GARFIELD AND THE IMPACT FACTOR: THE CREATION, UTILIZATION, AND VALIDATION OF A CITATION MEASURE

Part 2, The Probabilistic, Statistical, and Sociological Bases of the Measure

By

Stephen J. Bensman
LSU Libraries
Louisiana State University
Baton Rouge, LA 70803-3300
USA
E-mail: notsjb@lsu.edu
Telephone: (225) 578-6932

INTRODUCTION

Eugene Garfield laid one of the key empirical foundations for modern information science through the innovation of the citation indexing of science. He created the impact factor as part of the process of developing the *Science Citation Index* (*SCI*) produced by his company, the Institute for Scientific Information (ISI). In determining the coverage of this index, Garfield utilized two citation measures to analyze the structure of the scientific journal system: total citations and the impact factor. Both these measures of journal importance were subsequently incorporated into the *Science Citation Index* (*SC*I) and *Social Sciences Citation Index* (*SSCI*) *Journal Citation Reports* (*JCR*) published annually by ISI. Briefly defined, the total citations measure is the total number of references in the source journals indexed by ISI during the year covered by the *JCR* to all issues of a given journal dating from its origin. In contrast, the impact factor is a ratio calculated by dividing the number of references in ISI source journals to the issues of a given journal published in the two years preceding the *JCR* year divided by the number of citable source items in these issues. The difference in time framework makes total citations a measure of the historical significance of journals and the impact factor a measure of their current significance. Of the two measures, the impact factor came to be the one most widely applied and influential.

The first part of this article (Bensman 2007) comprised an intellectual biography of Garfield, which had the purpose of tracing the evolution of his ideas, mapping out the structure of his thought, and determining the place of the impact factor within this structure. This second part has a complementary purpose. It is first to translate Garfield's bibliometric concepts into those of probability and statistics. Then, using the latter concepts, the paper statistically explains and tests Garfield's findings in respect to total citations and the impact factor as measures of journal importance. The statistical tests are conducted with a sample of 120 journals considered by chemists as relevant to their discipline. Chemistry is the field, in which Garfield began his career. Four measures are used to gauge the importance of these journals: a subjective scoring constructed from the data of a 1993 survey of the Louisiana State University (LSU) chemistry faculty on their journal needs; their 1993 usage at the University of Illinois (UI) at Urbana-Champaign Chemistry Library; and both their total citations and impact factors in the 1993 *SCI JCR*. Since Garfield came to regard scientific journals as sociological entities, a special focus of the statistical analyses is elucidating the social significance of these journals. To accomplish this objective, there is employed data from the key assessments of the scholarly quality of US research-doctorate programs that have been conducted over the years. Here a critical role is played by the data collected for the 1993 National Research Council (NRC) evaluation of such programs (Goldberger, Maher, Flattau, 1995), which utilized for the first time ISI citations directly to the works of the program faculty to make the assessments.

This part's structure is dictated by its purpose. It begins with a recapitulation and, in certain respects, a refinement of the first part, summarizing Garfield's premises and research findings in respect to total citations and the impact factor as measures of journal importance. The key points of this recapitulation are the following. First, Garfield derived the concept of the citation indexing of science off the premise of the importance of the review article, and scientific review literature always had a special significance for him. Second, Garfield came to his most important contribution to information science theory, his Law of Concentration, by analyzing the patterns of total citations to scientific journals. According to this law, due to the interdisciplinary nature of science, whereby the major journals of one field comprise the minor journals of another field, the entire scientific journal system is dominated by a small, interdisciplinary core of large research journals. Garfield utilized total citations as a measure of journals as sociological entities to identify the journals forming this interdisciplinary core. Third, Garfield created the impact factor to counteract the size advantage of the large research journals comprising the core. By means of the impact factor, he made two important discoveries: 1) review articles have a higher citation rate than research articles, thereby validating his premise; and 2) the vast bulk of scientific articles have an extremely low citation rate. He formalized the latter discovery in what he termed "Garfield's constant" or the ratio of citations per cited item per year, which held fairly steady over time, rising from merely 1.33 in 1945 to 2.44 in 1997. Garfield utilized the impact factor to determine which journals outside the dominant core should be covered by the *Science Citation Index* (*SCI*).

The second section describes the construction of the database, which is utilized in this part to explain and test Garfield's findings in respect to total citations and the impact factor as well as his utilization of these measures. Two elements of this description are of primary theoretical importance. The first is the difficulty encountered in defining precisely what constitutes a "chemistry" journal due Garfield's Law of Concentration. This law ensures compound fuzzy journal sets, which are comprised of differing subject subsets and contain materials unrelated to the interests of the discipline under analysis. Due to Garfield's law, exogenous citations are an inevitable byproduct of any attempt to construct a set of journals representative of a given scientific discipline. The second element of theoretical importance is the complexities involved in defining a journal bibliographically due to title changes, divisions into parts, combinations of parts, mergers and absorptions of titles, etc. etc. Great care was taken in the construction of this database to obtain complete backfiles of journals in order to capture the full historical significance of the journals.

To translate Garfield's bibliometric concepts into those of probability and statistics, the third section does a distributional analysis of the four measures of journal importance under consideration—LSU faculty ratings, UI library use, *SCI* total citations, and *SCI* impact factors—testing the frequency distributions of the journals across these measures against the three following theoretical probability distributions: the normal, the binomial, and the Poisson. The purpose of these tests is to determine the stochastic model most appropriate for such data, and the proper model is ascertained to be the compound Poisson distribution, whose hallmark is an excess of variance over that expected as a result of random error. This excess variance is due to two stochastic processes: 1) the probabilistic heterogeneity of the elements and subsets of the set under analysis; and 2) contagion, by which the happening of event affects the probability of its subsequent occurrence. Special attention is paid here to the probabilistic heterogeneity and exogenous citations caused by the anomalous relationship of chemistry to biochemistry resulting from Garfield's Law of Concentration. The results of these tests are used to explain Garfield's findings, and it is demonstrated that the impact factor frequency distributions have a much higher proportion of their variance due to random error than the other measures.

Because of the significance of excess variance, the fourth section is dedicated to analyzing its sources in the frequency distributions of the journals across the four measures of their importance under analysis. Following the logic of the compound Poisson distribution, the journal sample is broken down into subsets defined by the following categorical variables: biochemistry vs. non-biochemistry titles; US vs. foreign titles; association vs. non-association titles; as well as US association, foreign association, US non-association, and foreign non-association titles. The differing probabilities governing these subsets are estimated. There is also explored the role of size as a source of variance in the distributions of the journals across the four measures of their importance. To provide a deeper understanding of the sources of variance, it is the fourth section that investigates the social significance of the journals, finding that US association journals have a higher social status than the other titles in the sample. This is proven by supplementing the analysis of the journal data with an analysis of peer rating, publication, citation, and other data from evaluations of the quality of US research-doctorate programs, particularly—as stated above—the one conducted under the auspices of the National Research Council (NRC) in 1993. The statistical analyses also indicate that the foreign association titles have a similar social function as the US association ones.

The fifth section analyzes the relationships among the four measures of journal importance under analysis by four different methods. First, there is employed the graphic method of plotting the measures against each other by means of scatter diagrams. To aid in the interpretation of these diagrams, hypothetical regression lines are drawn from the origin to the point representing the *Journal of the American Chemical Society*, the putative most important chemistry journal, so that positions of the points representing the other journals can be considered from the perspective of this line. This graphic method is preparatory to testing the relationships among the variables through means of the Pearson product-moment correlation coefficient or Pearson *r*, which gauges the strength of the relationship between two variables by how closely the points representing the observations fit a regression line. Before calculating the Pearson coefficients, the scatter diagrams are analyzed to identify outliers—or journals judged to be too far from the regression line—as a way of discovering factors interfering with the relationships. Then the coefficients are calculated both with and without the outliers to judge the effect of the interfering variables. Following the Pearson r analysis, the relationships among the four measures of journal importance are then explored by two nonparametric ranking techniques: the Spearman rank correlation coefficient or the Spearman rho; and the chi-squared test of independence. The Spearman rho is based upon comparing the ordinal ranks of given observations on two variables, and it is a crucial test, because most journal evaluations are based upon ranked lists of journals. To conduct the Spearman analysis, the ratio measures are first converted into ordinal ranks. The primary focus of this analysis is to explore the hypothesis that the measures of journal importance are too much subject to sampling variance particularly at the lower frequencies for the Spearman correlation to yield reliable results. This is accomplished through the utilization of Poisson confidence intervals and other techniques.   In general, the analysis validates the hypothesis particularly in respect to the impact factor, which is a function of two interacting random Poisson variables—number of citations to a two-year journal backfile divided by number of items judged to be citable in that backfile—instead of one random Poisson variable like the other measures. Due to the large component of random error found by the distributional and Spearman tests in the variance of the impact factors, it was judged necessary to employ the other nonparametric ranking method, the chi-squared test of independence, which is robust against such error. The chi-squared test of independence measures how well the observations fit within common categories. These categories can be defined in ordinal terms, and their size can be determined in accordance with one's tolerance for the amount of acceptable error. All the methods of testing the relationships among the measures of journal importance tend to corroborate each other in that they all find the total citations measure to be a better holistic measure of journal importance than the impact factor, but that it misses facets of journal importance captured by the latter measure. However, it is also shown that the better journals sets are defined in terms of the research vs. review function, the more the impact factor approximates total citations as a measure of journal importance.

As the final step of the analysis, the sixth section investigates the stability of both the total citations and impact factor rankings of the journals over the time period 1993-2003. Garfield had found not only that there is a high degree of stability over time at the higher levels of the total citations rankings, but also that highly cited articles tend to be published in the journals at these upper levels. This results in the historically significant journals being also the currently significant ones, causing total citations and the impact factor to approximate each other on the condition of the journal sets being better defined in terms of the research vs. review function. To investigate this phenomenon, the sixth section constructs sub-samples of journals representative of the journals at all levels of both the total citations and impact factor rankings. It then hypothesizes that a natural result of the interaction of the two stochastic processes underlying compound Poisson distributions, probabilistic heterogeneity and contagion, should be the continued or increasing dominance of the upper level journals as well as a high stability of the rankings over time. Both the total citations and the impact factor sub-samples act in accordance with the hypothesis, thereby not only corroborating Garfield's findings but also demonstrating the continued validity of the results based upon the 1993 data.

Each section has a subsection summarizing its findings and drawing conclusions from them. At the end there are presented general conclusions from the research as well as considerations that should be taken into account in utilizing citation measures of journal importance.

1. GARFIELD'S PREMISES, RESEARCH FINDINGS, AND UTILIZATION OF CITATION MEASURES

Garfield's View on the Importance of Review Articles

Garfield began in his conceptualization of a citation index for science from the premise of the importance of the review article. He was made aware of this importance by the person whose writings most influenced his early intellectual development—the British scientist and Communist, J. D. Bernal. In his monumental work *The Social Function of Science* Bernal (1940, pp. 297-298) stressed the importance of review literature in scientific communication. This importance was corroborated by Bernal (1948) in research he conducted for the 1948 Royal Society Scientific Information Conference. Garfield's belief in the importance of the review article was reinforced by Chauncey D. Leake, who was chairman of the advisory committee to the Johns Hopkins Welch Medical Library Indexing Project, where Garfield began his career in information science in the early 1950s. Leake constantly admonished Garfield (1970, Apr. 22; 1974, Oct. 30; 1978, Feb. 13) to study review articles and try to understand why they were so important in science. Garfield (1987, pp. 13-14) derived the concept of a scientific citation index by combining the structure of the review article with the method of the legal citator, to which he was introduced by William C. Adair (1955), a retired vice president of the company that published Shepard's Citations. As a result, Garfield (1987, May 4) had a heightened sense of the importance of the review article, which he once summed up in a discussion of review literature thus:

> …The "culture" of reviewing the literature is so fundamental to my own professional life that I too may forget that in comparison with research discoveries one reads about in the press, and for which Nobel Prizes are awarded, reviewing may seem to the

> uninitiated to be a relatively humdrum topic.
>
> But it is precisely this mistaken notion that I want to dispel. It is not an accident that so many of our greatest scientists have used, created, and contributed to the review literature. Like an important opinion rendered by the chief justice of the Supreme Court, reviews can have great value and influence…. (p.5)

However, Garfield was also aware that his opinion on the importance of review literature was at variance with that of many scientists. Thus, in the same discussion he (1987, May 4) reported that many authors of ISI Citation Classics believed that review articles should not be automatically granted this award and that some felt that their review articles should not be judged on the same criteria as their articles reporting original research.

