the use of these measures to evaluate people, either as individuals or in small formal groups, such as departmental faculties in academic institutions. Published descriptions of how citation analysis is being used to define the history of scientific development, or to measure the activity and interaction of scientific specialties, generate relatively little comment from the scientific community at large. And what is generated tends to be calm and reasoned. In contrast, any mention of using citation analysis to measure the performance of specific individuals or groups produces an automatic, and often heatedly emotional, response from the same people who otherwise remain silent. A case in point is a 1975 review in *Science*[11a] of the way citation analysis is being used, on an exploratory basis, by science administrators. The article in-

*Modification of a chapter in E. GARFIELD: *Citation Indexing: Its Theory and Application in Science, Technology and the Humanities, New York, Wiley, 1979.*

cluded a discussion of the use of citation measures to define and monitor changes in the specialty structure of science. This application could have a major impact on the development of science policies. But the spate of letters to the editor that commented on the article dealt only with the use of citation data to help measure individuals and academic departments in case of tenure, promotion, and grant awards.[11b]

It is not surprising that the published comments these applications elicit from the general-science community are almost always critical. After all, scientists are no less sensitive to performance measures than other people. And when you consider that some 25% of the scientific papers published are never cited even once[12] and that the average annual citation count for papers that are cited is only 1.7[13] it is not hard to understand why citation counts might seem a particularly threatening measure to some. Another reason for the defensive attitude might be the relative newness of citation data as a measure of performance. It will be interesting to compare the reactions of humanities scholars to the one of scientists when the *Arts & Humanities Citation Index TM.*[14] becomes better known. Being cited is already considered an important mark of scholarship in the humanities, so there is reason to believe that scholars in that area will find formal citation measurements more acceptable.

Interestingly enough, the comments from the science historians and sociologists, who specialize in the difficult task of figuring out how science works and how it can be measured usefully, show much more balance. Their general attitude about citation measures consists of approximately equal parts of healthy spekticism (they know how difficult it is to quantify the quality of scientific performance) and the scientific objectivity needed to accept the positive findings of properly conducted studies. In fact, the criticism coming from some quarters of the scientific community is in the face of a long list of studies[15-21] that show citation counts correlate very highly with peer judgements, which are widely accepted as a valid way of ranking scientific performance.

The large body of evidence supporting the use of citation measures to help evaluate scientists, individually and in groups, is no reason, however, to ignore the criticism of the partice. Discussion and controversy have the capacity for increasing our understanding of a subject, and there is much need to do just that where citation measures are concerned. Though the primary criticisms have been answered in detail many times, the answers are scattered through the literature of hundreds of papers, study reports, and letters to editors. Pulling all the salient points together in a single document, as I intend to do here, may well serve the ultimate goal of increasing and broadening the understanding of a method of measurement that is potentially very important to the way science is practiced.

Another important reason for continuing and intensifying the discussion is that none of the criticisms are unfounded. Most of them are based on facets of citation analysis that pose either theoretical or real problems in using the technique to evaluate people. Those using citation data to evaluate research performance at any level, but particularly at the level of individuals, must understand both its subtleties and its limitations. The position of those who advocate the use of citation data to evaluate people is not that it is simple and foolproof, but that the problems associated with it can be solved satisfactorily with a reasonable amount of methodological and interpretative effort. In other words, none of the grounds for criticism are insurmountable obstracles in the way of using citation data to develop fair, objective, and useful measures of individual or group performance, which is something I now will attempt to demonstrate.

What do citation counts measure?

The opposition to the use of citation counts to evaluate people is based on two sets of perceived weaknesses: one has to do with the mechanics of compiling the data; the other, with the intrinsic characteristics of the data. Some of these intrinsic characteristics have to do with what citation counts measure, others with what they do not measure.

Those who claim that citation counts measure too much to be valid talk about negative citations, self-citations, and methodological papers. The first two points represent problems that appear to be more theoretical than real. For example, while it is theoretically possible that a high citation count could be produced by publishing low-quality work that attracted a lot of criticism, the apparent reluctance of scientists to go to the trouble of refuting inferior work makes such a situation very unlikely. Two sociologists of sciences have commented on this trait. *G. M. Carter*[21] wrote, "Citations of articles for negative reasons are extremely rare and unlikely to distort the use of frequency counts of citations as measures of research output." *A. J. Meadows* was even more explicit about the trait:[22] "Surprisingly enough, despite its acceptance of the need for organized skepticism, the scientific community does not normally go out of its way to refute incorrect results. If incorrect results stand in the way of the further development of a subject, or if they contradict work in which someone else has a vested interest, then it may become necessary to launch a frontal attack. Otherwise, it generally takes less time and energy to bypass erroneous material, and simply allow it to fade into obscurity."

