The Skewness of Science

Per O. Seglen
Institute for Studies in Research and Higher Education, Munthes gt. 29, 0260 Oslo, and Institute for Cancer Research, The Norwegian Radium Hospital, Montebello, 0310 Oslo, Norway

Scientific publications are cited to a variable extent. Distributions of article citedness are therefore found to be very skewed even for articles written by the same author, approaching linearity in a semilog plot. It is suggested that this pattern reflects a basic probability distribution with some similarity to the upper part of a normal (Gaussian) distribution. Such a distribution would be expected for various kinds of highly specialized human activity, parallels being found in the distribution of performance by top athletes and in the publication activity of university scientists. A similar skewness in the distribution of mean citedness of different authors may combine with the variability in citedness of each author's articles to form a two-leveled citational hierarchy. Such a model would be capable of accounting for the extremely skewed distribution of citedness observed for all articles within a scientific field, which approaches linearity in a double-log rather than in a semilog plot.

The skewness implies that there will always be a large fraction of uncited publications, the size of the fraction depending on the citation practices (such as the number of references per publication) within the field in question. However, as part of a continuous probability distribution even uncited articles have a definite probability of contributing to scientific progress. Since it is furthermore impossible to eliminate uncited articles for statistical reasons, they should be the cause of neither worry nor remedy.

The citational variability between articles in a journal is less (semilog linearity) than in the corresponding field as a whole, suggesting that each journal represents a select, stratified sample of the field. However, the variability is still too large to make the journal impact factor (the average citedness of the journal's articles) suitable as a parameter for evaluation of science. Fifteen percent of a journal's articles collect 50% of the citations, and the most cited half of the articles account for nearly 90% of the citations. Awarding the same value to all articles would therefore tend to conceal rather than to bring out differences between the contributing authors.

The skewness in the citedness distribution of each author's articles, the large overlap between different authors and the existence of field-dependent systematic differences in citedness would seem to make even article citations unsuitable for evaluation of individual scientists or research groups. At the national level, citations may be more useful, provided due corrections are made for the field effects.

Introduction

It is well known that scientists differ greatly in productivity and scientific influence. In 1926, Lotka (1926) observed that the productivity distribution of scientists was extremely skewed, and formulated his famous "inverse square" law, stating that the number of scientists producing n papers is approximately proportional to $1/n^2$. Subsequent studies, which, like Lotka's, have been based on samples of indexed scientific literature, have confirmed the skewness of productivity distributions and shown that they can be well described by inverse power law functions, although the exponent may vary from sample to sample (Coile, 1977; Pao, 1986). Since deviations from an inverse power law distribution are usually observed in the region of the most productive authors (Price, 1963), various sophisticated distribution functions have been employed to achieve a more accurate description of the empirical bibliometric distributions, such as the generalized inverse Gaussian-Poisson distribution (Sichel, 1985) or the (generalized or two-parameter) Waring distribution (Glänzel & Schubert, 1985; Irwin, 1975). Several hypotheses have been invoked to account for the skewed distribution of scientific productivity, including differential ability or motivation, resource inequality, cumulative advantage, etc. (Price, 1976; Shockley, 1957), but supportive evidence for any of these is largely missing or ambiguous (Kyvik, 1991).

Citedness has not been subject to the same mathematical scrutiny as scientific productivity, but the studies available indicate that distributions of citations to scientific papers may be as skewed as productivity distributions (Folly, Hajtman, Nagy, & Ruff, 1981; Magyar, 1973; Naranan, 1971; Price, 1965). Citation data can be obtained from the Science Citation Index (SCI), an index of all references (citations) to the previous literature

Received March 16, 1992; revised May 25, 1992; accepted May 25, 1992.

given each year in the 3500 most important journals of science. The citing articles are listed under the first-author name of the cited article (alphabetically arranged), with bibliographic information provided both for the cited and citing article. The annual number of citations given to each cited article can thus be counted. In addition to the annual index, cumulative five-year indices are published, which may cover a considerable proportion of articles cited less than once a year.

Figure 1A shows the distribution of citations to a random sample of articles drawn from the 1985–1989 SCI (for visibility reasons only the lower end of the distribution is shown). To minimize age-dependent variability, only third-year citations were recorded; nevertheless, it is evident that the distribution is very skewed, with more than 50% of the articles being uncited. A double-log plot of the data over the full range provides reasonably good fit to a straight line (Fig. 1B; correlation coefficient = −0.99). As measured by this bibliometric sampling procedure, article citedness thus seems, like author productivity, to approximate an inverse power law distribution (Naranan, 1971). Which could be the factors responsible for this extraordinary skewness?

Does the Article Age Distribution Contribute to the Skewness?

The citedness of scientific articles changes with their age: after an early (third-year) peak (Seglen, 1989d), citedness declines steadily as a function of time since publication, probably reflecting the gradual obsolescence of the article contents (individual articles may of course vary greatly in their citational durability) (Moed, Burger, Frankfort, & Van Raan, 1987). Price (1965) has suggested that this obsolescence may, at least in part, account for the skewness of citation distributions. Relative to a fixed citation "window" (citations made during a single year or, e.g., during a defined five-year period), the variability in citedness would be expected to increase with the size of the publication window; i.e., the number of publication years to which citations are recorded. The data presented in Figure 1B, where both the citation window and the publication window were kept as narrow as possible (single years, three years apart), suggest, however, that article citedness may form skewed distributions independently of article obsolescence.

This issue was examined more directly by varying the width of the publication window. The period 1980–1984 was used as a constant citation window (having first established that a five-year citation window introduced no more variability than a one-year window) for collecting citations to articles from a single year (1979), from the five-year period 1980–1984, or from all years up to 1984. Somewhat surprisingly, the citation distributions were similar for all three publication windows (Fig. 2). The variability contributed by a broad article age distribution would thus seem to be negligible in comparison with other sources of variability.