Total Citations and Garfield's Law of Concentration

ISI conducted its most important citation analysis of the structure of the scientific journal system in 1971. It was an analysis of all the references published during the last quarter of 1969 in the 2,200 journals then covered by the *SCI*. Garfield (1972) reported on this project in an article published by the journal *Science*, splitting his discussion of it into two basic parts: the ranking of journals by total citations; and the ranking of journals by the impact factor. In respect to the first, the main finding was that the distribution of total citations over journals was highly and positively skewed. Garfield (p. 474) summed up this distribution thus: only 25 journals (little more than 1 percent of *SCI* coverage) were cited in 24 percent of all references; only 152 journals were cited in 50 percent of all references; that only 767 journals were cited in 75 percent of all references; and some 2000 or so journals were cited in 85 percent of all references. Moreover, according to him, only 540 journals were cited 1000 or more times, and only 968 journals were cited even 400 times. In commenting upon the structure of the *SCI* data, Garfield observed, "The predominance of cores of journals is ubiquitous" (p. 475).

The results of the total citations analysis caused Garfield (1971, Aug. 4; 1972, p. 476; 1983, pp. 21-23 and 160) to formulate his Law of Concentration. He derived this law off the Law of Scattering posited by Bradford (1934, Jan. 26; 1948; 1953, pp. 148-159), who had discovered in the early 1930s that articles on a given scientific subject were distributed across journals in the same fashion as Garfield had found total citations to be distributed. Thus, according to Bradford's Law, the articles on a given scientific topic concentrate in a small nucleus or core of journals and then scatter across other journals in zones that must increase exponentially in number of titles to contain the same number of articles on the topic as contained in the journals of the nucleus. For example, in one of Bradford's samples, the articles on applied geophysics were distributed across journals in 1928-1931 in the following manner: a small nucleus of 9 journals (2.8 %) containing 429 articles (32.2 %), a second zone of 59 journals (18.1 %) with 499 articles (37.5 %), and a third zone of 258 journals (79.1 %) publishing 404 articles (30.3 %). Bradford considered his law a function of the principle of the unity of science, according to which every scientific subject is related to every other scientific subject. This is demonstrated by his law mandating that individual journals have varying proportions of articles on different scientific subjects and his inability to determine the number of journals that could but did not have articles on a given subject, indicating that no clear boundaries exist between scientific disciplines. Due to these characteristics, Bensman (2001) described Bradford's Law as a mathematical description of a probabilistic model for the generation of fuzzy sets of the type defined by Zadeh (1965). In contrast to a standard or "crisp" set, whose members are either fully within the set with a membership value of 1 or fully outside the set with a membership value of 0, members of a fuzzy set are only proportionally within the set with membership values running from 0 to 1. From this perspective, Bradford sets can be defined as fuzzy sets, because the membership values of the journals in such a set exponentially decrease as the proportion of articles on a given scientific topic diminishes and scope is opened for articles on other scientific topics. Therefore, Bradford sets are not "crisp" sets with journals having memberships of either 0 or 1 but "fuzzy" sets of journals with memberships ranging from the 0 of an indefinable number of journals outside the set to the 1 of the journals in the Bradford core.

In formulating his Law of Concentration, Garfield transposed Bradford's Law from the level of a single discipline to that of science as a whole and then substituted the distribution of total citations across journals covering various subjects for the distribution of articles on various subjects across journals. Garfield (1971, Aug. 4) first announced his Law of Concentration in one of his weekly "Current Comments" essays with the following declaration:

> …At ISI, we are completing a study which has resulted in a generalization of Bradford's law which, in a sense, "unifies" the demonstration of its validity in studies of individual fields. Allow me the eponymic shorthand of calling this unified theory or generalization "Garfield's law of concentration." The name is intended to suggest that, in opposition to scattering, a basic concentration of journals is the common core or nucleus of all fields. (p. 5)

He described his law as postulating for science as a whole what Bradford's Law had postulated for a single discipline. In his *Science* article Garfield (1972) stated his law thus:

> The data reported here demonstrate the predominance of a small group of journals in the citation network. Indeed, the evidence seems so conclusive that I can with confidence generalize Bradford's bibliographical law concerning the concentration and dispersion of the literature of individual disciplines and specialties. Going beyond Bradford's studies, I can say that a combination of the literature of individual disciplines and specialties produces a multidisciplinary core for all of science comprising no more than 1000 journals. The essential multidisciplinary core could, indeed, be made up of as few as 500 journals…. (p. 476)

In his monograph on citation indexing Garfield (1983, pp. 21-23 and 160) used as a physical analogy for Bradford's law a comet, whose nucleus represents the core journals of a literature and whose tail of debris and gas molecules widening in proportion to the distance from the nucleus depicts the additional journals that sometimes publish material relevant to the subject. According to him, his Law of Concentration postulates that the tail of the literature of one discipline largely consists of the cores of the literatures of other disciplines. Despite the name he gave it, Garfield (1983, Nov. 21) did not consider his law to be a law, once writing, "Garfield's law of concentration is not really a law but a principle" (p. 11). It is obvious that Garfield's Law of Concentration reinforces the fuzzy characteristic of Bradford sets. It, thus, has serious implications for the statistical analysis of sets of citations in any given scientific discipline, because such sets contain citations from other disciplines that act as contaminants, causing outliers and distorting parameter estimates.

Over the years Garfield analyzed the characteristics of the journals comprising the multidisciplinary core identified by the total citations measure as dominating all of science. The following are some of the most important characteristics that he discovered. First, the journals of the dominant multidisciplinary citation core tend to be large in terms of number of articles published annually. Thus, in his *Science* article on the 1971 ISI project Garfield (1972, p. 476) reported that he rarely found among the 1000 journals most frequently cited one that was not also among the 1000 journals most productive in terms of articles published. Much later Garfield (1996, Sept. 2) found that, of the 50 journals highest in total citations in 1994, 28 were also among the 50 highest in number of source items that year. Second, the journals of the citation core maintain their dominance for decades, indicating a high degree of stability in the total citations rankings of journals. This can be seen in two comparisons Garfield made of the journals found highest in total citations in 1969 to those found highest on this measure in later years. In comparing 1969 to 1974 Garfield (1976b, p. 609) found what he considered a remarkable stability, in that of the 206 journals most-cited in the former year, 169 remained among the top most-cited 206 in the latter year. Some fifteen years later Garfield (1991, Sept. 2) confirmed this stability when he published two lists of journals: the 50 journals found most highly cited in his seminal study of 1969 *SCI* citations and the 50 titles most highly cited in the 1989 *SCI JCR*. Of the 50 titles on the 1969 list, 32 could be identified as still being on the 1989 list. And, finally, highly cited papers tend to concentrate in a few prestigious journals like the highly cited *Journal of Biological Chemistry*. When Garfield (1973, Sept. 26) examined a list of the 1000 papers most frequently cited during the previous decade, he found that only about 200 journals had accounted for these 1000 articles, of which half had been published in only 15 journals. Summing up his main point on this matter, Garfield wrote that "it is remarkable that of the 1000 or more most heavily cited articles in the literature, not one appeared in an 'obscure' journal" (p. 6).

The Impact Factor, Garfield's Constant, and Review Journals

Garfield created the impact factor explicitly to counteract the size and age advantages of the journals comprising the interdisciplinary citation core posited by his Law of Concentration to be dominating all of science. In a recent history of this measure, Garfield (2006) stated that the reason why Irving H. Sher and he created the impact factor was to help select additional source journals for the *SCI* and that sorting by the impact factor allows for the inclusion of many small but influential journals, whose importance would be missed by rankings based upon article publication and total citation counts. This purpose was made clear in the article by Garfield and Sher (1963), in which the basic principle of calculating the journal impact factor was established. Here the authors pointed out that in the usual citation count methods the importance of a journal is determined by the absolute number of citations to it. However, according to them, this count is largely a reflection of the number of articles published, and this approach is not much more sophisticated than ranking the importance of journals by the quantity of the articles they publish. Garfield and Sher then declared, "The first step in obtaining a more meaningful measure of importance is to divide the number of times a journal is cited by the number of articles that journal has published" (p.200). The first experiment in constructing the impact factor measure at ISI was made as part of the 1971 project that utilized 1969 citation data to analyze the structure of the scientific journal system. Garfield (1972, Feb. 23) first discussed this experiment in a "Current Comments" essay, in which he once again noted that citation frequency is biased in favor of the large journals. In this essay he again advanced the view that the ratio of citations to sources provides an overall measure of impact, but this time he stipulated that the ratio needs to be limited by chronological criteria, lest it be skewed by a few super-cited classics. Garfield then reported the method chosen during the project to construct a measure of journal importance that would control for both size and age: divide the citations of the source year of the ISI project—1969—to the issues of a given journal published during the two years preceding the source year—1967 and 1968—by the total number of articles published in these issues in 1967 and 1968. This was the basic method that ISI would henceforth use to calculate the impact factor.

Garfield (1972, pp.476-477 and p. 478-479, n. 27 and n. 28) developed more fully the above considerations in the discussion of the impact factor in his *Science* article on the 1971 ISI project. He began this discussion by emphasizing the relationship of citation frequency to journal size, declaring, "In view of the relation between size and citation frequency, it would seem desirable to discount the effect of size when using citation data to assess a journal's importance" (p. 476). Garfield once again presented the method that had been selected to calculate the impact factor but in two footnotes went into certain of the difficulties involved in the calculation of this measure in greater detail. Of the difficulties he discussed, the most important concerned the nature of the divisor of the ratio. In describing this problem Garfield (1972 pp. 478-479, n. 28) pointed out that construction of the divisor is complicated by the variety items published by scientific journals. According to him, whereas many journals publish only full-length reports of original research, many others publish, in addition, editorials, technical communications, letters, notes, general correspondence, scientific news surveys and notes, book reviews, and so on. He then stated that in calculating the impact factor, he had not attempted to limit the divisor to citable items such as lead articles, original communications, etc., in contrast to ones with little likelihood of being cited. Garfield even doubted whether it is possible "to construct an acceptable classification that would accommodate all of the different kinds of published material" (p. 479, n. 28). However, Garfield (1976a) changed this policy in constructing the divisor of the impact factor in compiling the data for the first *JCR* published as part of the 1975 *SCI*. Here the divisor of the impact factor was limited to "citable" items" (p. 6). In an article discussing the first *JCR* Garfield (1976b, p. 613) warned that ISI had revised its definition of source items used to calculate the impact factor in contrast to 1969, when there had been included in the divisor many materials such as editorials and news notes that were not citable by their very nature. This change in policy not only caused major shifts in the impact factor rankings but introduced the potential for a lot of random error in the calculation of the impact factor through a process of classification, whose feasibility Garfield himself doubted.