These observations add a dimension of subtlety to the negative-citation question. If scientists tend to ignore inferior work that is of little importance, then the work that they do go to the trouble of formally criticizing must be of some substance. Why, then, should negative citations be considered a sign of discredit? Criticism, as well as communication, is one of the fundamental functions of the process of scientific publication. Many new theories and findings of importance are criticized initially. It seems presumptuous to assume that the critics are always right. They are just as likely to be wrong. A significant number of early papers of Nobel Prize winners are rejected for publication by the leading journals of their fields. And even when the paper being criticized is wrong, does the mistake diminish to zero the contribution of the scientific work being described? Do not mistake important enough to be formally refuted serve the constructive purpose of clarifying, focusing, and stimulating

The question of whether negative citations invalidate citation counts as a measure of individual performance evokes the more fundamental question of what facet of scientific performance do citation counts measure. If citation statistics were purported to be a precise measure of the number of times an individual was right, negative citations certainly would be an unacceptable aberration. But citation counts are not that kind of measure. What they are is a very general measure of the level of contribution an individual makes to the practice of science. Since scientists tend to ignore the trivial, negative citations seem to say as much about that rather abstract facet of scientific performance as positive ones.

The question of the validity of self-citation is a simpler one. Theoretically, self-citations are a way of manipulating citation rates. On the other hand, the practice of citing oneself is also both common and reasonable. Studies show that at least 10% of all citations are self-citations, when self-citations are defined as a scientist citing work on which he or she appeared as primary author. If the definition were expanded to include references to work on which the scientist was the secondary author, or to the work of a collaborator (team self-citation), the percentage undoubtedly would be much greater. Since scientists tend to build on their own work, and the work of collaborators, a high self-citation count, more often than not, indicates nothing more ominous than a narrow specialty.

The reason why this is almost always the case is that it is quite difficult to use self-citation to inflate a citation count without being rather obvious about it. A person attempting to do this would have to publish very frequently to make any difference. Given the refereeing system that controls the quality of the scientific literature in the better known journals, the high publication count could be achieved only if the person had a lot to say that was at least marginally significant. Otherwise, the person would be forced into publishing in obscure journals. The combination of a long bibliography of papers published in obscure journals and an abnormally high self-citation count would make the intent so obvious that the technique would be self-defeating.

The third point of criticism, which is the high citation counts of some methodological papers, deserves more attention. Many scientists feel that methodological advances are less important than theoretical ones. Some of them who feel that way conclude that citation counts cannot be a valid measure because they favor those who develop research methods over those who theorize about research findings.

Such a conclusion overlooks several important points. The most obvious one is the questionable validity of the judgement that methods are inherently less important than theories. It may be that they are, but no one can deny that some methods, and instruments too, have opened up major new areas of research. Whether such methods are less important than theories that have had the same impact is a classic subject for debate, rather than a scientific truth. Some indication of the interaction between methodology and theory can be found in *Citation Classics*, a weekly series of statements by authors of highly cited papers, which is published in *Current Contents*®. The authors of methods papers discuss, among other things, the impact their work has had on theory and practice.

Another, less contentious, point that is overlooked is that methods papers do not inevitably draw a large number of references. Thousands of them are never cited. If you look at the 100 most cited works in the chemical literature as compiled from *Science Citation Index*® *(SCI*®*)* data for any given time period, for example, you will find that roughly 73% do not deal primarily with experimental methodology.[23] So, while some methods papers are highly cited, certainly most are not. It varies according to the orientation of the field. In fields highly oriented to methodology, such as analytical chemistry, methods papers do tend to be highly cited. But in those that do not have a particularly strong methodological orientation, high citation counts are as much the exception, rather than the rule, for methods papers as they are for theoretical papers.

The most subtle point that is overlooked concerns the question raised earlier about the quality that citation counts measure. People talk about citation counts being a measure of the "importance", or "impact" of scientific work, but those who are knowledgeable about the subject use these words in a very pragmatic sense: what they really are talking about is utility. A highly cited work is one that has been found to be useful by a relatively large number of people, or in a relatively large number of experiments. That is the reason why certain methods papers tend to be heavily cited. They describe methods that are frequently and widely used. *O. H. Lowry*'s 1951 paper on protein measurement[24] is a classic example. It was cited 50 000 times between 1961 and 1975, a count that is more than five times as high as the second most highly cited work. The only thing the count indicates about this particular piece of *Lowry*'s work was best said by him: "It just happened to be a trifle better or easier or more sensitive than other methods, and of course nearly everyone measures protein these days."[25]

Conversely, the citation count of a particular piece of scientific work does not necessarily say anything about its elegance or its relative importance to the advancement of science or society. The fact that *Lowry's* paper on protein determination is more highly cited than *Einstein's* paper on his unified field theory certainly does not indicate that *Lowry's* contribution is more significant than *Einstein's*. All it says is that more scientists are concerned with protein determination than studying unified field theory. In that sense it is a measure of scientific activity.