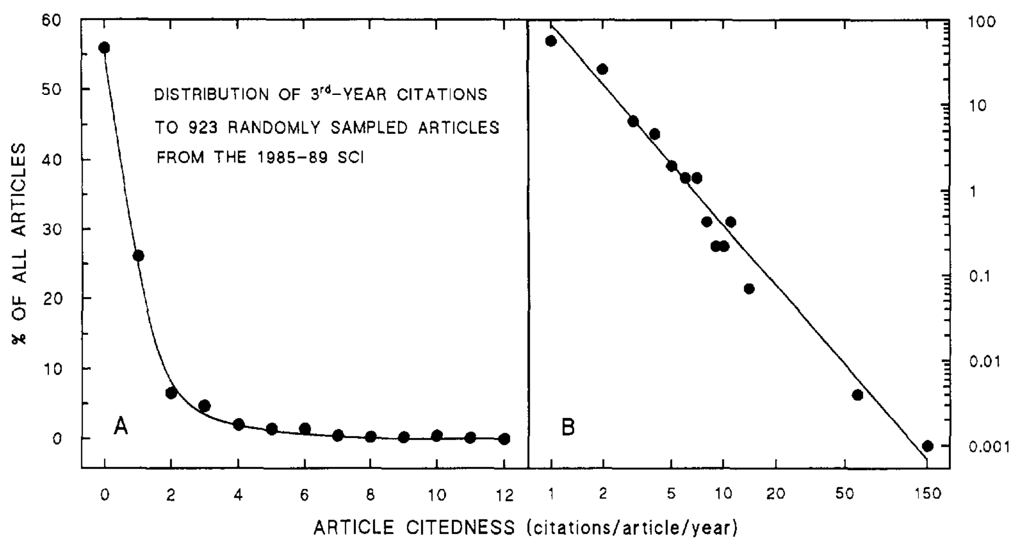

FIG. 1. Distribution of third-year citations to articles in a random SCI sample. A total of 923 articles were drawn from the 1985–1989 SCI (one random page from every second volume of the 38 volumes), and the number of citations to each article during the third year after publication was recorded. The percentages of articles in each citation cohort (receiving 0, 1, 2, etc. citations) were arranged in order of increasing citedness. (A) Linear plot. Only the low-end part of the distribution (up to 11 citations/article/year) is shown. (B) Double-logarithmic plot covering the whole data range. To avoid the problem of zero values on a log scale, the citation cohorts have now been defined as "less than," i.e., the uncited articles are in cohort 1. For citedness values above ten (where only a few of the integer cohorts are filled) the cohort size has been adapted symmetrically to the articles found, with a mean value (number of articles per integer-cohort) plotted in the middle of the defined cohort.

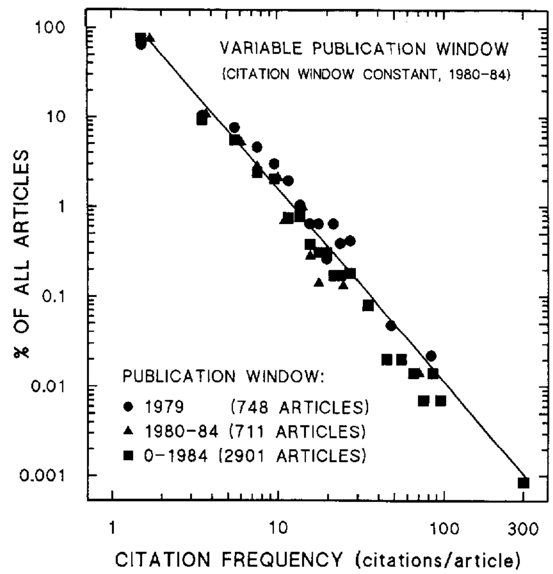

FIG. 2. Effect of publication window on distribution of article citedness. A constant citation window (1980–1984) was used to collect SCI citations to random samples of articles from 1979 (748 articles, ●), 1980–1984 (711 articles, ▲), or from all years up to 1984 (■). Citation cohorts and data plotting in a double-logarithmic diagram as in Figure 1. A few of the triangular symbols have been slightly displaced horizontally for better visibility.

Distribution of Citations to Articles From Single Journals

One way to identify sources of citational variability may be to examine well circumscribed article assemblies such as scientific journals. It has been shown that the variability in citedness between different journals is no less than that between individual articles (Naranan, 1971). As shown in Figure 3, the distribution of journals according to citedness is extremely skewed, producing a straight line in a double-log plot. Nevertheless the average citedness of each journal has been found to be remarkably constant from year to year, justifying the assignment of a "journal impact factor" value (mean annual citedness of the journal articles) to each journal (Garfield, 1987). It might thus seem as if the scientific literature is neatly stratified into citational cohorts by the journals.

To see just how homogeneous these citational cohorts are, citations to articles sampled from three different biochemical journals, differing considerably in mean citedness, were recorded. The journals were *J. Biol. Chem.* (journal impact factor 6.4; impact of sample 6.9); *Biochem. J.* (journal impact factor 3.8; impact of sample 4.2), and *Biochim. Biophys. Acta* (journal impact factor 2.4; impact of sample 2.0). As shown in Figure 4, citations to all three journals formed remarkably skewed distributions, uncited or once-cited articles representing the largest citation cohort in all journals. By normalizing the citation cohorts to fractions of the journal mean, it became evident that all journals exhibited the same intrinsic variability in article citedness, independently of the journal impact (Fig. 5). However, the variability within journals was less than observed for the bibliometric distributions of all articles (Figs. 1 and 2), the data conforming much better to linearity in a semilog plot (Fig. 5A) than in a double-log plot (Fig. 5B). The journals would thus seem to impose some degree of citational stratification upon the article population.

Although the most cited articles constitute a minor fraction of the journal contents, they contribute heavily to the journal impact. Figure 6 is a cumulative contribution function that gives the percentage of journal cita-

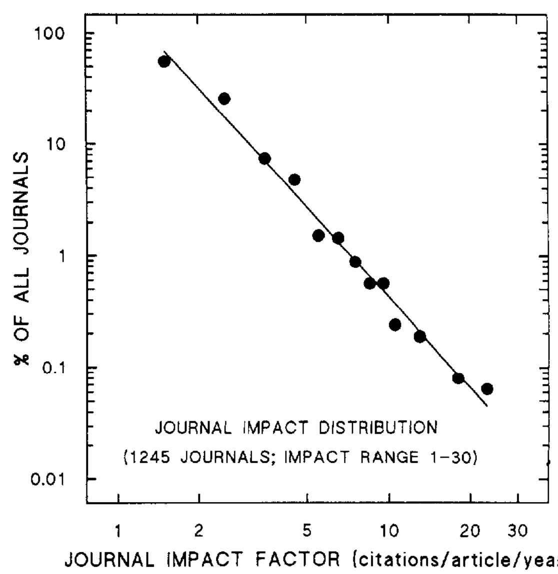

FIG. 3. Citedness distribution of journals. Journals indexed in the SCI were enumerated in broad citation cohorts (1.00–1.99, 2.00–2.99, etc.) on basis of their mean citedness (journal impact factor) (Garfield, 1987), and plotted in a double-logarithmic diagram.