With the aid of the impact factor Garfield was able to uncover other characteristics of the scientific journal system that both complemented and contrasted with the findings he made with the use of total citations. One of the most important of these was the extremely low citation rate of the typical scientific paper. In assessing the 1971 ISI project Garfield (1973, Sept. 26, p. 5) noted that one of its surprising discoveries was the relatively low impact of articles published in most journals, including journals that seem almost universally accepted as preeminent. In his write-up of the project in *Science* Garfield (1972, pp. 474-475) observed that the highly skewed distribution of citations over journals resulted in the impact of the average paper being relatively slight. His proof of this was that the average paper was found to be cited only 1.7 times per year. He gave a statistical explanation of this low citation rate in a footnote, where he presented statistics showing that from 1964 to 1970 the number of *SCI* citations per cited item per year was consistently around 1.7 (p. 478, n. 19). Later Garfield (1976, Feb. 9) was to call this ratio "Garfield's constant" and puzzle about its significance. Here he gave the following account of the origins of the name:

> … By the time we had completed the 1964 *Science Citation Index* (*SCI*), we were aware that there was a surprising near-constancy in the ratio of 1.7 between references processed each year and the number of different items cited by those references. Very early we began to call the 1.7 ratio *the citation constant*… (p. 5)

However, Garfield was well aware that his so-called "constant" was mutable, and he calculated that doubling the growth rate of scientific literature or doubling the number of references per paper would have the same effect of doubling the supposed "constant." Garfield then came to the following interesting conclusion:

> Obviously a changing number cannot be called a constant. But if the *SCI* were a real random sample of the total literature or achieved 'complete' coverage, we then would observe a constant, I believe, or at least be able to explain why we didn't. (p. 7)

Commenting upon Garfield's Constant in later years, Garfield (1998, p. 72; 1999, Nov. pp. 10-11) emphasized that his "constant" is actually a ratio and observed that it was remarkably stable over time, given the growth of scientific literature, rising from merely 1.33 in 1945 to only 2.44 in 1997. He attributed the increase in this ratio to the inflation of scientific literature over the years. One of the statistical consequences of Garfield's Constant is to constrict the range of impact factors. This can be seen in the data presented in the first *JCR*, which was based upon 1974 *SCI* citations and for the first time presented the results of calculating the impact factor as a ratio of citations to citable items instead of to all items (Garfield,1976a; Garfield,1976b). Here the top journal in total citations is the *Journal of the American Chemical Society* with 98,995, whereas the top journal in the impact factor was *Transplantation Reviews* at 25.579.

Despite the constricted range, Garfield found that the distribution of journals by the impact factor was highly and positively skewed in the same fashion as the distribution of journals by total citations. In his *Nature* article introducing the new *JCR* Garfield (1976b, p. 613) pointed out that of the 2,443 titles covered by the 1974 *SCI* only 150 journals had an impact factor above 3 and that the average impact of all the journals was only 1.015, i.e., below Garfield's Constant. Given the skewed nature of the impact factor distribution, this meant that vast bulk of the titles had to have an impact constricted to the range below Garfield's Constant that rose from merely from merely 1.33 in 1945 to only 2.44 in 1997. Given this constriction of a large number of titles to such a short range, it is not surprising that Garfield (2000; 2005) admitted in later years that the only reason ISI calculated the impact factors reported in the *JCR*s out to three decimal places was to avoid the large number of ties that would have resulted in listing many journals alphabetically in the impact factor rankings. Garfield himself considered the impact factor accurate only up to one decimal place, and this fits in with his knowledge of the random error involved in classifying source items into "citable" and "non-citable" to construct the divisor of the measure. Garfield (2006) wrote:

> …The precision of impact factors is questionable, but reporting to 3 decimal places reduces the number of journals with the identical impact rank. However, it matters very little whether, for example, the impact of *JAMA* [*Journal of the American Medical Association*] is quoted as 24.8 rather than 24.831. (p. 90)

The last statement is inaccurate, and it will be shown below that it has a profound effect particularly at the lower frequencies on ordinal rankings by the impact factor, on which most journal evaluations are based.

However, there is one characteristic of the impact factor that is particularly significant in terms of Garfield's intellectual development. In analyzing the scientific journal system with this measure, Garfield (1972, Feb. 23; 1976b; 1990, pp. 7-8) found that review articles have a higher average citation rate than other types articles and that therefore small review journals comprise a significant proportion of the journals highest on this measure. This feature of impact factor distributions stands in sharp contrast to the total citation distributions, where the dominant titles are usually large research journals. Thus, the impact factor served to validate the premises, off which Garfield developed the *Science Citation Index*.

Total Citations and the Impact Factor as Determinants of ISI Journal Coverage

Garfield (1990, May 28) published his most cogent description of ISI journal selection policies in a "Current Comments" essay that was an adaptation of a talk he gave in Tapei at the Symposium on Science Journal Evaluation sponsored by the National Science Council of the Taiwan Science and Technology Information Center. From this essay it is possible to see that he used total citations and the impact factor to capture different facets of journal importance in making the decision on which titles should be covered by ISI's indexes. At the beginning of his discussion Garfield stated that ISI took three types of information into account when evaluating journals for coverage that ranged from the quantitative to the qualitative: citation data, journals standards, and expert judgment.

The first citation measure presented by Garfield in this essay was total citations. This was the measure he utilized to identify those journals comprising the small multidisciplinary core of journals posited by his Law of Concentration as dominating all science. In terms of this law, these journals form the cores or nuclei of the Bradfordian citation comets of the individual scientific disciplines. To illustrate these journals, Garfield gave a table, which listed the 25 most-cited journals in the 1988 *SCI JCR* in descending rank order. He then ascribed to this group the following characteristics: 1) they tended to be cited year after year, surviving and prospering for decades; 2) they were mostly large journals with 14 of them also ranking among the top 25 in numbers of articles published; and 3) they were dominated by titles in the big life-sciences specialties. Most interestingly, Garfield considered these journals highest in total citations as those journals which scientists intuitively identify as the most important journals of science, writing that this list of 25 journals "…probably agrees closely with most readers' mental list of the most important scientific journals" (p. 5). Here it should be pointed out that total citations should be considered a measure of the "historical importance" of journals, since it counts the number of citations in a given year to the entire backfiles of titles.

Having dealt with total citations, Garfield then took up the impact factor's role in ISI journal selection. One can fully understand his utilization of the impact factor in this process only after comprehending the relationship of this measure to his Law of Concentration. Garfield (1997) outlined the framework of this relationship in the following statement:

> …Once the core journals are selected, the remainder of one's effort is spent selecting from thousands of relatively small and low-impact journals published, both in the advanced as well as in the developing countries. (p. 640)

Garfield (1997) described these journals as comprising "the tail of a long hyperbolic curve" (p. 640). It is thus evident that Garfield used total citations to identify the core journals or the nucleus of the Bradfordian citation comet of a given discipline and the impact factor to select which journals to cover from the journals comprising the long tail of this comet.

In his essay on ISI journal selection policies Garfield (1990, May 28) began his discussion of the impact factor by declaring its purpose was to compensate for the "putative size advantage" (p. 8) of the journals dominant by total citations through a measure designed to estimate the average number citations per article. He implicitly acknowledged the amount of random error in this estimate by calculating the impact factor in the tables ranking journals by this measure only to the first decimal place and not to three decimal places as is done in the *JCR*s. Garfield presented two such tables. The first listed in descending rank order the 25 journals highest on the impact factor in the 1988 *SCI JCR*. He characterized this list by describing it as "obviously dominated by review journals, which tend to publish fewer contributions than original research journals, but these are cited much more frequently" (p. 8). Thus, for Garfield, one key advantage of the impact factor was its capability to identify the all-important review journals among those comprising the long tail of the Bradfordian citation comet of a given discipline. Garfield's second impact factor table was most interesting but not for reasons stated in this essay. This table listed in descending rank order the 25 journals highest on the impact factor after restricting the set to only those journals publishing at least 100 articles to exclude most review journals. Such a restriction in effect reintroduced the element of size. In commenting upon this list he noted only that the impact factor tended to favor research areas like the life sciences that more heavily cite recent research published in the previous two years. However, as noted above, in a previous essay Garfield (1972, Feb. 23) had pointed out that the impact factor ratio could be highly skewed by a few super-cited classics unless limited chronologically, and he justified the two-year limit as resulting in "a current impact factor which discounts the effect of most super-classics" (p. 6). To demonstrate the possible effect of such a citation classic, Garfield (1996, Sept. 2) reported that Oliver Lowry's classic 1951 protein determination paper alone accounted for about 7,000 or 3 % of the 265,000 citations in 1994 to the *Journal of Biological Chemistry*, which was the journal most highly ranked in total citations that year. Thus, in effect, by restricting the set to journals publishing 100 or more articles and then ranking these journals by the impact factor, Garfield was presenting a method of comparing the "current importance" of research journals to their "historical importance," which total citations measure. As a sign that high-impact research articles tend to concentrate in a highly stable core of journals dominating all science, it was found that, of his list of 25 journals highest in total citations, 12 were also on his list of the 25 journals highest in the impact factor but publishing more than 100 articles per year, indicating a considerable overlap between "current importance" and "historical importance." At the end of his discussion of the impact factor in his essay on ISI journal selection policies, Garfield (1990, May 28) declared that impact factors were not "the sole or single most important criterion for coverage" and that journal impact was only one of "several quantitative and qualitative factors" (p. 9) taken into consideration.

The Nature of Journals: Bundles of Articles vs. Sociological Entities

Garfield's views on the nature of scientific journals evolved over time. Initially he was influenced by the ideas of Bernal (1940, pp. 292-308), who considered the article as the chief vehicle of scientific communication and journals as inefficient bundles of articles. Bernal sought to reform the scientific communication system by abolishing journals and replacing with them with a national document delivery system that would deliver to individual scientists packages of articles specific to their needs and interests. One of the chief motives of Garfield (1956) for developing a citation index for science was that he perceived it as a means of defining such packages. This view of the relative importance of the article vs. the journal fits in with the impact factor, which attempts to assess the importance of journals by the importance of the articles they publish.

However, from the very beginning, Garfield was aware of the potential of citations for evaluating scientists and their work. This ultimately led to his adopting a sociological approach toward journals. By the time of the publication of the first *JCR*, Garfield (1976a) had come to view journals as sociological entities. In his preface to this *JCR* he stated that he had begun "to study journals as socio-scientific phenomena as well as communications media" and declared his hope that the *JCR* would prove uniquely useful in exploring the relatively new field of the sociology of science (p. ix). Garfield had been intellectually prepared for such a transition by Bernal, who as a Marxist had pioneered the study of the social aspects of science. Besides Bernal, two other persons were highly influential in Garfield's adoption of the sociological perspective on scientific journals. The first was Robert K. Merton, and the other was Derek J. de Solla Price.

Merton is generally considered the founder of the sociology of science. His most important contribution to information science is the concept of "the Matthew Effect." Merton (1968) first advanced this concept in an article published in *Science*, where he derived it from the Gospel according to Matthew (13:12 and 25:29), which states in the King James translation he preferred: "For unto every one that hath shall be given, and he shall have abundance; but from him that hath not shall be taken away even that which he hath." In this article Merton discussed two aspects of the Matthew Effect. His main focus was on psychosocial processes affecting the allocation of awards to scientists for their contributions. Here Merton used the Matthew Effect as a model for a complex pattern of the misallocation of credit for scientific work, and he summed up this misallocation thus:

> …the Matthew effect consists in the accruing of greater increments of recognition for particular scientific contributions to scientists of considerable repute and the withholding of such recognition from scientists who have not yet made their mark. (p. 58)

In discussing the other aspect, Merton utilized the highly stratified institutional structure of US academic science to point out that the Matthew Effect embodies the principle of cumulative advantage operative in many systems of social stratification and producing the same result: the rich get richer at a rate that makes the poor become relatively poorer. This aspect was more thoroughly studied by Merton's students, Harriet Zuckerman, Jonathan Cole, and Stephen Cole. In her study of US Nobel laureates Zuckerman (1977) found that their careers were advanced by a multiplicative process of cumulative advantage, because the scientific system judged them on functionally relevant criteria. As a result, she wrote, laureates are "advantaged in the sense of being more able to begin with, of getting more of what is needed to perform their roles, and of consequently achieving more" (p. 60). For their part, Cole and Cole (1972) used citation analysis to describe the social stratification system of physics, where most of the work of scientific merit was produced by a small elite.