The only responsible claim made for citation counts as an aid in evaluating individuals is that they provide an objective measure of the utility or impact of scientific work. They say nothing about the nature of the work, nothing about the reason for its utility or impact. Those factors can be dealt with only by content analysis of the cited material and the exercise of knowledgeable peer judgment. Citation analysis is not meant to replace judgment, but to make it more objective and astute.

What citation counts do not measure

While one school of critics is concerned with what citation counts do measure, another is equally concerned about what they do not measure. The inability of citation counts to identify premature discoveries – work that is highly significant but so far ahead of the field that it goes unnoticed – is one reason this school gives for questioning their validity. This criticism could appropriately be called the "Mendel syndrome", since those who voice it invariably refer to the long-dormant work of Gregor *Mendel*.

It is true, of course, that citation counts will not identify significance that is unrecognized by the scientific community. They are, after, all, nothing more, nor less, than a reflection of that community's work and interests. To go beyond that is to begin questioning the validity of the community's perception of things, which is another area that calls for peer judgment.

The fact is, other forms of citation analysis can be helpful in going beyond the scientific community's general perception of things.[26] There are techniques that might be useful in identifying not only premature work, but the more prevalent phenomenon of immature fields, which are characterized by being small, young, and potentially much more important than their level of activity or citation rate would indicate. But this is another subject.

As far as evaluating individuals is concerned, the inability of citation counts to go beyond the general perceptions of the scientific community seems to be irrelevant to the question of how accurately they reflect those perceptions.

Another issue that is relevant is caused by the phenomenon of obliteration, which takes place when a scientist's work becomes so generic to the field, so integrated into its body of knowledge that people frequently neglect to cite it explicitly.[27] This happens, of course, to all high quality work eventually, but in some cases it happens within a relatively short time period. For example, *Lederberg's* work on the sexual reproduction of bacteria was first published in 1960[28] but became a part of the field of genetics so quickly that the rate at which it is now cited is much lower than its importance would lead one to except. When this happens, the long-term citation count of the scientist responsible for the work may fail to reflect the full magnitude of his contribution. That is why I suggested in 1963 that a PERT-type measure be used to establish the current impact of old work.[29]

There is, however, not much chance of obliteration causing inequities. It happens only to work that makes a very fundamental and important contribution to the field; and before the obliteration takes place, both the citation count and reputation of the scientist responsible for the work usually reach a level that makes additional citation credits superfluous. Of course, obliteration might lead to bad judgments by people unfamiliar with the field, but this possibility is just another reason why evaluations should always be made by, or in consultation with, people knowledgeable in the fields of the scientists involved.

Some people are also concerned because raw citation counts do not take into account the standing, or prestige, of the journal in which the cited work was published. This is true, although the *Journal Citation Reports* section of *SCI* and the *Social Sciences Citation Index® (SSCI®)* provides rankings of journals, based on several different citation measures, that can be used for this purpose. Theoretically, it is possible, as *Narin* has shown,[30] to weight the citation counts to reflect this factor, but it is not very clear how the weights should be used. Should citations to a paper published in *Science* count for more, to reflect the accomplishment of publishing in *Science,* or less, to reflect the possible increase in citation potential that may be attributed to the high visibility of *Science* material. And, what about the journals that published the citing articles? Is not the prestige of the journal that published the citing work just as important as the one that published the cited work?

Though it is easy to speculate on the effects that journal prestige may have upon citation counts, it does not seem to be a very important factor. Since abstracting and indexing services generate visibility for the material published in most journals, it is doubtful whether publication in any particular journal increases the probability of being cited enough to justify a negative weight. Even in a journal as well known and regarded as *The Physical Review,* 47% of all the articles published in

1963 had a 1966 citation rate of 0 or 1.[31] Another reason for discounting the prestige of the journal is that most papers cited highly enough to make any difference are published in such a small group of journals that all enjoy a high level of prestige and visibility.

Another, more relevant, concern of the "can't-do" school of critics is that citation counts cannot be used to compare scientists in different fields. This is partially true, depending on the methodology used to make the comparison. It certainly is improper to make comparisons between citation counts generated in different fields.

What makes it improper is that citation potential can vary significantly from one field to another. For example, papers in the biochemistry literature now average 30 references, whereas those in the mathematical literature average fewer than 10. The potential for being cited in the biochemistry literature, then, is three times that of the mathematical literature. Work by *Koshy*[12] shows that the variations in citation rates and patterns that exist from one discipline to another extend to such citation characteristics as how quickly a paper will be cited, how long the citation rate will take to peak, and how long the paper will continue being cited.