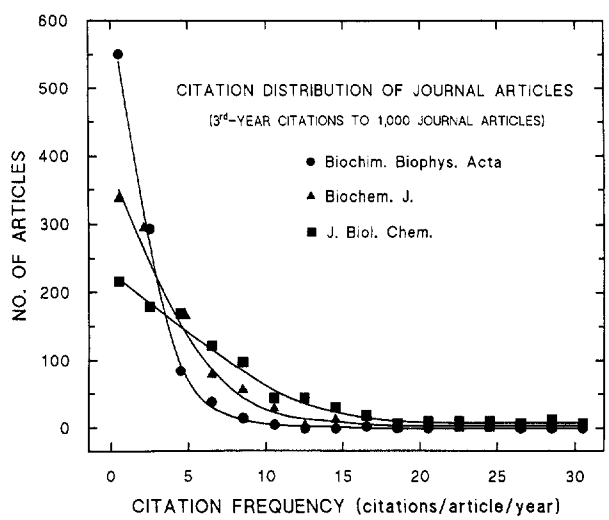

FIG. 4. Distribution of citations to journal articles. Third-year citations (in SCI) to *Biochim. Biophys. Acta* (general subjects) 1984, all issues (323 articles, ●); *Biochem. J.*, 1983, vols. 211–214 (447 articles, ▲); and *J. Biol. Chem.*, 1984, issues 11–14 (402 articles, ■) were normalized to 1000 articles and enumerated in citation cohorts 0–1, 2–3, 4–5, etc. One of the triangular symbols has been slightly displaced horizontally for better visibility.

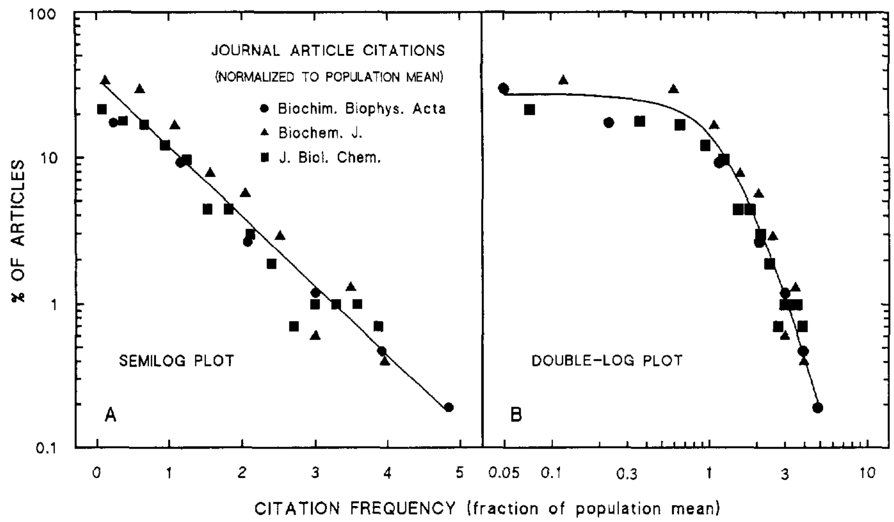

FIG. 5. Semilogarithmic distribution of journal article citedness. The citation data from Figure 4 were enumerated in citation cohorts defined as fractions of the population mean (mean sample impact), and plotted in a semilogarithmic (A) or a double-logarithmic (B) diagram. (●), *Biochim. Biophys. Acta;* (▲), *Biochem. J.;* (■), *J. Biol. Chem.*

tions as a function of an increasing fraction of the articles, starting with the most cited article cohort. The contribution function is very similar for all journals investigated (Seglen, 1991); data for the three biochemistry journals previously discussed have therefore been combined. The figure shows that 15% of the articles collect 50% of the citations, and that the most cited half of the articles account for about 90% of the citations given to the journal.

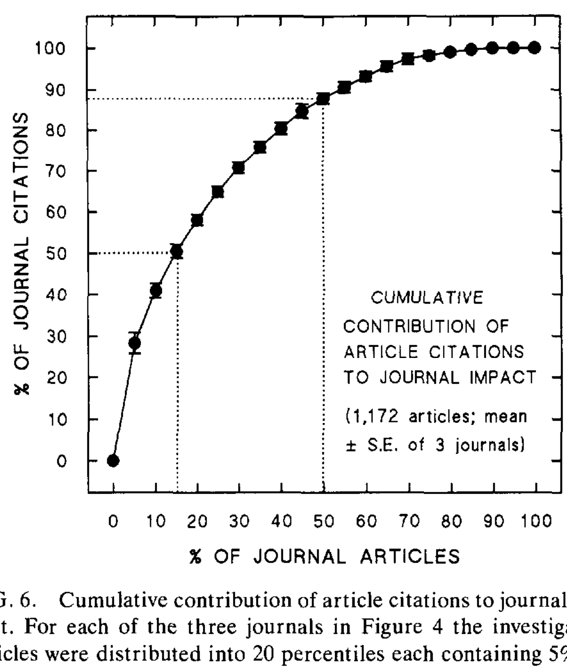

FIG. 6. Cumulative contribution of article citations to journal impact. For each of the three journals in Figure 4 the investigated articles were distributed into 20 percentiles each containing 5% of the total article number, arranged in order of decreasing citedness. The number of third-year citations in each cohort was recorded and expressed as percent of the total number of citations given to the journal sample. Each value is the mean ±SE of the three journals. The dotted lines indicate that 50% of the citations are given to 15% of the articles and almost 90% to 50% of the articles.

The great variability in citedness within a journal has important implications for the significance attached to the journal impact factor. In several countries, this easily available factor has been used in academic evaluations of individual scientists, on the implicit premise that the impact factor of the journal is representative of its constituent articles, and hence, of the article authors. The skewness of the journal article distributions shows that this premise does not hold true: only a minor fraction of the articles are cited anywhere near the journal mean. Studies of citations to articles written by the same author but published in different journals have furthermore revealed that the correlation between journal impact and real article citedness usually is poor (Seglen, 1989b, 1989c). Assigning the same value to all articles in a journal will overestimate the less influential and underestimate the more influential articles, thus effectively leveling out the very differences that evaluation procedures should seek to identify.

Distribution of Citedness of Articles Written by the Same Author

While individual journals still display a large variability in article citedness, the individual scientific author would be expected to represent the ultimate in citational homogeneity. Nevertheless, the few published examples of citations to articles by individual authors indicate that these article distributions are also skewed (Folly et al., 1981; Seglen, 1991). To obtain more representative material, an analysis of citations to all articles by 13 productive senior scientists in a major Norwegian biomedical institute was undertaken, based on complete publication lists provided by the authors. Examination of individual distribution profiles (Fig. 7) indicated a considerable degree of skewness in all cases. By express-

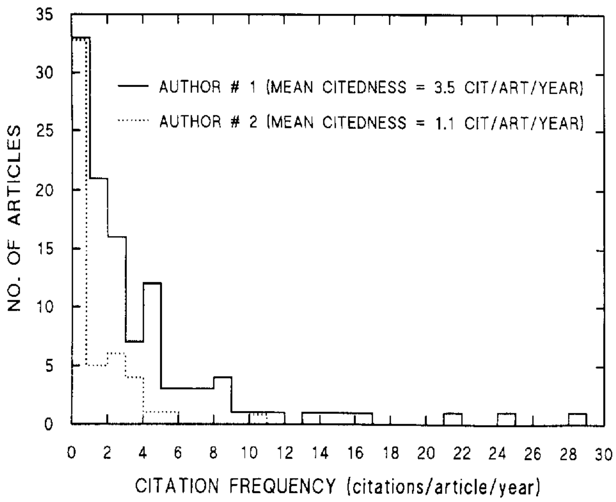

FIG. 7. Distribution of article citedness, individual authors. Second- to fourth-year citations to all journal articles by two biomedical scientists (whole and dotted lines, respectively) were recorded, and the number of articles per citation cohort (0, 1, 2, etc.) enumerated.

ing each distribution in terms of its mean value, the different individual distributions could be averaged. Figure 8A shows the resulting mean distribution, which has relatively small standard errors and which should be representative of articles by individual authors in general (at least at this level of productivity and citedness). The distribution is highly skewed, generating a straight-line in a semilog plot (Fig. 8B).