Price complemented Merton in two basic ways. First, whereas Merton was a sociologist of science, Price was a historian of science. Second, in contrast to Merton, whose approach to science was qualitative and conceptual, Price's approach was largely quantitative and statistical. This enabled the latter to place the former's concepts on firm quantitative and statistical foundations. The most influential work of Price (1963) was his *Little Science, Big Science*, whose approach was described by him in the preface as "to deal statistically, in a not very mathematical fashion, with general problems of the shape and size of science and the ground rules governing growth and behavior of science-in-the-large" (p. viii). In this book Price analyzed the exponential character of scientific growth and frequency distributions, likening them to similar patterns of growth and frequency distributions found operative in nature and human society. He characterized the distributions modeling the productivity of scientists and the library use of scientific journals as "the same Pareto curve as in the distributions of incomes or sizes of cities" (p. 75). In two key articles Price (1976; 1978) created and explained a stochastic model of the cumulative advantage process underlying Merton's Matthew Effect. He called his model the "*Cumulative Advantage Distribution (CAD)*" (1976, p. 292) on the basis of the Coles' work on scientific social stratification. Price considered the beta function as best modeling the stochastic processes underlying bibliometric phenomena and based his CAD on it. Noting that in statistical terms the cumulative advantage process is called "contagion," he compared his CAD to Merton's Mathew's Effect, whose stochastic model, he pointed out, is the negative binomial distribution (NBD). Price noted that in this model contagion is "double-edged" (1976, p. 293) in that success is rewarded by increased chance of further success, but failure is punished by increased chance of further failure. In contrast, Price stated that with his CAD contagion is "single-edged" in that success increases the chance of further success, but failure has no subsequent effect in changing probabilities.

Summary and Conclusions

The purpose of this part of the paper is to explain and test statistically Garfield's premises and research findings in respect to citation measures of journal importance as well as his utilization of them to determine which journals should be covered by ISI's citation indexes. Before this can be done, it is necessary summarize his premises, findings, and utilization of citation measures in a number of clearly defined points.

Two basic premises underlay Garfield's thinking in respect to scientific journals. First, review articles have an extremely important function in scientific literature. It was off this premise that Garfield developed the citation indexing of science. In his opinion, review articles are not only as important as research articles but may even be more important, because they serve as arbiters of the findings presented in the latter type of article. Therefore, Garfield was partial toward any measure that validates the importance of the review article. Second, scientific journals are not only means of communications but also sociological entities.

Garfield used two citation measures to make his major explorations of the structure of the scientific journal system: total citations and the impact factor. Each of these measures revealed different facets of the scientific journal system, and Garfield summarized his finding for each by formulating a bibliometric law. His law summarizing his findings with total citations was Garfield's Law of Concentration, which he derived off Bradford's Law of Scattering. The main points of this law are the following. First, total citations to the journals of a given discipline are highly and positively skewed and concentrate on a relatively few large research journals, which form the citation core of that discipline. Garfield likened this citation core to the nucleus of a comet and the other journals of the discipline to the comet's tail, which expands exponentially in number of titles containing exponentially decreasing numbers of citations relevant to the discipline. Second, due to the interdisciplinary nature of science, the journals forming the citation tail of one discipline are the journals forming the citation cores of other disciplines. The result is a small interdisciplinary core of large research journals dominating the entire scientific journal system. Garfield found these core journals highly stable over time in that they maintained their citation dominance for decades, and he identified these core titles as those intuitively considered by scientists as the most important journals of science. Garfield's Law of Concentration has three important statistical consequences for disciplinary sets of scientific journals: 1) the frequency distribution of titles across total citations is highly and positively skewed; 2) these sets are compound ones consisting of disciplinary subsets; and 3) these sets are fuzzy with many of the journal citations exogenous to the discipline and causing journal outliers.

The impact factor was created by Garfield to counter the size advantage of the large research journals of the citation core. He did this by equalizing the journal backfiles at two years and then calculating the ratio or arithmetic mean of the citations to the citable items in these two years. Garfield made three major discoveries through using the impact factor as his measure. First, the frequency distribution of journals across the impact factor is highly and positively skewed in the same fashion as the frequency distribution of journals across total citations. Second, in line with this, the citation rate or "impact" of most scientific journal articles is extraordinarily low. And, third, review articles have a much higher citation rate than other articles, thereby validating his premise of the importance of review articles. Garfield formalized his findings in respect the low impact of most scientific articles with his Garfield's Constant, which is the ratio or arithmetic mean of the number of citations to the number of cited items in a given year. This constant was low and rather stable across time—rising from 1.33 in 1945 to 2.44 in 1997—and, given the skewed distribution of the impact factors, it set a limit below which most journal impact factors had to fall. Due to the difficulty in defining precisely what is a "citable" item and other such problems, Garfield never considered the impact factor a very accurate measure, and, as a reflection of this, he shortened the measure to one decimal place from the three decimal places published in the *JCR*s when evaluating journals with it.

Garfield did not use the impact factor as a holistic measure of journal importance. Instead he used it in conjunction with total citations to capture facets of journal importance missed by the latter measure. Since total citations measure the full size and temporal range of journals, these may be considered a measure of journals as sociological entities. Garfield used total citations to identify the titles comprising the small, interdisciplinary citation core posited by his Law of Concentration as dominating the entire scientific journal system. These are large research journals for the most part. The impact factor measures journals by their importance of their articles, and Garfield used this measure to identify which journals outside the dominant core should be covered by the ISI citation indexes. Since the impact factor controls for size, its main role here is to identify review journals, which generally publish few articles per year but have an extremely important function in scientific literature according to one of Garfield's main premises. Garfield also used the impact factor to measure the current significance of research journals versus their historical significance captured by total citations. He did this by restricting his set to the larger titles to exclude review journals. This is one way to identify newly established, significant titles.

2. CONSTRUCTION OF THE DATABASE

Louisiana State University (LSU) Chemistry Faculty Ratings

The journal set, which will be used to test Garfield's premises, findings, and utilization of citation measures, originated in a survey of the faculty of the Department of Chemistry of Louisiana State University (LSU) in Baton Rouge on their journal needs. This survey was carried out in April 1993 as part of a pilot study in preparation for a restructuring of the serials holdings of LSU Libraries. Twenty-five persons, or roughly 71% of the department's approximately 35 professors and instructors, responded to the survey. It should be noted here that only the faculty of the Department of Chemistry were surveyed; the Departments of Biochemistry and Chemical Engineering were not included in the pilot study. However, there were organizational links between the Departments of Chemistry and Biochemistry. One person served as distinguished professor in both departments, while an associate professor in the Department of Chemistry was also a member of the adjunct faculty of the Department of Biochemistry. The Department of Chemistry has traditionally been one of LSU's strongest departments, and it scored in the middle of the second highest quartile of the ranking of US chemistry programs by peer ratings of the scholarly quality of program faculty in the 1993 evaluation of US research-doctorate programs by the National Research Council (NRC) (Goldberger, Maher, Flattau, 1995, pp. 316-322).

In the survey the members of the LSU chemistry faculty were asked to identify those journals important to them from the entire serials universe without restricting themselves to titles not on subscription at LSU. Their responses certainly reflected Bradford's and Garfield's laws. The selections of the chemistry faculty classed in a broad spread of the subject categories, into which ISI grouped the journals in the 1993 *SCI JCR*, including some of the following: Engineering, Electrical & Electronic; Environmental Sciences; Geosciences; Materials Science, Ceramic; Nutrition and Dietetics; Physics; as well as Radiology & Nuclear Medicine. In fulfillment of Garfield's Law, one of the 25 respondents picked the prestigious *New England Journal of Medicine*—a result that certainly would have been different had the 25 respondents been medical researchers. Two years later, the LSU chemistry faculty was again surveyed on their journal needs together with the faculty of all the university's other scientific and technical units. For this survey, Bensman and Wilder (1998, pp. 179-183) constructed 33 subject categories called "curriculum cores" by assigning subclass groups from the Library of Congress class groups Q (Science), S (Agriculture), and T (Technology) to LSU scientific and technical academic units on the basis of course descriptions in the university's catalog. During this survey, the LSU chemistry faculty selected 191 journals, which distributed themselves across 11 of these curriculum cores in the following manner: General Science-4 (2.1%), Biology-8 (4.2%), General Technology & Engineering-1 (0.5%), Biochemistry-31(16.2%), Chemistry-99 (51.8%), Computer Science-1 (0.5%), Physics-22 (11.5%), Plant Biology-5 (2.6%), Chemical Engineering-5 (2.6%), Mechanical Engineering-1 (0.5%), and Environmental Studies 1 (0.5%). In addition, the LSU chemistry faculty picked 13 (6.8%) journals that classed in R (Medicine).

Due to this response pattern, it was considered necessary to increase the homogeneity of the sample set by restricting the journals to those that classed in subject categories that could be regarded as subsets of the chemistry set. The main reason for this homogenization was to reduce the number of contaminants from exogenous variables that could distort the intended analysis. The Library of Congress (LC) schedule for class QD, Chemistry, was utilized as a guide to these subject categories. This policy was violated only in the case of Spectroscopy, which is located in LC class QC, Physics, due to the heavy emphasis in the LSU Department of Chemistry on this topic. However, such a method could not eliminate all sources of contaminants due to the fuzziness of the LC class schedules. The main source of contaminants was the anomalous relationship between chemistry and biochemistry already visible in the personnel overlap between the LSU chemistry and biochemistry departments. The LC class schedules treat Biochemistry as a subset of Organic Chemistry within the Chemistry (QD), but they also have class groups for various facets of Biochemistry as subsets of other disciplines. Thus, there is one for Physical Biochemistry as a subset of Physical and Theoretical Chemistry within Chemistry (QD), another for General Biochemistry of Plants and Animals under Biology (General) (QH), another for Plant Biochemistry under Botany (QK), and final class group for Animal Biochemistry under Physiology (QP). In contrast to the LC classification, the Dewey Decimal Classification (DDC) entirely separates Biochemistry from Chemistry by making it a subclass of Life Sciences; Biology (570), and this policy was followed in the 1993 evaluation of US research-doctorate programs by the National Research Council (NRC), which classed Chemistry under the rubric of Physical Sciences and Mathematics but combined Biochemistry with Molecular Biology in the Biological Sciences (Goldberger, Maher, Flattau, 1995). Garfield (1972, Feb. 2) himself explored the relationship of chemistry to biochemistry as part of the 1971 ISI analysis of the scientific journal system, comparing the titles most cited by the *Journal of the American Chemical Society* (*JACS*) in 1969 to the titles most cited by *Biochemistry* in that same year. He noted that *Biochemistry* cited heavily other biochemical journals plus the important biomedical titles, whereas *JACS* cited very little the biochemistry literature.

The journals thus selected for inclusion in the sample set were then subjected to technical and historical analysis, whose primary purpose was to trace the journals through all title changes and divisions into sections back to their initial volume and year of origin to determine their age. The primary criterion for whether a serial remained the same publication through all these vagaries was the consistency and continuity of the volume numbering. During the course of this analysis, it was decided to aggregate all the sections of a given serial into a single unit, because usually the LSU chemistry faculty did not distinguish among different sections of a journal in naming it. Thus, the five sections of the *Journal of the Chemical Society—Chemical Communications*, *Dalton Transactions*, *Faraday Transactions*, *Perkins Transactions 1*, and *Perkins Transactions 2*—were treated as one entity in terms of statistical measures. The journal sample resulting from the above steps contained 154 observations or titles.