In fact, there is reason to believe that the disciplinary distinctions made between fields may not always be fine enough to avoid unfair comparisons. Work on co-citation analysis[26] suggests quite strongly that the literature varies consistently from one specialty to another in characteristics that affect the potential of being cited. The characteristics identified are the size, degree of integration, and age of the literature. Interestingly enough, the size of the field is measured by the size of its core literature, rather than by the number of researchers. Probably the most common misconception held about citation counts is that they vary according to the number of researchers in the field.[7] This is simply not true. Citation potential appears to be an expression of something considerably more complex than the number of people who are theoretically available to cite a paper – though that number does affect the probability of generating extremely high citation rates. It also seems to have much to do with the ratio of publishing authors to total research population, the distribution of published papers over the population, and the distribution of references over the existing literature. While we do not know very much about these variables, it is quite likely that they vary from field to field according to social factors, degree of specialization, and the rate of research progress. In the absence of any detailed knowledge of either the variables or the factors that influence them, the most accurate measure of citation potential is the average number of references per paper published in a given field, and that number does not necessarily correlate with field population.

Reasons aside, citation potential does vary among fields, and the boundaries of fields may be drawn much more finely and narrowly than one might expect. In a controversy over the use of cats in research on the physiology of sexual behavior, Dr. *Lester R. Aronson* claimed that his work on cats may have received relatively few citations because rats are used in most of the research on reproduction, and these people read and cite only studies on rats or other rodents.[32]

Evaluation studies using citation data must be very sensitive to all divisions, both subtle and gross, between areas of research; and when they are found, the study must properly compensate for disparities in citation potential. This can be done very simply. Instead of directly comparing the citation count of, say, a mathematician against that of a biochemist, both should be ranked with their peers, and the comparison should be made between rankings. Using this method, a mathematician who ranked in the 70th percentile group of mathematicians would have an edge over a biochemist who ranked in the 40th percentile group of biochemists, even if the biochemist's citation count was higher.

It can be said that this type of analysis is an involved one, and there is no doubt that it is. But making comparisons across disciplines or specialties is a complicated matter. Presumably, the resons for doing so are significant enough to justify a reasonable amount of effort to make the evaluation fair.

Still another school of criticism discounts citation measures on the grounds that they are too ambiguous to be trusted. One ambiguity that bothers them is that although all Nobel Prize winners and most members of the U.S. National Academy of Sciences have high citation rates,[33] there are other people with equally high rates who have not won this type of peer recognition. And they point out the ambiguities that are inherent in a measure that makes no distinction between a scientist who was cited 15 times a year for two years and one who was cited six times a year for five years. *Crosbie* and *Heckel*[7] make the additional point that citation measures of departmental performance are extremely sensitive to the time period covered by the analysis and can easily produce ambiguous results for multiple time periods; thus, careful interpretation is necessary.

All of these points are perfectly valid ones. There are ambiguities associated with the use of citation counts as a measure of individual performance that prevent them from being completely definitive. They very definitely are an interpretative tool that calls for thoughtful and subtle judgments on the part of those who employ them.

Can citation counts be accurate?

All the rest of the grounds for criticism are concerned with the mechanics of compiling the data. The mechanical weaknesses result from characteristics of the *SCI* and the *SSCI*, the most frequently used sources of citation data, that can affect the accuracy of the citation rate compiled for an individual.

One such characteristic is that the *Citation Index* of *SCI* and *SSCI* lists cited items only by the first author. If you search the *Citation Index* for the cited work of a given scientist, you will find only those publications in which the scientist was listed as first author. Thus, the citation data compiled for that scientist will not reflect work on which he or she was a secondary author. Obviously, this characteristic can affect the accuracy of someone's citation rate.

How greatly this inaccuracy distorts relative citation measurements is a matter of considerable debate.[34] One study by *Cole* and *Cole* showed that the omission of citations to secondary-author publications "does not affect substantive conclusions".[36] However, that study dealt only with physicists.

Lindsey and *Brown* take the opposite position. They theorize that limiting citation counts to primary-author papers does introduce a measurement error if the primary-author papers are a unique subset of an author's publication record; that would be the case if co-author sequence were based on importance of contribution.[37] The size of the error would depend on the extent to which the primary-author papers are not a random, representative sample of all the papers.

Work done by *Roy*[38] suggests that the problem might not exist if the comparisons are being made within and between small, homogenous groups, such as faculty departments. Reasoning that the citation counts for primary- and secondary-authorship papers are either roughly the same, or that any deviation that does exist is constant within the homogeneous universe, he has worked out the following formula for computing total citation counts:

$$\mathrm{CT} = \mathrm{CF} \cdot \frac{\mathrm{TP}}{\mathrm{FP}}$$

where CT – total citations,
CF – citations to primary-author papers,
TP – total papers,
FP – primary-author papers.