There would thus seem to be a fundamental, irreducible variability in article citedness, which is independent even of the publishing author. The skewed distribution observed should probably be regarded as the basic probability distribution of citations, reflecting both the wide range of citedness values potentially attainable and the low probability of achieving a high citation rate.

Distribution of Citations Within a Field: The Transition From Semilogarithmic to Double-Logarithmic Linearity

Given the semilog distribution of article citedness when individual authors or journals are examined, how is the transition to the double-log linearity of article distributions in general (Figs. 1 and 2) achieved, and at what organizational level does it occur? An examination of a well-defined scientific field may be useful: Magyar's (1973) analysis of dye laser research shows that the distribution of article citedness within this field conforms reasonably well to linearity in a double-log plot (Fig. 9). Thus, within a relatively small research field (454 papers), the transition from semilog to double-log linearity has already occurred. The same distribution was found by Naranan (1971) within the much larger field of genetics.

The scientific field can be considered as an assembly of authors, each of whom contributes a collection of articles with a semilog distribution of citedness. Since the authors differ in mean citedness they will form another distribution at a higher hierarchical level, thus increasing the overall citedness variability in a multiplicative manner. Unfortunately, clean distributions of author citedness have apparently not been worked out (our own group of thirteen scientists being too small to form a distribution), but the analysis by Cole and Cole (1973) of the lifetime citedness of a sample of U.S. physicists

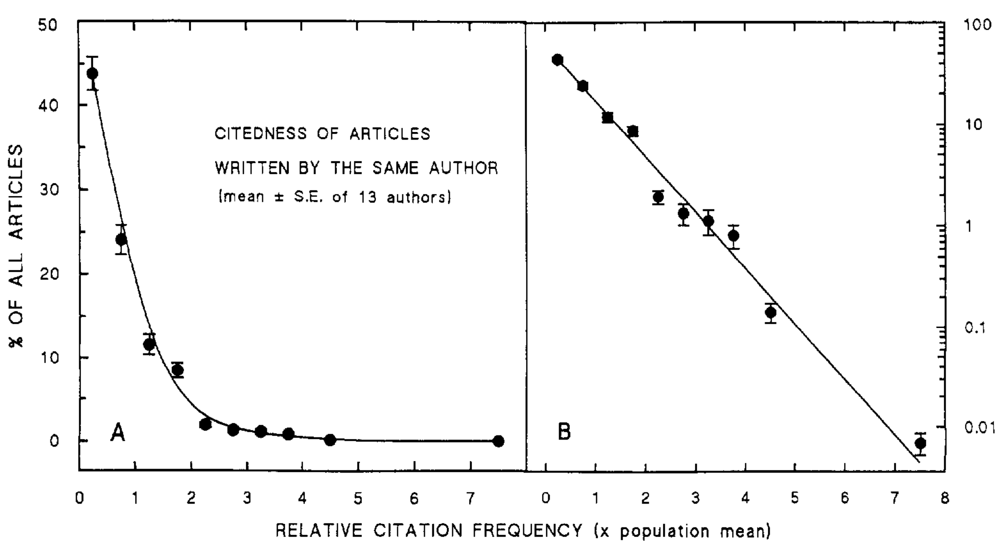

FIG. 8. Distribution of citedness of articles written by the same author. Second- to fourth-year citations to all articles by each of 13 authors were enumerated in citation cohorts defined as fractions of the author's distribution mean (mean article citedness). The number of citations in each cohort was expressed as percent of all citations given to the author and plotted in a linear (A) or semilogarithmic (B) diagram. Each value is the mean ±SE of the 13 authors. Some of the SEs are hidden by the symbols.

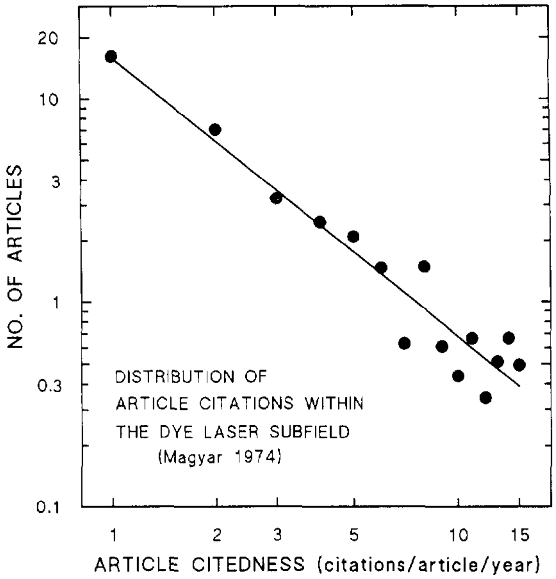

FIG. 9. Distribution of article citedness within a small scientific field. Data from Magyar's (1973) study of the dye laser field (publications and citations within the same six-year period) have been plotted in a double-logarithmic diagram.

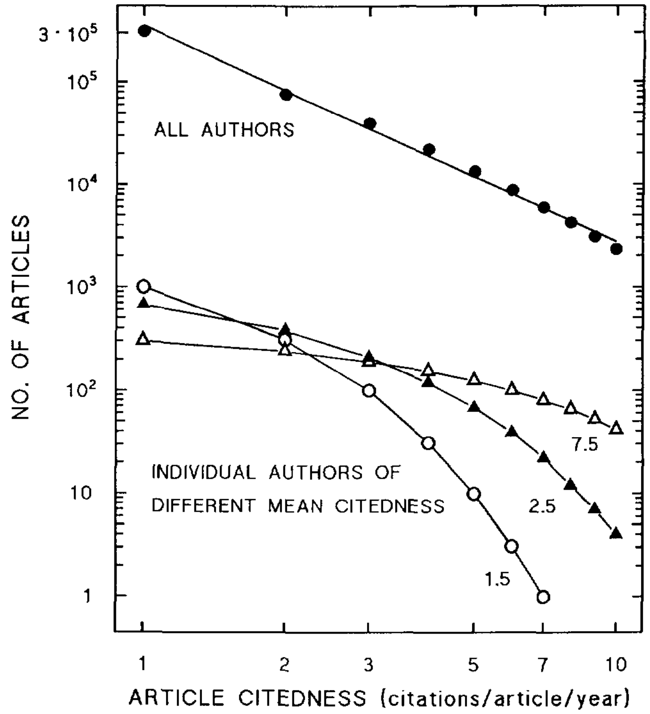

FIG. 10. Transition from semilog to double-log linearity in a two-level hierarchical model of the scientific field. A hypothetical scientific field was constructed, consisting of 3333 authors semilogarithmically distributed (1000, 700, 490, etc.) into 19 citation cohorts corresponding to mean citedness values from 1 to 10 (1.0, 1.5, 2.0, etc.). Each author was assumed to have published 145 articles, distributed semilogarithmically, with regard to citedness. The citedness distribution of articles by different authors (○, ▲, △) as well as the distribution of all articles in the field (●) is shown.