A measure of scientific value called faculty score was developed for these journals from the information provided by the respondents to the April 1993 survey of the LSU Department of Chemistry in the following manner. In this survey the chemistry faculty members were asked to prioritize their serials needs by identifying the titles important to them and dividing these titles into the three following groups: 1) those titles used frequently enough for teaching purposes to be needed on campus; 2) those titles used frequently enough for research purposes to be needed on campus; and 3) titles for both teaching and research that could be located off campus and satisfactorily accessed through rapid document delivery. Within each group the faculty were asked to limit themselves to 10 titles, and for the first two groups they were asked to rank the titles in descending order of importance from 1 to 10. The faculty members were also requested to estimate the frequency with which the titles would be used.

Inspection of the responses to the April 1993 survey did not reveal whether the LSU chemistry faculty as a whole regarded teaching or research as the more important in respect to serials. Therefore, it was decided to ignore this distinction, regroup the titles as to whether they were needed on campus or could be located off campus, and eliminate any double counting of titles by individual faculty members. Then each title was assigned 10 points for every faculty member who chose it and another ten points for every faculty member who wanted it on campus. If a title was placed in the off-campus group by a faculty member, it was given no extra points. The titles were also allocated points on how the faculty ranked them, with 10 points given for every rank of 1, down to 1 point for every rank of 10. Finally titles were assigned points on the basis of faculty estimates of the frequency with which they would be used: 10 points for each faculty estimate of monthly or more; 5 points for each estimate of less than monthly up to yearly; and 1 point for each estimate of yearly or less often. Under this system the highest number of points a faculty member could give a title was 40, and the maximum score a title could achieve was 1000. The *Journal of the American Chemical Society* came closest to this maximum with a score of 755.

University of Illinois (UI) Chemistry Library Use

The possibility of testing the validity of LSU faculty score against actual library use presented itself, when Chrzastowski (1993), head of the University of Illinois (UI) Chemistry Library, made available to this author data from a journal use study she had conducted at her library during the period January 4 – March 31, 1993, contemporaneously with the survey of the LSU chemistry faculty. The UI Chemistry Library serves a program that traditionally has been among the top US chemistry programs in terms of peer ratings of the quality of its program faculty. This chemistry program was among the top 9 US chemistry programs in the first such rating conducted by Cattell (1910, p. 685), and it remained consistently among the top 15 US chemistry programs in every subsequent major peer rating from 1924 through 1993. In the 1993 NRC evaluation of US research-doctorate programs the UI chemistry program was eighth highest in the ranking of 168 US chemistry programs by peer rating of the scholarly quality of program faculty (Goldberger, Maher, Flattau, 1995, p. 316). It was also a large program, according to the 1993 NRC data, with 44 faculty members and 277 graduate students. The UI Chemistry Library use data was compiled by having student workers count titles as journals were re-shelved, returned from 2-hour loan, or sent out to other libraries via interlibrary loan. Unbound issues were also counted. The titles on the UI Chemistry Library list were compared to the list of titles in the LSU survey sample of 154 observations and were adjusted to the same technical bases as those in the LSU sample, when a match was found. This process resulted in 120 observations common to both the LSU and UI lists, and these 120 observations comprise the sample analyzed in this paper.

Total Citations and the Impact Factor

The two citation measures of scientific value under analysis, total citations and the impact factor, were derived from the bibliometric data contained in the 1993 *SCI JCR*. Before one can fully understand these measures, it is necessary to know a number of basic facts on how they were constructed. The first is the disciplinary sources of the citations. ISI (1994a, pp.11-12) defined the term "source item" for the *SCI JCR* as an item published in any of the journals covered not only by the *SCI* but also by the company's other two citation indexes, the *Social Sciences Citation Index* (*SSCI*) and the *Arts & Humanities Citation Index* (*A&HCI*). To construct the *JCR*, ISI counted only research articles, review articles, and technical notes. The number of citations a journal is stated as having in the *SCI JCR* is a count of the references to this journal in all the source items processed by ISI for the complete *SCI/SSCI /A&HCI* database. Given Garfield's Law of Concentration, the number of citations to a journal can therefore be regarded as a measure of this journal's importance for all fields of human knowledge. However, this method of compiling the citation data also opens a broader scope for the action of citations exogenous to the subject field under analysis. To give some idea of the scale of the data, the Institute for Scientific Information (1994b, p. 61) stated that the 1993 *SCI* was based upon 3,291 source publications containing 652,532 source items by 819,087 unique source authors.

Second, in compiling the *JCR*, ISI (1994a, p. 7) did not combine journal counts on the basis of "lineage," except where a title change had been so minor that it did not affect the title's position alphabetically. Nor did ISI aggregate the counts for the different sections of a given journal. Therefore, it was necessary to manipulate the *SCI JCR* data to make it conform to the way the journals had been technically defined in constructing the chemistry sample under analysis. To do this, the *SCI* total citations measure was compiled by aggregating into one statistic all the citations from all the different title changes and sectional divisions over time. By this method it was hoped to capture the full size, age, and, thus, "historical significance" of the journals.

This method made the *SCI* total citations measure stand in sharp contrast to the *SCI* impact factor measure. The impact factor represents an attempt by ISI to create a normalized measure of value by controlling the citation frequency of a journal for size and age. This is done by limiting the backfile of a serial to the two years preceding the processing year of the *JCR* and then dividing the references during the processing year to this two-year backfile by the number source items in it to create an average citation rate per source item. The Institute for Scientific Information (1994a) succinctly summarized the purpose and effect of the impact factor thus:

> The impact factor is useful in understanding the significance of absolute citation frequencies. It tends to discount the advantage of large journals over small ones, of frequently issued journals over less frequently issued ones, and of older journals over newer ones. In each such case the first is likely to produce or have produced a larger citable body of literature than the second. All things being equal, the larger that body, the more often a journal will be cited. By providing some qualification of the quantitative data in the *JCR*, the impact factor is an important tool for journal evaluation. (p. 11)

When required by the policy of aggregating journal title changes and sectional divisions into single statistics, the necessary adjustments were made to the corresponding impact factors in the 1993 *SCI JCR*. The restriction of the impact factor to the two years immediately preceding the year under analysis makes it a measure of "current significance" in contrast to the "historical significance" being gauged by total citations. As will be seen, controlling for journal size by estimating the arithmetic mean of citations per citable source item also enables the impact factor to identify a different functionality of scientific journals than that identified by total citations.

Summary and Conclusions

The problems encountered in constructing the database highlight two important theoretical issues in using citation analysis and *JCR* data to evaluate scientific journals. First, there are the difficulties in defining a disciplinary set of journals due to Bradford's Law of Scattering and Garfield's Law of Concentration. The defining of precise subject set is absolutely essential, because conventional statistics is based upon the frequency theory of probability. The most cogent development of the frequency theory was done by Von Mises (1951/1957), who based probability on relative frequencies within what he termed the "collective" but may also be considered a "set." Von Mises defined the collective as "a sequence of uniform events or processes which differ by certain observable attributes, say colours, numbers, or anything else" (p. 12), and he admonished, "It is possible to speak about probabilities only in reference to a properly defined collective" (p. 28). Von Mises (pp. 16-18) used as an example of this requirement the fact that a person's probability of dying at a given age is dependent on whether this person is defined as belonging to a collective containing both men and women or only men. However, due to the interdisciplinary nature of science, it was extremely difficult define a coherent set of journals relevant to the LSU chemists, whose selections classed in a broad array of scientific subjects. The most obvious difficulties involved the anomalous relationship of chemistry to biochemistry and the heavy emphasis of the LSU Department of Chemistry upon spectroscopy, which is generally considered a subfield of physics. From the problems encountered in constructing the set, it becomes clear that sets of scientific journals are both composite ones consisting of different subject subsets and fuzzy ones, whose member journals are recipients of citations exogenous to the subject under consideration.

Second, there are the problems of bibliographically defining a journal due to the policy followed in the *JCR*s of not taking into account journal title changes, continuations, supersessions, etc., and not aggregating the citation accounts of different journal sections. This problem existed from the very beginning. For example, in the article, in which Garfield (1976b) introduced the first *JCR* to the broader scientific community, the total citation rankings had significant deviations from those of the *JCR*, because unlike the *JCR* the citation counts for sections, re-titled continuations, etc., were combined. Tempest (2005) analyzed the effect of journal title changes on impact factors due to the *JCR* policy. He notes that a title change reduces the impact factor of a journal over a three-period—the time necessary for the new title variant to replace fully the old title variant as the basis of the calculation of the impact factor. More surprisingly, Tempest also found that title changes negatively affect impact factors for periods longer than three years, with the impact factors of journals in the physical and chemical sciences taking the longest to recover. The method of constructing the database for this paper attempted to minimize such problems as much as possible by combining the different bibliographic units of titles into single entities and adjusting the citation data accordingly.

Table 1a. Frequency Distributions of Titles by LSU Faculty Score for the Full Set, Biochemistry Subset, and Remaining Subsets

Faculty Score Class	*Full Set*	*Biochemistry Subset*	*Remaining Subsets*
1 to 80	71	16	55
81 to 160	27	5	22
161 to 240	8	3	5
241 to 320	6	1	5
321 to 400	6	0	6
401 to 480	1	0	1
481 to 560	0	0	0
561 to 640	0	0	0
641 to 720	0	0	0
721 to 800	1	0	1
Sum	**120**	**25**	**95**

Table 1b. Normality Tests for Distributions of LSU Faculty Score over Titles Comprising Full Set, Biochemistry Subset, and Remaining Subsets Frequency Distributions

Statistical Measures	*Full Set*	*Biochemistry Subset*	*Remaining Subsets*
Actual Distributions			
Arithmetic Mean (a)	103.4	81.3	109.3
Median (a)	55.5	53	56
Mode (a)	37	10	37
Skewness (b)	2.6	1.4	2.5
Shapiro-Wilk Variance Test p (c)	0.00	0.00	0.00
Distributions Transformed to the Lognormal			
Geometric Mean	66.3	59.3	68.3
Skewness (b)	0.2	-0.3	0.2
Shapiro-Wilk Variance Test p (c)	0.08	0.34	0.15

(a) With the normal distribution the three measures of central tendency—arithmetic mean, median, and mode—equal each other.

(b) With the normal distribution skewness is zero, indicating symmetry.

(c) At the standard 0.05 level of significance the Shapiro-Wilk test rejects the null hypothesis of the normal distribution at a p below this level.

Table 2a. Distribution of Titles by University of Illinois (UI) Chemistry Library Use for the Full Set, Biochemistry Subset, and Remaining Subsets

UI Chemistry Library Use Class	*Full Set*	*Biochemistry Subset*	*Remaining Subsets*
1 to 270	104	23	81
271 to 540	10	2	8
541 to 810	1	0	1
811 to 1080	2	0	2
1081 to 1350	2	0	2
1351 to 1620	0	0	0
1621 to 1890	0	0	0
1891 to 2160	0	0	0
2161 to 2430	0	0	0
2431 to 2700	1	0	1
Sum	**120**	**25**	**95**

Table 2b. Normality Tests for Distributions of University of Illinois (UI) Chemistry Library Use over Titles Comprising Full Set, Biochemistry Subset, and Remaining Subsets Frequency Distributions

Statistical Measures	*Full Set*	*Biochemistry Subset*	*Remaining Subsets*
Actual Distributions			
Arithmetic Mean (a)	148.3	85.7	164.7
Median (a)	53.5	51	56
Mode (a)	40	9	2
Skewness (b)	5.3	2.5	4.9
Shapiro-Wilk Variance Test p (c)	0.00	0.00	0.00
Distributions Transformed to the Lognormal			
Geometric Mean	50.9	39.6	54.4
Skewness (b)	-0.2	-0.6	-0.2
Shapiro-Wilk Variance Test p (c)	0.21	0.46	0.44

(a) With the normal distribution the three measures of central tendency—arithmetic mean, median, and mode—equal each other.