If his hypothesis is correct, it may not always be necessary to go to the trouble of compiling the citation data on secondary-author papers to arrive at a total citation count. Using a bibliography to obtain the total number of papers published, it would be possible to calculate the total citation count from primary-author data alone. To test the formula, *Roy* used it to calculate the total citation counts for two faculty departments from a compilation of primary-author data and then compared the counts with compilations of both primary- and secondary-author data. The correlation between the calculated and compiled counts was 0.98 for a materials-science department and 0.94 for one in physics. Much more data is needed, however, to verify the accuracy of the formula.

Another element of uncertainty is introduced by the uneven incidence of multi-authored papers from field to field. Although the trend toward collaborative science and multi-authored papers is a strong one,[39] the *Lindsey* and *Brown* study[37] showed that these papers accounted for only 17–25% of samples of published papers in the fields of economics, social work, and sociology, but 47–81% of samples of published papers in gerontology, psychiatry, psychology and biochemistry.

Preliminary stratification studies at ISI have shown that failure to include secondary-author papers in citation counts introduces a substantial distortion at the very highest stratum of cited scientists. A list of the 250 most-cited scientists,[34] taken solely from a compilation of primary-author counts, had only 28% of its names in common with one taken from a compilation of all-author counts.[35] This latter preliminary study was confined to papers published between 1961–76 in journals received by ISI. Thus its data base was not as broad as the primary-author study which included papers published prior to 1961 and journals cited in *SCI* source journals but not received by ISI. The results of the two studies are, therefore, not directly comparable. Partly due to this, only 77 names from the 250 most-cited primary authors list were represented among the top 300 when each co-author was treated equally.[35] The organic and inorganic chemistry portion of the 300 most-cited scientists list is shown in Table 1. It illustrates that the effect of counting only primary authors is not great for a small portion of these scientists (e.g. *Corey, Olah,* and *Paquette),* significant for most, and of great consequence for some (e.g., *Davidson, Roberts,* and *Witkop).*

Now that we have compiled this all-author file, we can better study patterns of self-citation. Indeed, we are in a position to measure degrees of citation among teams of researchers. Previously, with only first-author data available, a person's self-citations could only be identified when the person was the *first* author of a cited work. Now we can identify citations to publications on which a person was a co-author. We might use this new capability to explore the inbreeding patterns of certain groups who tend to cite each other.

At ISI these stratification studies are conducted on a scale that requires computer compilation but, obviously, the same thing can be done manually in small-

Table 1

Modified organic and inorganic chemistry portion of list from Ref.[35] on the 300 most-cited authors of papers published 1961–1976. Based on *Science Citation Index*® data, the list shows citations for *co-authored* as well as primary-authored articles

Author (Year of Birth)	Total citations	Total papers	Citations as 1st author	1st authored papers	Citations as co-author	Co-authored papers
*Bender M. L. (1924)	5 131	148	3 029	69	2 102	79
*Benson S. W. (1918)	4 359	157	2 239	52	2 120	105
*Brown H. C. (1912)	10 288	400	8 337	289	1 951	111
*Clementi E. (1931)	5 440	92	4 819	61	621	31
*Corey E. J. (1928)	8 500	247	7 646	229	854	18
*Cotton F. A. (1930)	10 292	350	7 664	250	2 628	100
*Cram D. J. (1919)	3 827	164	2 057	58	1 770	106
Davidson E. R. (1936)	3 757	60	436	24	3 321	36
*Dewar M. J. S. (1918)	6 635	224	4 805	168	1 830	56
*Djerassi C. (1923)	11 027	431	2 118	77	8 909	354
Drago R. S. (1928)	4 178	165	984	38	3 194	127
*Flory P. J. (1910)	5 538	133	2 079	49	3 459	84
Grant D. M. (1931)	3 869	90	896	12	2 973	78
Gray H. B. (1935)	4 526	175	988	20	3 538	155
Hammond G. S. (1921)	5 129	141	1 859	38	3 270	103
Hoffmann R. (1937)	7 969	125	5 761	61	2 208	64
*Huisgen R. (1920)	4 996	242	3 965	166	1 031	76
Ibers J. A. (1930)	6 452	209	919	29	5 533	180
Jortner J. (1933)	4 821	197	1 144	42	3 677	155
*Karplus M. (1930)	6 193	128	3 063	25	3 130	103
Khorana H. G. (1922)	6 620	174	770	12	5 850	162
*King R. B. (1938)	4 583	252	3 656	207	927	45
Kochi J. K. (1928)	3 919	159	2 151	55	1 768	104
Li C. H. (1913)	3 908	248	1 212	65	2 696	183
Lipscomb W. N. (1919)	6 364	218	495	20	5 869	198
Muetterties E. L. (1927)	3 883	128	2 193	58	1 690	70
Nemethy G. (1934)	3 927	43	2 214	17	1 713	26
*Olah G. A. (1927)	7 451	380	6 683	346	768	34
Paquette L. A. (1934)	3 819	270	3 448	235	371	35
*Pople J. A. (1925)	10 479	121	6 287	33	4 192	88
*Roberts J. D. (1918)	6 088	196	118	6	5 970	190
Robins R. K. (1926)	4 239	247	167	6	4 072	241
Samuelsson B. (1934)	5 849	148	1 019	27	4 830	121
Scherage H. A. (1921)	9 232	280	315	14	8 917	266
Schleyer P. V. (1930)	5 806	169	1 484	29	4 322	140
Sŏrm F. (1913)	5 858	492	261	17	5 597	475
Stewart R. F. (1936)	3 894	52	3 219	42	675	10
Sweeley C. C. (1930)	4 424	85	2 124	14	2 300	71
*Tanford C. (1921)	5 888	107	1 638	28	4 250	79
*Winstein S. (1912)	4 522	162	1 302	26	3 220	136
Witkop B. (1917)	4 341	194	70	5	4 271	189
*Woodward R. B. (1917)	4 044	48	2 292	24	1 752	24