may provide a clue. The physicists were found to form a highly skewed distribution, approaching linearity in a double-log plot. However, since lifetime citedness incorporates productivity as well as citedness per se, the recorded variability would be higher than for citedness alone, corresponding to a combination of two independent, skewed distributions.

Purely bibliometric (database-derived) distributions of author productivity alone are generally highly skewed, conforming to linearity in double-log plots (Coile, 1977; Lotka, 1926; Pao, 1986), but in the case of defined populations of scientific staff in universities or research institutes distributions approaching linearity in semilog plots have been observed (Kyvik, 1989; Shockley, 1957). If we assume that author citedness similarly has a semilog distribution, would our two-level hierarchical model of the scientific field achieve the transition from semilog to double-log linearity in the distribution of article citedness? The answer seems to be a qualified *yes.* Figure 10 indicates that if both the mean citedness values of the authors themselves and the citedness of the articles written by each author are assumed to have semilog distributions, the citedness distribution of all articles by all authors in the field will approach linearity in a double-log plot. The more skewed the author distribution, the better the fit with linearity, an inverse power law distribution providing a perfect fit.

While authors and authors' articles can be taken to represent actual, functional units of the scientific system, it should be noted that a complex citational field can be also be stratified (or, perhaps more appropriately, sliced) according to other criteria. One approach might be to follow the process of citation (i.e., the citers) rather than the property of citedness (i.e., the citees); another to follow, e.g., journals through various levels of organization. As we have already seen, journals behave very much like scientists: within the whole realm of science, journal citedness shows a very skewed (double-log) distribution (Figs. 1, 2) that extends even to circumscribed scientific fields (Fig. 9), yet within each journal, the variability has been reduced to a semilog distribution of article citedness (Fig. 5A). Reduced variability may thus be achieved by a reduction in system complexity that does not necessarily correspond to a functional dissection: journals represent select, stratified samples rather than functional units of scientific activity.

Possible Model Distributions of Citedness: Topnormal and Extreme-Property Distributions

Our operational distinction between semilogarithmic and double-logarithmic linearity may help to indicate major differences in the causative basis of distributions, but it should be remembered that these plots are nothing but crude approximations of actual probability distributions. The citation distributions present striking parallels to productivity distributions, and it is therefore likely that the most complex of them can be well represented mathematically, e.g., by the Waring distribution (Braun, Glänzel, & Schubert, 1990). However, it might be helpful if the buildup of complex citation distributions from simpler ones could be explained in terms of probability distributions which were more tractable sociologically, even at the risk of sacrificing some mathematical precision for the sake of conceptual simplicity.

The normal (Gaussian) distribution may serve as a convenient starting point: this probability distribution, generated by arithmetically additive causes of variability, is familiar to most of us in the form of a bell-shaped curve, e.g., body height distributions (Fig. 11). At first glance, the normal distribution would seem to bear little resemblance to the skewed distributions we have been considering so far, but if we consider only a part of it, e.g., the probability of being taller than 1.85 m, it is evident that this upper part of the normal distribution forms a skewed distribution curve, which in fact approaches linearity in a semilog plot. Therefore, it should be considered as a general possibility that certain phenomena which represent an extreme kind of human property or ability may tend to form semilog distributions resembling (although not necessarily being identical to) the "topnormal" distribution, i.e., the upper part of the normal distribution.

An an additional illustration of the concept we may look at short-distance running, a specialized capability which is likely to approach a normal distribution in the general human population. If we plot a distribution of the world's best short-distance runners we obtain a linear semilog curve, consistent with an "extreme-property" type of probability distribution (Fig. 12).

Science may be the prime example of an extreme type of human effort, where numerous mental capabilities interact to generate scientific creativity (Shockley, 1957). Although scientific ability may well be normal-distributed in the general population, scientists are

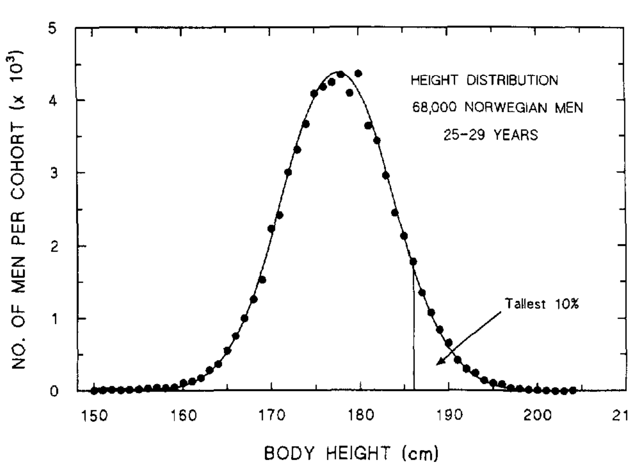

FIG. 11. The normal distribution of body height. Tables showing the body heights of 68,000 Norwegian men aged 25–29 years (~40% of the total age cohort) were obtained from the Norwegian Cancer Registry and plotted in a linear diagram. The tallest 10%, forming a skewed distribution, are indicated.