(b) With the normal distribution skewness is zero, indicating symmetry.

(c) At the standard 0.05 level of significance the Shapiro-Wilk test rejects the null hypothesis of the normal distribution at a p below this level.

Table 3a. Distribution of Titles by Science Citation Index (SCI) Total Citations for the Full Set, Biochemistry Subset, and Remaining Subsets

SCI Total Citation Class	*Full Set*	*Biochemistry Subset*	*Remaining Subsets*
1 to 23500	99	17	82
23501 to 47000	11	3	8
47001 to 70500	5	3	2
70501 to 94000	2	1	1
94001 to 117500	1	0	1
117501 to 141000	0	0	0
141001 to 164500	1	0	1
164501 to 188000	0	0	0
188001 to 211500	0	0	0
211501 to 235000	1	1	0
Sum	**120**	**25**	**95**

Table 3b. Normality Tests for Distributions of Science Citation Index (SCI) Total Citations over Titles Comprising Full Set, Biochemistry Subset, and Remaining Subsets Frequency Distributions

Statistical Measures	*Full Set*	*Biochemistry Subset*	*Remaining Subsets*
Actual Distributions			
Arithmetic Mean (a)	14627.6	25827.4	11680.3
Median (a)	5100.5	6436	5062
Mode (a) (b)	11750	11750	11750
Skewness (c)	4.7	3.5	4.2
Shapiro-Wilk Variance Test p (d)	0.00	0.00	0.00
Distributions Transformed to the Lognormal			
Geometric Mean	5139.0	7267.5	4691.0
Skewness (c)	0.3	0.2	0.2
Shapiro-Wilk Variance Test p (d)	0.23	0.37	0.62

(a) With the normal distribution the three measures of central tendency—arithmetic mean, median, and mode—equal each other.

(b) No value was found for mode, so mode was calculated as the center point of the class with the most observations.

(c) With the normal distribution skewness is zero, indicating symmetry.

(d) At the standard 0.05 level of significance the Shapiro-Wilk test rejects the null hypothesis of the normal distribution at a p below this level.

Table 4a. Distribution of Titles by Science Citation Index (SCI) Impact Factor for the Full Set, Biochemistry Subset, and Remaining Subsets

SCI Impact Factor Class	*Full Set*	*Biochemistry Subset*	*Remaining Subsets*
0.000 to 1.600	60	9	51
1.601 to 3.200	37	7	30
3.201 to 4.800	11	5	6
4.801 to 6.400	6	1	5
6.401 to 8.000	1	1	0
8.001 to 9.600	0	0	0
9.601 to 11.200	2	1	1
11.201 to 12.800	1	0	1
12.801 to 14.400	1	1	0
14.401 to 16.000	1	0	1
SUM	**120**	**25**	**95**

Table 4b. Normality Tests for Distributions of Science Citation Index (SCI) Impact Factor over Titles Comprising Full Set, Biochemistry Subset, and Remaining Subsets Frequency Distributions

Statistical Measures	*Full Set*	*Biochemistry Subset*	*Remaining Subsets*
Actual Distributions			
Arithmetic Mean (a)	2.384	3.177	2.175
Median (a)	1.599	2.307	1.519
Mode (a) (b)	0.800	0.800	0.800
Skewness (c)	3.2	2.3	3.6
Shapiro-Wilk Variance Test p (d)	0.00	0.00	0.00
Distributions Transformed to the Lognormal			
Geometric Mean	1.705	2.319	1.572
Skewness (c)	0.4	0.6	0.4
Shapiro-Wilk Variance Test p (d)	0.07	0.33	0.14

(a) With the normal distribution the three measures of central tendency—arithmetic mean, median, and mode—equal each other.

(b) No value was found for mode, so mode was calculated as the center point of the class with the most observations.

(c) With the normal distribution skewness is zero, indicating symmetry.

(d) At the standard 0.05 level of significance the Shapiro-Wilk test rejects the null hypothesis of the normal distribution at a p below this level.

Figure 1. Distribution of Titles by LSU Faculty Score for Full Set, Biochemistry Subset, and Remaining Subsets

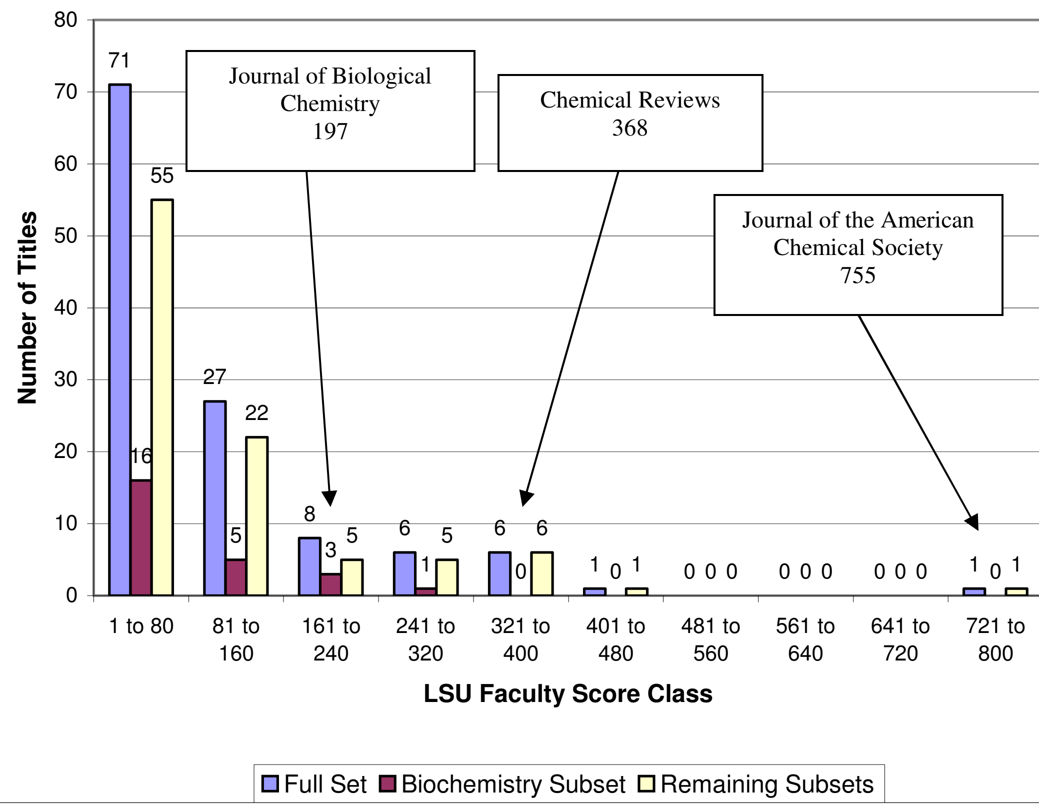


Figure 2. Distribution of Titles by University of Illinois (UI) Chemistry Library Use for Full Set, Biochemistry Subset, and Remaining Subsets

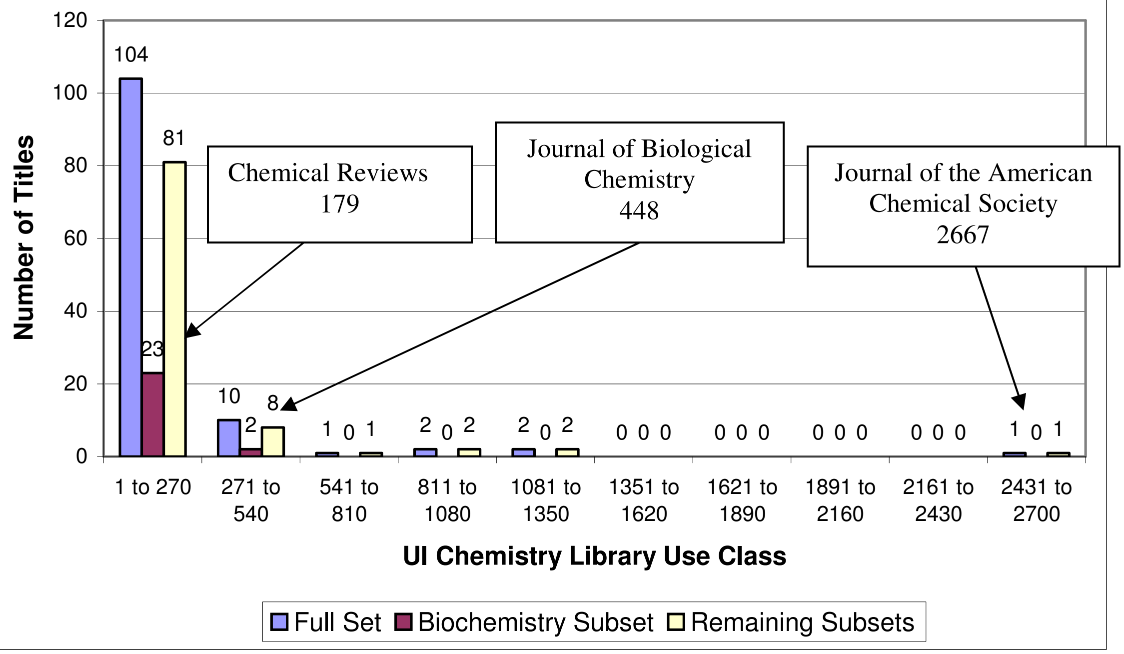


Figure 3. Distribution of Titles by Science Citation Index (SCI) Total Citations for Full Set, Biochemistry Subset, and Remaining Subsets