*Author also appeared in the 250 most-cited primary authors list.[34]

scale studies. If a bibliography is not available, the *SCI/SSCI Source Index,* which shows for each author listed all the items published in the journals covered by the index during the given time period, can be used to compile one. Since the source coverage of *SCI/SSCI* is not exhaustive, this approach does entail some risk of incompleteness. On the other hand, *SCI/SSCI* covers all the highly cited journals; thus, omitted journals items would be those with a relatively low probability of being cited a significant number of times. Nevertheless, because it is always possible that a highly cited paper will be published in a journal that has a low citation rate, the most thorough way of compiling someone's citation rate is by working from a bibliography known to be complete.

Citation analyses for comparative purposes should also take into consideration the impact that co-authorship has on writing productivity. *Price* and *Beaver*[40] found that in their study of collaboration in an information exchange group for a 5-year period four papers was the maximum for authors working alone or with only one co-author. The overall patterns was for productivity to increase as the number of collaborators increased. Thus, collaboration increases productivity, which can affect total citation count. This potential distortion can be handled in a number of different ways, depending upon the situation. In some cases, it may be enough to calculate the average citation count per paper and make that the basis for comparison. *Lindsey* and *Brown* suggest the procedure of allocating equally the citation count of a given paper among all its authors.[37] If the authors and their work are well known by the person doing the evaluation, it may be possible to allocate citation credits among the co-authors of a given paper on the basis of a subjective judgment of their relative contributions.

The second characteristic of *SCI/SSCI* that can affect the accuracy of an individual's citation count is the homograph problem of distinguishing between two or more people with the same last name. An example of the problem is the name *R. A. Fisher,* which identifies both the well known theoretical statistician and a lesser known physicist. The chances are that any annual edition of *SCI* during the past 10 years will list under that name cited works for both the statistician and the physicist.

There are two solutions to the problem, depending upon the size of evaluation study being done. If the study involves few enough people to make it practical to compile their citation counts from the printed index, the distinction between people with the same name frequently can be made by examining the titles of the journals in which the cited work and the citing work were published. For example, the *Citation Index* of the 1974 *SCI* lists 137 cited works under the single name of *J. Cohen,* but an examination of the titles of the cited and source journals involved clearly identifies eight different people: a psychologist, surgeon, physicist, chemist, ophthalmologist, gynecologist and obstetrician, mathematician, and a biostatistician.

The second solution to the name-homograph problem is the same simple one that eliminates the error potential of counts based only on primary-author credits: a complete bibliography of the person being evaluated. This should always be used in any evaluation study large enough to justify a computer analysis of the *SCI/SSCI* data base. And, of course, there is no reason why it cannot be used by researchers on smaller manual analyses to avoid whatever trouble may be involved in matching journal titles against fields of study and to eliminate the small possibility of mistakes or ambiguities in such an operation.

Pros and cons

Any fair appraisal of citation analysis as an aid in evaluating scientists must acknowledge that there is much about the meaning of citation rates that we do not know. We are still imprecise about the quality of scientific performance they measure. We still know very little about how sociological factors affect citation rates. There is still much uncertainty about all the possible reasons for low citation rates. And there is still much to learn about the variations in citation patterns from field to field.

On the other hand, we know that citation rates say something about the contribution made by an individual's work, at least in terms of the utility and interest the rest of the scientific community finds in it. We know that high citation rates correlation with peer judgments about scientific excellence and the importance of contributions. And we know enough about overall citation patterns and the variables that affect them to devise a useful statistical model to predict a scientist's lifetime-citation rate in terms of the average number of citations per paper.

Such a model has been developed and tested by *Geller, de Cani,* and *Davis.*[41] It is based on our knowledge of gross citation patterns and the annual growth of the scientific literature. The input to the model is a citation history, covering at least four years, of all of an individual's existing papers. From this, the model projects the total citation count of each paper for a 40 year time period, which is considered the lifetime of a paper. The average lifetime-citation count per paper is calculated from the aggregate 40 year total. A validation technique is included to identify papers whose history indicates a citation pattern that differs enough from the norm to require special attention.