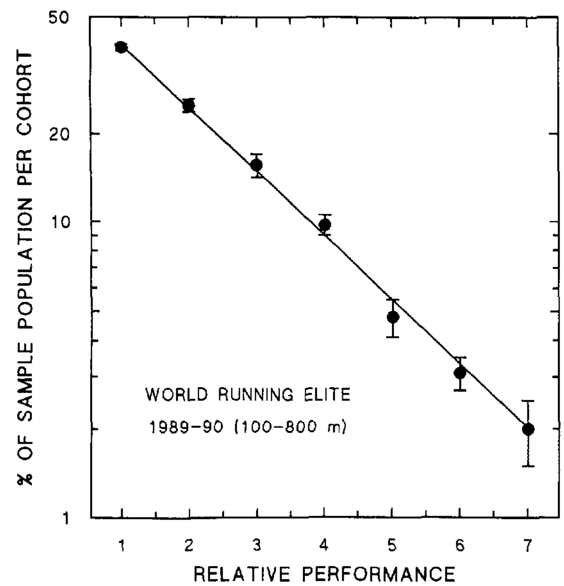

FIG. 12. Distribution of athletic performance of the world running elite. The personal best results of the ~150 top runners at each of the distances 100 m, 200 m, 400 m, and 800 m for the years 1989 and 1990 were taken from the *Athletics* yearbooks (1990, 1991) and distributed into seven equally wide time (performance) cohorts, cohort 7 representing the best performance. The number of runners in each cohort was recorded and expressed as percent of all runners listed for that distance. Each value is the mean ±SE of the eight distributions assembled. Some of the SEs are hidden by the symbols.

likely to form an extreme-property distribution with regard to their speciality (just like the short-distance runners), be it in terms of citedness or in terms of productivity. It should be emphasized, incidentally, that although these two indicators of scientific ability may behave similarly in the statistical sense, at the individual level they are not necessarily coincident with each other (Moed et al., 1985; Dewitt, Nicholson, & Wilson, 1980), nor with scientific quality in general.

As an alternative to time-independent multiple causation it has been suggested that skewed scientometric distributions may be generated by sociological reinforcement mechanisms operating over time, on the principle of "cumulative advantage" (Cole & Cole, 1973) or "the Matthew Effect" (Merton, 1968). While such a model can in theory generate the required skewness (an inverse power law distribution) (Price, 1976), the evidence for the actual operation of cumulative advantage is controversial (Kyvik, 1991).

Like the probability of possessing scientific talent, the probability of having an article highly cited is low (Dieks & Chang, 1976). A very large number of factors influence the extent to which an article is cited (MacRoberts & MacRoberts, 1989; Seglen, 1989a); it is therefore no surprise that the probability declines rapidly with increasing citedness, resulting in an extreme-property type of probability distribution. The distributions of author article citedness, author productivity, and athletic top performance are strikingly similar to the distribu-

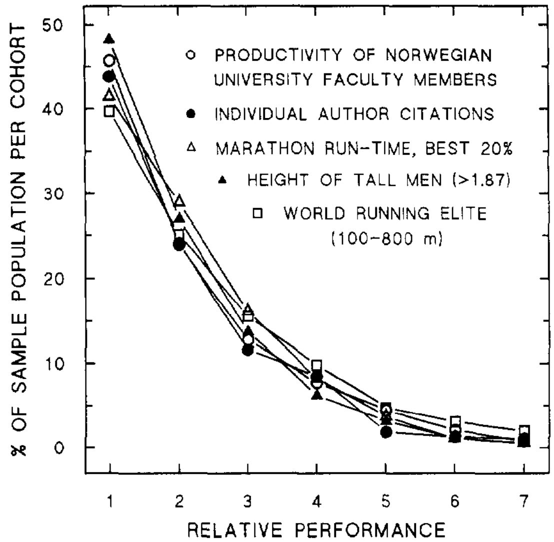

FIG. 13. Distributions of scientific and physical performance. The following kinds of activity/property were distributed into seven equally wide performance cohorts (cohort 7 representing the best performance). (○), Publication productivity of Norwegian university staff (Kyvik, 1991); (●), citations to individual authors' articles; (△), best 20% male half-marathon runners (17–44 years), 1991 Oslo Marathon; (▲), height of Norwegian men (25–29 years) >1.87 m; (□), world running elite 100–800 m.

tion of physical tallness (Fig. 13), suggesting that the extreme-property distribution and the topnormal distribution may be closely related.

The attractiveness of athletics as an analogy to science lies in its ability to illustrate how highly deterministic individual efforts must yield to the stochastic laws of the population. No matter how hard we train, how much we publish, or how often we are cited, the statistical distribution to which we belong remains the same. If someone moves up through the distribution, others must move down. This statistical inertia should not, however, be taken as a discouragement: the purpose of scientific effort is not to alter the skewed shape of the "body of science," but to help it move forward.

Are Uncited Articles Dispensable?

One consequence of the skewness of citedness distributions is that a large fraction of the published work within any field of science will be cited very little or not at all. The size of the uncited fraction depends primarily on the referencing practices of the field: if the number of references per published item and/or the emphasis on current work is low (Moed et al., 1985), most of the recent literature will be uncited. More than 99% of the papers, e.g., in architecture or theatrical science, thus remained uncited during the first five years after publication; for the SCI database as a whole the figure was 55% (Hamilton, 1991).

The large fraction of uncited papers has been the cause of some concern both among sociologists and administrators of science, on the assumption that these papers might represent research of low quality that could be dispensed with at little loss to the scientific community (Cole, 1983; Hamilton, 1990). Therefore, it ought to be pointed out that uncitedness should be cause of neither worry nor remedy. The fact that a minority of the published articles collect the majority of citations is an inevitable statistical phenomenon that cannot be altered by intervention. A reduction in the size of the uncited fraction can only be achieved by raising the number of references per publication. If the least-cited papers or authors were to be eliminated, the result would only be a decrease in the overall volume of published science and in the total number of citers and citations, whereas the distribution of citedness would be just as skew as before, and the fraction of uncited papers would remain the same. We could imagine this sanitation process to be repeated until there were only one top scientist left, but with due correction for self-citations even he would remain uncited. Therefore, one should be very wary about using citedness, in absolute terms, as a guide for scientopolitical action.

The fact that a large fraction of the scientific literature is uncited does not mean that it is not being read, or that it does not contribute to scientific progress, it means that the total number of citations given (determined by the average number of references per article) is simply too small to give room for all published articles. As part of a continuous probability distribution, each uncited paper has a definite probability of being influential. If nothing else, it is safe to assume that all publications exert some influence upon their authors, an influence which may become manifest in other, more heavily cited articles.

The Ortega Hypothesis

Although neither productivity nor citedness suffice as general indicators of scientific quality, they suggest a generality of skewed distributions which can probably be extended to other quality parameters. Therefore, it would seem reasonable to assume that scientific capability is distributed in a topnormal manner, and that scientists contribute to scientific progress according to a contribution function like the one shown in Figure 6; i.e., about 15% of the scientists accounting for 50% of the scientific development.

Previous estimates have probably somewhat exaggerated the contributive heterogeneity by assuming an inverse power law (double-logarithmic) distribution rather than an extreme-property (semilogarithmic) distribution of scientific influence (Cole, 1983; Price, 1976), but there is still enough hetergeneity left to justify the continuing debate about how the distribution should be interpreted (several authors, 1987). The "Ortega hypothesis" asserts that eminent scientists depend on the work of the more mediocre ones (Ortega y Gasset, 1932), whereas, particularly, Cole (1983) suggested, on the basis of citation patterns, that the work of top scientists is built primarily on the work of other top scientists. However, relative to the skewed distribution of author citedness in general, the tendency of highly cited and less cited scientists to overemphasize their own kind is not very striking (Cole, 1983; Snizek, 1986), particularly when considering that field effects (see below), in-house citations, and shared journal preference would be expected to impose some clustering and stratification. Besides, if all segments of the distribution have a similar self-bias, the biases will cancel out, and the primary distribution function will determine the contributions to scientific progress fairly accurately.