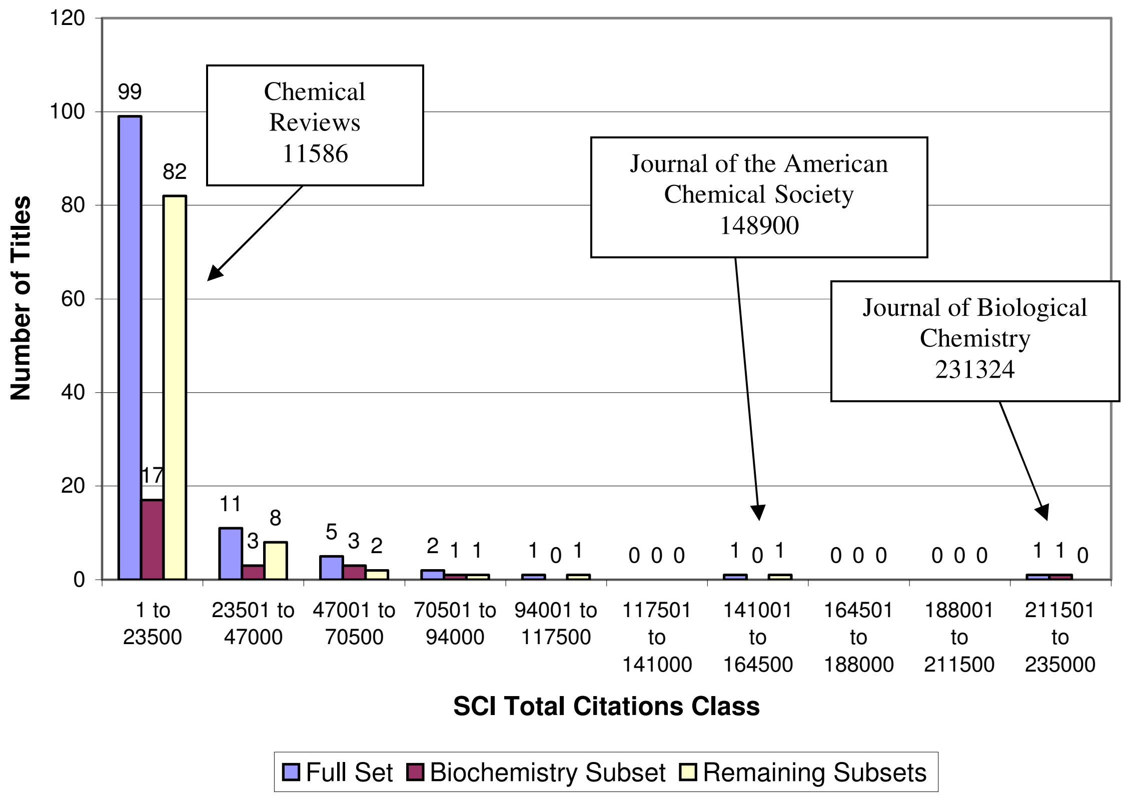
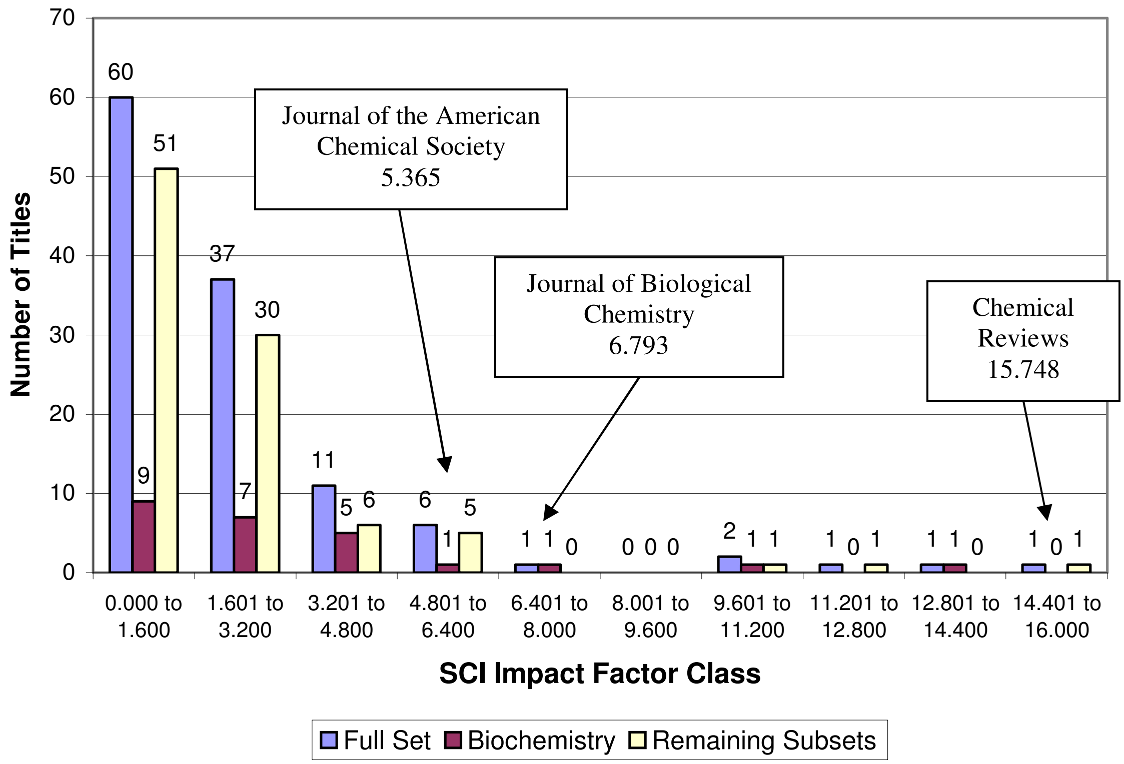


Figure 4. Distribution of Titles by Science Citation Index (SCI) Impact Factor for Full Set, Biochemistry Subset, and Remaining Subsets



Above : TABLES 1-4 and FIGURES 1-4]

3. DISTRIBUTIONAL ANALYSIS OF THE DATA

Frequency Distributions and Tests for Normality

Tables 1-4 above describe the frequency distribution of the 120 journal titles over each of the four measures of scientific value under consideration: LSU faculty score, UI library use, *SCI* total citations, and *SCI* impact factor. To make these tables comparable, the respective ranges of the measures were divided into deciles, and the journals were distributed across these deciles. Each of these four tables is divided into two parts: part a describing the frequency distribution; and part b testing the frequency distribution against the normal distribution. Figures 1-4 above are histograms of these four distributions, and these also show the relative rank positions on the four measures of the three journals which scored highest on at least one of these measures: the *Journal of the American Chemical Society*, which scored highest on LSU faculty score and UI library use; the *Journal of Biological Chemistry*, which scored highest on *SCI* total citations; and *Chemical Reviews*, which scored highest on *SCI* impact factor.

Two interesting characteristics of the data are indicated by the nature of the three journals scoring highest on at least one of the measures. First, the *Journal of Biological Chemistry* is not a chemistry journal but a biochemistry one, and its appearance as the top journal in terms of *SCI* total citations is a sign of the influence of variables exogenous to purposes of the analysis. Therefore, it was decided to break out the biochemistry subset from the full sample and do a comparative analysis of this set in respect to the full set and the remaining subsets left after the removal of the biochemistry subset. This comparative analysis has been structured into the above tables and figures. Of the 120 journals in the sample, 25 or 20.8% classed in biochemistry, showing the close but anomalous relationship of this discipline to chemistry. The second interesting characteristic concerns the different nature of the journal highest in *SCI* impact factor, *Chemical Reviews*, from the two journals highest on the other three measures of scientific value. *Chemical Reviews* was a small review journal, publishing 108 articles in 1993, all of them review articles. Both the *Journal of the American Chemical Society* and the *Journal of Biological Chemistry* were large research journals. The former published 2,276 articles, of which none were review articles, whereas the latter published 3,916 articles, of which 30 (0.8%) were review articles. This is a pattern that was consistently found by Garfield, and it indicates that the impact factor is capturing a different functionality of the scientific journal literature than the other measures.

Inspection of the above tables and figures reveals that the frequency distributions of the titles across all four measures of scientific importance are of the same type, whether for the entire set, the biochemistry subset, or for the remaining subsets excluding the biochemistry subset. They are all highly and positively skewed with the titles heavily concentrated in the lower deciles and then extending exponentially in a long tail rightward. What is particularly noteworthy is the compressed range of the impact factor, whose entire extent is merely 0.000 to 16.000 with 108 of the 120 of the titles located in an extremely short range from 0.000 to 4.800. One way to understand the characteristics of these distributions is to test them against the requirements of the normal distribution.

The normal distribution was developed in the eighteenth century as the law of error in point estimation in astronomical and geodetic observations. Eisenart (1983) states that the purpose of a law of error is "to demonstrate the utility of taking the arithmetic mean of a number of measurements or observed values of the same quantity as a good choice for the value of the magnitude of this quantity on the basis of the measurements or observations in hand" (p. 530). According to K. Pearson (1956, p. 108), three conditions are necessary for the normal distribution to arise: 1) an indefinite number of "contributory" causes; 2) each contributory cause is in itself equally likely to give rise to deviation of the same magnitude in excess and defect; and 3) the contributory causes are independent. The normal distribution is completely defined by its arithmetic mean and standard deviation, and it functions as a law of random measurement error due to its following characteristics; 1) the three measures of central tendency—arithmetic mean, mode, and median—are equal to each other; and 2) the observations are symmetrically distributed about the mean in the form of a bell-shaped curve with a skewness of zero, giving the observations a 50/50 chance of being on either side of the mean. In his treatise on probability, Keynes (1921, pp. 196-209) explored various measures of central tendency as a basis for a law of error, finding that not only the arithmetic mean but also the geometric mean, harmonic mean, and the median could serve as such a basis. In his opinion, the arithmetic mean occupies no unique position in this respect. Keynes further pointed out that statisticians were finding the normal distribution descriptive of neither error nor reality and that the arithmetic mean and the normal law of error can be applied to only certain special classes of phenomena. His opinion in this matter was seconded by Geary (1947), who once recommended that the following warning be printed in bold type in every statistics textbook: ***"Normality is a myth; there never was, and never will be, a normal distribution"*** (p. 241).

Tests for normality, whose results are set forth in parts b of Tables 1-4, reveal how the distributions of the journals across the four measures of scientific value deviate from the assumptions of the normal distribution. These deviations follow the same pattern for the full set, the biochemistry subset, and the remaining subsets after the exclusion of the biochemistry subset. First, the three measures of central tendency—arithmetic mean, median, and mode—are not equal to each other, and the arithmetic mean is much greater than the other two measures of central tendency. It should be pointed out that no value for the mode was found for the two citation measures, and this value was estimated by calculating the center point of the decile containing the most titles—in both cases, the lowest one. The inability to find the modal point was probably due to the vast range of the total citations measure and the calculation of the impact factor to three decimal places. Second, all the distributions are positively skewed instead of symmetrical. And, third, the hypothesis of normality was resoundingly rejected for all distributions by the Shapiro-Wilk test (Shapiro & Wilk, 1965; Shapiro, 1980, pp. 19-24). This test uses regression analysis to compare actual variance to theoretical variance, and all the distributions failed on one of the two key parameters—variance in the form of the standard deviation—that define the normal distribution.

The primary reason for the failure of the normal distribution to describe the frequency distributions of the titles across all four measures of scientific significance can be located in the characteristic of these distributions, upon which Price (1963) focused his analysis in his seminal work, *Little Science, Big Science*—their exponential nature. All the distributions appear to conform to some exponential law, resulting in the appearance of a small number of titles at the high values of the measure. Being so far above the arithmetic mean, these titles caused this mean to be much higher than the other measures of central tendency and increased the variance beyond that expected under the conditions of the normal distribution. This, in turn, invalidates the primary theoretical justification of the arithmetic mean in statistics. According to Moroney (1956), the primary purpose of any average or measure of central tendency is to serve as "as the *representative* of a homogeneous group in which the members are recognizably similar" (pp. 40-41). Moroney describes distributions of the type under discussion here as following "the 'gangster law of growth', i.e. a geometric progression or the exponential law" (p. 38), and, for such distributions, he recommends using as the measure of central tendency the geometric mean, median, or mode. To calculate the geometric mean, the data was logarithmically transformed, which converted the distributions into lognormal ones centered on the geometric mean. According Aitchison and Brown (1957), the lognormal distribution "arises from a theory of elementary errors combined by a multiplicative process, just as the normal distribution arises from a theory of elementary errors combined by addition" (pp. 1-2), and they point out that there are many situations in nature where it is more reasonable to assume that the process underlying change or growth is multiplicative rather than additive. The effect of the logarithmic transformation on the data is shown in part b of the Tables 1-4. In all cases, the geometric mean more closely approximates the median and mode than the arithmetic mean; the skewness is sharply reduced, making the distributions much more symmetrical; and, finally, the Shapiro-Wilk test does not reject the hypothesis that the actual variance matches the theoretical variance expected under the conditions of the normal distribution.