The development of such a model is an important step forward in systematizing the use of citation data and reducing the incidence of methodological errors. But there is still a need to be careful.[42] There is still a need to understand the limitations of citation data as a measure of relative scientific performance.[43] As with any methodology, citation analysis produces results whose validity is highly sensitive to the skill with which it is applied. The apparent simplicity of counting citations masks numerous subtleties associated with comparing citation counts. Superficial citation studies that ignore this dimension of subtlety can be very misleading. Valid citation studies call for a thorough understanding of the intricacies of making comparisons,[33] particularly when dealing with citation counts that are not extraordinarily high.

Finally, two elements stand out as being fundamental to the debate on citation rates as an evaluation tool. One is that as the scientific enterprise becomes larger and more complex, and its role in society more critical, it presumably becomes more difficult, more expensive, and more necessary to make the evaluations that identify those people and groups who are making the greatest contribution. The second is that citation measures have been demonstrated to be a valid form of peer judgment that introduces a useful element of objectivity into the evaluation process and involves only a small fraction of the cost of surveying techniques. While citation analysis may sometimes require significantly more time and effort than judgments made on nothing but intuition, professional evaluations certainly are important enough to justify such an investment.

References

1. A. A. JOHNSON, R. B. DAVIS. The Research Productivity of Academic Materials Scientists, *Journal of Metals* 27 (1975) No. 6, 28–29.
2. D. SHAPLEY. NSF: A 'Populist' Pattern in Metallurgy, Materials Research? *Science,* 189 (1975) No. 4203, 622–624.
3. C. A. WERT. The Citation Index Revisited, *Journal of Metals* 27 (1975) No. 12, 20–22.
4. T. GUSTAFSON, The Controversy Over Peer Review, *Science,* 190 (1975) No. 4219, 1060–1066.
5. D. SHAPLEY, Materials Research: Scientists Show Scant Taste for Breaking Ranks, *Science,* 191 (1976) No. 4222, 53.
6. R. ROY, Comments on Citation Study of Materials Science Departments, *Journal of Metals,* 28 (1976) No. 6, 29–30.
7. G. C. CROSBIE, R. W. HECKEL, Citation Criteria for Ranking Academic Departments, *Journal of Metals,* 28 (1976) No. 9, 27–28.
8. N. ARBITER, Letter to the Editor, *Journal of Metals,* 28 (1976) No. 12, 33.
9. J. C. AGARWAL, et al., Letter to the Editor, *Journal of Metals,* 28 (1976) No. 12, 33.
10. C. J. ALTSTETTER, Letter to the Editor, *Journal of Metals,* 28 (1976) No. 12, 33–35.

11a. N. WADE, Citation Analysis: A New Tool for Science Administrators, *Science,* 188 (1975) No. 4187, 429–432.

11b. Citation Analysis. Letters in response to Wade by M. Klerer, H. J. M. Hanley, J. Arditti, R. E. Machol, *Science,* 188 (1975) No. 4193, 1064. [N. WADE, Citation Analysis: A New Tool for Science Administrators, *Science,* 188 (1975) No. 4187, 429–432.]

12. G. P. KOSHY, The Citeability of a Scientific Paper. *Proc. of Northeast Regional Conference of American Institute for Decision Sciences,* Philadelphia, Pa., April/May 1976, 224–227.

13. E. GARFIELD, Is the Ratio Between Number of Citations and Publications Cited a True Constant? *Essays of an Information Scientist,* Vol. 2, Philadelphia: ISI Press, 1977, 419–421.

14. E. GARFIELD, Will ISI's *Arts & Humanities Citation Index* Revolutionize Scholarship? *Current Contents,* No. 32 (August 8, 1977) 5–9.

15. I. H. SHER, E. GARFIELD, New Tools for Improving and Evaluating the Effectiveness of Research, *in Research Program Effectiveness,* M. C. YOUITS, D. M. GILFORD, R. H. WILCOX, E. STAVELY, H. D. LEMER, (Eds). New York; Gordon and Breach, 1966, p. 135–146.

16. J. P. MARTINO, Citation Indexing for Research and Development Management, *IEEE Transactions on Engineering Management,* EM-18 (1971) No. 4, 146–151.

17. A. E. BAYER, J. FOLGER, Some Correlates of a Citation Measure of Productivity in Science, *Sociology of Education,* 39 (1966) 381–390.

18. J. A. VIRGO, A Statistical Procedure for Evaluating the Importance of Scientific Papers, *Library Quarterly* 47 (1977) No. 4, 415–430.