Given a skewed, but continuous quality distribution of scientists it is obvious that any line between eminence and mediocrity will have to be drawn rather arbitrarily. Mediocrity may be hard to identify in practice, the quality output of individual authors being so heterogeneous and overlapping. With emphasis on continuity it may be appropriate to consider a somewhat diluted (and hardly controversial) formulation of the Ortega hypothesis: mediocre as well as eminent scientists contribute to scientific progress, but both categories rely more heavily on the work of the eminent, who make the greatest contributions. Since we are dealing with a probability distribution that is not altered significantly by truncation, the same statistical reasons that make it impossible to eliminate uncited articles make it inevitable that, at any level of quality, there will be a majority fraction of scientists who contribute less to scientific progress than the small, dominant elite. Rather than engaging in futile attempts to change the statistical realities, science policymakers should seek to optimize the scientific enterprise through a distribution of funds which is in accordance with the distribution of scientific ability. The rate of scientific progress can furthermore be stimulated by an increase in the overall volume of science, but not by sanitation.

Is Citedness a Useful Parameter for Evaluation of Science?

Citations represent a measure of *utility* rather than of quality—and a limited kind of utility at that. The citation rates that publications achieve are determined by their usefulness to other scientists rather than to society in general. The basic purpose of science, on the other hand, is *the production of new knowledge* in general, the eventual utility of this knowledge being as often unanticipated as it is intended. The progress of science would probably be better served by evaluation parameters more directly related to the basic qualities of science, emphasizing the *novelty, solidity,* and *magnitude* of the knowledge produced.

With sufficient awareness of its limitations, citedness might still seem to hold some potential as at least a partial indicator of scientific quality. However, there are some fundamental problems which seriously restrict the applicability of citedness for evaluation purposes. These problems can be broadly classified into four major categories:

(1) *Citation bias.* Citations are not issued as an entirely fair and objective record of influence, but reflect both the needs and the idiosyncrasies of the citer, including such factors as utility, quality, availability, advertising (self-citation), collaboration or comraderie (in-house citations), chauvinism, mentoring, personal sympathies and antipathies, competition, neglect, obliteration by incorporation, argumentation, flattery, convention, reference copying, reviewing, and secondary referencing (MacRoberts & MacRoberts, 1989; Seglen, 1989a). While the sheer number of factors may help to achieve some statistical balance, we all know of scientists who are cited either much less (ourselves) or much more than they deserve on the basis of their scientific achievements.

(2) *Incomplete journal coverage and other registration problems.* SCI and other citation databases sample only a fraction of the total scientific literature, and although the majority of citations are included (Price, 1976), the coverage may vary between fields and countries (Sivertsen, 1991) as well as between individuals. The citation window (registration period) may furthermore be important: while some authors publish mostly ephemeral papers which are optimally covered by a short-term window, others may be characterized by a slowly developing, but more sustained impact better served by a wide window (Moed et al., 1987). The fact that citations are registered only to the first author of a paper is unproblematic in large-scale studies (Roy, Roy, & Johnson, 1983), but necessitates the use of complete reference lists when individual authors are examined. Misspellings, omissions, and other database errors may likewise be unimportant in a wide perspective, but can have considerable consequence for the evaluated individual (Moed et al., 1987).

(3) *Author variability.* The skewed distribution of citations to individual authors' articles may make it difficult to fulfill the purpose of a citation-based evaluation, i.e., to obtain a representative measure of author citedness. The need for a relatively large material as well as for a time lapse before citations can be reliably recorded makes citedness more suitable for lifetime evaluations than for evaluation of current scientific activity. A considerable degree of overlap between different authors is inevitable, making it necessary to use nonparametric statistical methods to establish whether there are significant differences between the evaluated authors. Furthermore, the evaluator has to face the problem of how to accord credit for multiauthored papers as well as how to weight many infrequently cited papers against a few more heavily cited ones, taking into account that citedness may be a function of, e.g., article length (Hayes, 1983; Seglen, 1991). If citations are simply added up, one obtains a

combined citedness/productivity measure rather than a measure of citedness alone.

(4) *Field effects.* By far, the greatest impediment to the use of citedness as an evaluation parameter is the difference in citation properties between different scientific fields (Moravcski, 1973). If we consider the field as a closed system (all citations being given by papers within the same field as the cited article), each field will have a characteristic mean value of citedness which is determined (primarily) by the average number of references per article and by article age (Moed et al., 1985; Pravdic & Pekorari, 1985). These effects alone have been shown to account for a four-fold difference in citedness between biochemistry and mathematics (Moed et al., 1985). Another factor is the developmental dynamics of the field (expansion, constant size, or contraction) (Hargens & Felmlee, 1984), which—due to the time lag between publication and citation—may alter the ratio between the number of citers and the number of citees. Since this ratio is independent of field size, so is the mean citedness of the field (Cole & Cole, 1974), but it should be noted that a large field provides a wider range, and hence, a better opportunity for the individual article to attain a high citation rate.

However, scientific fields are not closed systems, and the most important source of citational variability is probably the extent to which an article is cited outside its core field. Once the field boundaries are broken there is virtually no limit to the number of citations an article may accrue. While "fieldbusting" is typically a property of the singular article, there are also consistent interrelations between fields; e.g., compare the nonreciprocal relationship between basal and applied sciences. Clinical medicine draws heavily on basal biomedicine but not vice versa (Narin, Pinski, & Gee, 1976), which may account for the three- to five-fold differences in citedness that have been demonstrated between these fields (Folly et al., 1981; Seglen, 1989c). Large differences between disciplines within major fields have also been recorded (Sivertsen, 1991), and there is evidence that systematic field differences may extend even to the subspeciality level (Seglen, 1991). Therefore, it must be considered that each individual scientist may constitute a "microfield" with a characteristic citation probability determined by that individual's research profile. If the choice of research subject is a major determinant of future citedness, it should be obvious that citedness is no basis for a fair evaluation of individual scientists.

At higher aggregate levels, like in evaluations of national science activity, field differences between individuals become relatively unimportant. Nevertheless, it is necessary to consider field effects at the discipline level, and to make the appropriate corrections. For example, in a study of biomedicine in different OECD countries Sivertsen (1991) observed that the basal discipline microbiology received twice as many citations as, e.g., the more applied discipline, biotechnology. This would tend to favor a country like Norway, in which the fraction of microbiology within biomedicine was two to three times higher than in Sweden. To obtain valid comparisons between nations, corrections for field-specific citedness have to be made at the discipline level before a mean citedness value for the whole mainfield is constructed (Sivertsen, 1991).