19. E. GARFIELD, Citation Indexing for Studying Science, *Nature,* 227 (1970) No. 5259, 669–671.

20. C. L. BERNIER, W. N. GILL, R. G. HUNT, Measures of Excellence of Engineering and Science Departments: A Chemical Engineering Example, *Chemical Engineering Education,* 9 (1975) 94–97.

21. G. M. CARTER, Peer Review, Citations, and Biomedical Research Policy: NIH Grants to Medical School Faculty, *Rand Report,* R-1583-HEW, Santa Monica, California; Rand Corporation, 1974.

22. A. J. MEADOWS, *Communication in Science,* London Butterworths, 1974, p. 45.

23. H. G. SMALL, *Characteristics of Frequently Cited Papers in Chemistry,* Final Report on NSF Contract #C795, 1974. See also E. GARFIELD, A list of 100 Most Cited (Chemical) Articles, *Current Contents,* No. 10 (March 9, 1977) 5–12.

24. O. H. LOWRY, N. J. ROSEBROUGH, A. L. FARR, R. J. RANDAL, Protein measurement with the Folin phenol reagent, *Journal of Biological Chemistry,* 193 (1951) 265–275.

25. O. H. LOWRY, Personal communication to D. de S. PRICE, November 11, 1969.

26a. H. G. SMALL, B. C. GRIFFITH, The Structure of Scientific Literatures, I: Identifying and Graphing Specialties, *Science Studies,* 4 (1974) 17–40.

26b. B. C. GRIFFITH, H. G. SMALL, J. A. STONEHILL, S. DEY, The Structure of Scientific Literatures, II· Towards a Macro- and Microstructure for Science, *Science Studies* 4 (1974) 339–365.

27. R. K. MERTON, *On the Shoulders of Giants-A Shandean Postcript,* New York, Harcourt Brace & World, 1965, p. 218–219.

28. J. LEDERBERG, A View of Genetics, *Science,* 131 (1960) 269–276.

29. E. GARFIELD, I. H. SHER, Citation Indexes in Sociological and Historical Research, *American Documentation,* 14 (1963) 289–291.

30. F. NARIN, *Evaluative Bibliometrics: The USE of Publication and Citation Analysis in the Evaluation of Scientific Activity,* Cherry Hill, N.J.; Computer Horizons, Inc., 1976, p. 500, NTIS-PB252339/AS.

31. J. R. COLE, S. COLE, The Ortega Hypothesis, *Science,* 178 (1975) 368–375.

32a. E. GARFIELD, Citation Analysis and the Anti-Vivisection Controversy, *Current Contents,* No. 17 (April 25, 1977) 5–10.

32b. E. GARFIELD, Citation Analysis and the Anti-Vivisection Controversy. Part II. An Assessment of Lester R. Aronson's Citation Record, *Current Contents*, No. 48 (November 28, 1977) 5–14.
33. E. GARFIELD, The 250 Most-Cited Primary Authors, 1961–1975. Part II. The Correlation Between Citedness, Nobel Prizes, and Academy Memberships, *Current Contents*, No. 50 (December 12, 1977) 5–15.
34. E. GARFIELD, The 250 Most-Cited Primary Authors, 1961–1975. Part III. Each Author's Most Cited Publication, *Current Contents*, No. 51 (December 19, 1977) 5–20.
35. E. GARFIELD, The 300 Most-Cited Authors, 1961–1976, Including Co-Authors at Last. Part I. How the Names Were Selected, *Current Contents*, No. 28 (July 10, 1978) 5–18.
36. J. R. COLE, S. COLE, *Social Stratification in Science*, Chicago; University of Chicago Press, 1973, p. 32–33.
37. D. LINDSEY, G. W. BROWN, Problems of Measurement in the Sociology of Science: Taking Account of Collaboration, (unpublished, 1977).
38. R. ROY, Approximating Total Citation Counts From First-Author Counts and Total Papers, Working Paper, April 1977.
39. D. de S. PRICE, *Little Science, Big Science*, New York, Columbia University Press, 1963.
40. D. de S. PRICE, D. B. BEAVER, Collaboration in an Invisible College, *American Psychologist*, 21 (1966) 1011–1018.
41a. N. L. GELLER, J. S. CANI, R. E. DAVIES, Lifetime-Citation Rates as a Basis for Comparisons Within a Scientific Field, *Proc. of the American Statistical Association. Social Statistics Section*, Washington, D. C., 1975. p. 429–433.
41b. N. L. GELLER, J. S. de CANI, R. E. DAVIES, Lifetime-Citation Rates to Compare Scientists' Work, *Social Science Research*, (in press).
42. E. GARFIELD, Caution Urged in the Use of Citation Analysis, *Trends in Biochemical Sciences*, 2 (1977) No. 4, N84.
43. D. LINDSEY, G. W. BROWN, The Measurement of Quality in Social Studies of Science: Critical Reflections on the Use of Citation Counts, (unpublished, 1977).