In conclusion, field effects, as well as several other considerations, would seem to preclude the use of citations in evaluation of individual scientists or research groups. However, citedness can be a useful indicator of scientific impact at the national level, provided due corrections are made for field effects.

Acknowledgment

I thank Dr. Ing. Olav Kaalhus for helpful discussions.

References

Braun, T., Glänzel, W., & Schubert, A. (1990). Publication productivity: from frequency distributions to scientometric indicators. *Journal of Information Science, 16,* 1–8.

Coile, R. C. (1977). Lotka's frequency distribution of scientific productivity. *Journal of the American Society for Information Science, 28,* 366–370.

Cole, J. R., & Cole, S. (1973). *Social stratification in science.* Chicago: The University of Chicago Press.

Cole, J. R., & Cole, S. (1974). Citation analysis. *Science, 183,* 32–33.

Cole, S. (1983). The hierarchy of the sciences. *American Journal of Sociology, 89,* 111–139.

Dewitt, T.W., Nicholson, R. S., & Wilson, M. K. (1980). Science Citation Index and chemistry. *Scientometrics, 2,* 265–275.

Dieks, D., & Chang, H. (1976). Differences in impact of scientific publications: Some indices derived from a citation analysis. *Social Studies of Science, 6,* 247–267.

Folly, G., Hajtman, B., Nagy, J. I., & Ruff, I. (1981). Some methodological problems in ranking scientists by citation analysis. *Scientometrics, 3,* 135–147.

Garfield, E. (1987). *SCI journal citation reports.* Philadelphia: Institute for Scientific Information.

Glänzel, W., & Schubert, A. (1985). Price distribution. An exact formulation of Price's "square root law." *Scientometrics, 7,* 211–219.

Hamilton, D.P. (1990). Publishing by—and for?—the numbers. *Science, 250,* 1331–1332.

Hamilton, D.P. (1991). Research papers: Who's uncited now. *Science, 251,* 25.

Hargens, L. L., & Felmlee, D. H. (1984). Structural determinants of stratification in science. *American Sociological Review, 49,* 685–697.

Hayes, S.C. (1983). When more is less: Quantity versus quality of publications in the evaluation of academic vitae. *American Psychologist, 38,* 1398–1400.

Irwin, J.O. (1975). The generalized Waring distribution. Part I. *Journal of the Royal Statistical Society, A138,* 18–31.

Kyvik, S. (1989). Productivity differences, fields of learning, and Lotka's law. *Scientometrics, 15,* 205–214.

Kyvik, S. (1991). *Productivity in academia: Scientific publishing at Norwegian universities.* Oslo: Norwegian University Press.

Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Science, 16,* 317–323.

MacRoberts, M. H., & MacRoberts, B. R. (1989). Problems of citation analysis: A critical review. *Journal of the American Society for Information Science, 40,* 342–349.

Magyar, G. (1973). Bibliometric analysis of a new research subfield. *Journal of Documentation, 30,* 32–40.

Matthews, P. (Ed.) (1990). *Athletics 1990: The international track and field annual.* London: Sports World Publications, Ltd.

Matthews, P. (Ed.) (1991). *Athletics 1991: The international track and field annual.* Windsor, UK: Burlington Publications, Ltd.

Merton, R. K. (1968). The Matthew effect in science. *Science, 159,* 56–63.

Moed, H. F., Burger, W. J. M., Frankfort, J. G., & Van Raan, A. F. J. (1985). The application of bibliometric indicators: Important field- and time-dependent factors to be considered. *Scientometrics, 8,* 177–203.

Moed, H. F., Burger, W. J. M., Frankfort, J. G., & Van Raan, A. F. J. (1987). *On the measurement of research performance: The use of bibliometric indicators.* Leiden, The Netherlands: Science Studies Unit, LISBON-Institute, University of Leiden.

Moravcsik, M. J. (1973). Measures of scientific growth. *Research Policy, 2,* 266–275.

Naranan, S. (1971). Power law relations in science bibliography—a self-consistent interpretation. *Journal of Documentation, 27,* 83–97.

Narin, F., Pinski, G., & Gee, H. H. (1976). Structure of the biomedical literature. *Journal of the American Society for Information Science, 27,* 25–45.

Ortega y Gasset, J. (1932). *The revolt of the masses.* New York: W.W. Norton.

Pao, M. L. (1986). An empirical examination of Lotka's law. *Journal of the American Society for Information Science, 37,* 26–33.

Pravdic, N., & Pekorari, R. (1985). The citing practices of the authors to the national journals in mathematics, physics, and chemistry. *Scientometrics, 8,* 233–246.

Price, D. J. de S. (1963). *Little science, big science.* New York: Columbia University Press.

Price, D. J. de S. (1965). Networks of scientific papers. *Science, 149,* 510–515.

Price, D. J. de S. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science, 27,* 292–306.

Roy, R., Roy, N. R., & Johnson, G. G. (1983). Approximating total citation counts from first author counts and from total papers. *Scientometrics, 5,* 117–124.

Seglen, P. O. (1989a). Evaluering av forskningskvalitet ved hjelp av siteringsanalyse og andre bibliometriske metoder. *Nordisk Medicin, 104,* 331–335.

Seglen, P. O. (1989b). From bad to worse: evaluation by journal impact. *Trends in Biochemical Science, 14,* 326–327.

Seglen, P. O. (1989c). Bruk av siteringsanalyse og andre bibliometriske metoder i evaluering av forskningsaktivitet. *Tidsskrift for den Norske Lægeforening, 31,* 3229–3234.

Seglen, P. O. (1989d). Kan siteringsanalyse og andre bibliometriske metoder brukes til evaluering av forskningskvalitet? *NOP-Nytt (Helsingfors), 15,* 2–20.

Seglen, P. O. (1991). Die Evaluierung von Wissenschaftlern anhand des 'journal impact.' In P. Weingart, R. Sehringer, & M. Winterhager (Eds.), *Indikatoren der wissenschaft und technik. Theorie, methoden, anwendungen* (pp. 72–90). Frankfurt: Campus Verlag.

Several authors (1987). *Scientometrics, 12,* 293–353.

Shockley, W. (1957). On the statistics of individual variations of productivity in research laboratories. *Proceedings of the Institute of Radio Engineers, 45,* 279–290.

Sichel, H. S. (1985). A bibliometric distribution which really works. *Journal of the American Society for Information Science, 36,* 314–321.

Sivertsen, G. (1991). *Norsk forskning på den internasjonale arena. En sammenlikning av 18 OECD-lands artikler og siteringer i Science Citation Index 1973–86.* Oslo: Institute for Studies in Research and Higher Education.

Snizek, W. E. (1986). A re-examination of the Ortega hypothesis: the Dutch case. *Scientometrics, 9,* 3–11.