Having established the characteristics of the various measures of central tendency, it is now possible to utilize them to compare the biochemistry subset to both the full journal set and the remaining subsets after the removal of the biochemistry subset. This will be done with the geometric means and arithmetic means set forth in part b of Tables 1-4. Analyzing these means reveals a sharp dichotomy between LSU faculty score and UI library use, on the one hand, and the two citation measures, on the other. For the first two measures, the geometric mean of the biochemistry subset is lower than the geometric means of the full set and remaining subsets, and this effect is the same and exaggerated with the arithmetic means. For the two citation measures, it is precisely the opposite: the geometric mean of the biochemistry subset is higher than the geometric means of the full set and remaining subset, and this effect is the same and exaggerated with the arithmetic means. A number of conclusions can be drawn from this pattern. First, the exaggerated effect with the arithmetic means indicates that much of the causation of the differences stems from the high values of the dominant journals on each measure. Second, this pattern is a reflection of the relationship of biochemistry to chemistry that Garfield (1972, Feb. 2) found during the 1971 ISI citation analysis of the structure of the scientific journal system. As noted above, he found that the journal *Biochemistry* cited heavily other biochemical journals plus the important biomedical titles, whereas the *Journal of the American Chemical Society* cited very little the biochemistry literature. Taken together with this finding, the above shifts in the relative position of the biochemistry subset permits the deduction that many of the journal citations under analysis are exogenous to the interests of the LSU chemistry faculty as well as of the patrons of the UI Chemistry Library and that their appearance in the counts is a function of Bradford's Law of Scattering and Garfield's Law of Concentration.

Binomial Tests

A number of key characteristics of the distributions of the journals across the measures of scientific importance under analysis here can be elucidated by testing these distributions against the assumptions of the binomial distribution. From the perspective of this analysis, the most important feature of the binomial distribution is its close connection of the arithmetic mean with probability. Probability—traditionally designated by p—can be defined as the proportion of successes or occurrences in a sample or set of events. For example, if an urn contains 9 black balls and 1 white ball, the p of a white ball is 0.1. The proportion of failures or non-occurrences is designated by q, and, in this urn, the q of a white ball is 0.9. Together, p+q = 1. The binomial distribution arises, when samples of size s are repeatedly drawn from a universe. In Excel notation, the expected probabilities of the distribution are calculated by the expansion of the binomial (p+q)^s, and its arithmetic mean, variance, and standard deviation are estimated thus:

AVERAGE = s*p
VARbinomial = s*p*q
STDEVbinomial = SQRT(s*p*q) = SQRT(VARbinomial)

Historically the normal distribution was derived off the binomial distribution. According to Snedecor & Cochran (1989, pp. 117-119 and 130), the discrete binomial distribution approximates the continuous normal distribution as sample size s increases, and the size of s required for this approximation is dependent on the value of p, being smallest at p = 0.5.

The binomial distribution is the central distribution of a system of distributions originated by the German economist, Wilhelm Lexis. A. Fisher (1922, pp. 117-126) and Rietz (1924; 1927, pp. 146-155) have written the most cogent summaries of Lexian theories in English. The central measure of the Lexian system of distributions is the Lexis Ratio (L), which compares the actual standard deviation of a distribution (STDEV) to its theoretical binomial standard deviation thus:

L = STDEV/STDEVbinomial.

STDEVbinomial is calculated with the above formula. Rietz (1927, pp.24-25) stipulates two conditions necessary for the binomial: 1) the underlying probability p remains constant from sample to sample; and 2) the drawings are mutually independent in the sense that the results of drawings do not depend in any significant sense on what happened in previous drawings.

Urn models are used by both Rietz (1924) and A. Fisher (1922, pp. 117-126) to demonstrate the Lexian system of distributions, and this method will be implemented here. The urn model for the binomial distribution can be a single urn with black and white balls in constant proportions, where the drawing of a white ball is considered a success. Samples are drawn from this urn and replaced to ensure a constant proportion of white balls and independence of trials. Under these conditions—probabilistic homogeneity and independent trials—the Lexis Ratio should equal or approximate one, indicating that the actual variance equals the theoretical binomial variance. Given the binomial's close relationship to the normal distribution, a Lexis Ratio of one indicates the dispersion around the mean is primarily due to random error.

According to Lexian theory, the variance of the Poisson distribution is less than the corresponding variance of the binomial, so that a Lexis Ratio significantly less than 1 is indicative of the former distribution. The urn model for the Poisson distribution consists of a series of urns with differing proportions or probabilities of white balls. Samples are constructed by taking one ball at a time from each urn and replacing the balls in the urns to maintain the independence of trials. This has the effect of randomizing the heterogeneous probabilities of the urns and reducing the variance below that expected under the conditions of the binomial. However, it is evident from the demonstration by Rietz (1924) that, while randomizing the heterogeneous probabilities in this manner significantly reduces the variance below that expected under the binomial, this procedure does not necessarily result in the Poisson distribution.

A Lexis Ratio greater than one means that the actual variance is greater than the theoretical binomial variance and is the sign of the Lexis distribution. Simply defined, a Lexis distribution is a distribution that results, when the set under analysis is not probabilistically homogeneous but consists of subsets governed by differing probabilities, whose effects are manifest. The urn model for the Lexis distribution is similar to the one for the Poisson distribution in that it, too, consists of a series of urns containing different proportions or probabilities of white balls. However, unlike the Poisson model, the samples are constructed not by taking balls sequentially one from each urn but taking full samples from each urn in rotation. As the balls are drawn, they are replaced to maintain trial independence. This method of sampling emphasizes the probabilistic heterogeneity of the urns or subsets, raising the variance above that expected under the conditions of the binomial. Lexian theory divides the variance into two components: the "ordinary or unessential" binomial component and the "physical" component (Rietz 1924, p. 86). The binomial component may be defined as that variance due to random error, whereas the excess variance is interpreted as resulting from the differing probabilities of the component subsets and is a sign of the Lexis distribution. This is actually a good model for the binomial sampling of citations to journals. Under it journals can be conceptualized as citation samples of size s drawn fully in rotation from urns or subsets with differing proportions of citations and non-citations.

In his classic textbook R.A. Fisher (1970, 68-70) presents a test for the binomial that utilizes the chi-squared distribution as an index of dispersion. An examination of Fisher's equation for chi-squared in his binomial test reveals it to be based upon a comparison of the actual variance of a set of data to its theoretical binomial variance. The relationship to the Lexis Ratio is obvious, and Fisher himself states, "In the many references in English to the method of Lexis, it has not, I believe, been noted that the discovery of the distribution of [chi-squared] in reality completed the method of Lexis" (p. 80). He then presented a method by which a given chi-squared could be transformed into its equivalent Lexis Ratio. Fisher's index of dispersion test was further developed by Cochran (1954), who placed it within the system of hypothesis testing which is the standard method in statistics today. This system involves null and alternative hypotheses. Given Fisher's linking of his binomial index of dispersion test with the Lexis Ratio, one can define the hypotheses for his binomial index of dispersion test in accordance with Lexian theory. The null hypothesis is the binomial distribution. If the actual variance is significantly less than the theoretical binomial variance, the alternative hypothesis is that the distribution has the subnormal dispersion indicative of the Poisson distribution; if the actual variance is significantly greater than the theoretical binomial variance, the alternative hypothesis is that the distribution has the supernormal dispersion characteristic of the Lexis distribution. Thus, Fisher's binomial index of dispersion test can be considered from the Lexian viewpoint a test for whether a set is homogeneous or composed of subsets governed by differing probabilities. This test is discussed in Snedecor & Cochran (1989) under the name "variance test for homogeneity of the binomial distribution" (p. 204).

Table 5 below presents the results of the binomial tests of the distributions of the titles comprising the full set, biochemistry subset, and remaining subsets over the measures of scientific value. The method of testing involved calculating the Lexis Ratio, converting this ratio into its equivalent chi-squared, and using the chi-squared to judge the significance of the ratio. A major problem concerning the binomial with this type of data is that, in order to calculate p, one has to know not only the number of successes or occurrences but also the number of failures or non-occurrences. The latter are invisible by definition and uncountable. One way around this

Table 5. Binomial Tests of the Distributions of the Titles Comprising Full Set, Biochemistry Subset, and Remaining Subsets over the Measures of Scientific Value			
Statistical Measures	***Full Set***	***Biochemistry Subset***	***Remaining Subsets***
LSU Faculty Score			
Estimated Probability	0.103	0.081	0.109
Arithmetic Mean	103.4	81.3	109.3
Standard Deviation	112.4	66.2	121.3
Estimated Binomial Standard Deviation	9.6	8.6	9.9
Lexis Ratio	11.7	7.7	12.3
Chi-Squared p (One-Tailed) (a)	0.00	0.00	0.00
University of Illinois (UI) Chemistry Library Use			
Estimated Probability	0.056	0.032	0.062
Arithmetic Mean	148.3	85.7	164.7
Standard Deviation	318.5	119.2	351.4
Estimated Binomial Standard Deviation	11.8	9.1	12.4
Lexis Ratio	26.9	13.1	28.3
Chi-Squared p (One-Tailed) (a)	0.00	0.00	0.00
Science Citation Index (SCI) Total Citations			
Estimated Probability	0.062	0.110	0.050
Arithmetic Mean	14627.6	25827.4	11680.3
Standard Deviation	29386.3	48248.3	21371.2
Estimated Binomial Standard Deviation	117.1	151.6	105.3
Lexis Ratio	250.9	318.3	202.9
Chi-Squared p (One-Tailed) (a)	0.00	0.00	0.00
Science Citation Index (SCI) Impact Factor			
Estimated Probability	0.036	0.048	0.033
Arithmetic Mean	2.384	3.177	2.175
Standard Deviation	2.559	3.095	2.373
Estimated Binomial Standard Deviation	1.516	1.739	1.450
Lexis Ratio	1.7	1.8	1.6
Chi-Squared p (One-Tailed) (a)	0.00	0.00	0.00
(a) At the standard 0.05 level of significance the chi-squared test rejects the null hypothesis of the binomial distribution in favor of the alternative hypothesis of the Lexis distribution at a p below this level.			

[Above: TABLE 5]

difficulty is to utilize the technique suggested by Grieg-Smith (1983, pp. 57-58) and recommended by Elliott (1977, p. 17). This technique requires that one first define the size of the binomial sample s by either determining or hypothesizing the maximum possible number of occurrences for any given member of the set. The method will now be demonstrated with LSU faculty score for the full set, and the results will be used to calculate the theoretical binomial parameters of the distribution. The largest number of faculty points any journal could have received was 1000, which thus becomes the size of the sample s. Each journal, thus, represents a binomial sample of 1000 points, and, since there were 120 journals in the full set, the total number of points the set could have received was 120000. Since the set actually received only 12412 points, the calculation of p and the binomial parameters is as follows.

p = 12412/120000 = 0.103
q = 1.000-0.103 = 0.897
AVERAGE = s*p = 1000*0.103 = 103.4
VARbinomial = s*p*q = 1000*0.103*0.897 = 92.7
STDEVbinomial = SQRT(1000*0.103*0.897) = SQRT(92.7) = 9.6

For UI library use, s was assumed to be 2667, which was the number of uses of the *Journal of the American Chemical Society*; for *SCI* total citations, s was taken to be 234319, which was the number of citations received by the title highest on this measure in 1993, the *Proceedings of the National Academy of Sciences of the USA*; and, for *SCI* impact factor, s was hypothesized to be 66.273, which was the impact factor the title highest on this measure in 1993, *Clinical Research*.

Due to the assumptions required for the calculation of p, the binomial parameters and Lexis Ratios set forth in Table 5 can be considered to be only crude estimates. Yet, even so, they are highly informative. First, it is quite evident that we are dealing with extremely low probabilities. The range of p is from a minimum of 0.032 for the UI library use of the titles of the biochemistry subset to a maximum of 0.110 for the *SCI* total citations of the titles of this same subset. Second, in all cases, the Lexis Ratio is significantly above one indicating a rejection of the null hypothesis of the binomial and acceptance of the alternative hypothesis of the Lexis distribution. This indicates not the probabilistic homogeneity required by the binomial but probabilistic heterogeneity due to the existence of subsets with differing probabilities and therefore arithmetic means. A conclusion to be drawn from this is that even the low probabilities and arithmetic means in Table 5 are overestimates of the probabilities and consequent arithmetic means of most of the titles in the set and subsets, since these overall probabilities and arithmetic means are being skewed upward by those of the dominant titles and subsets

Special attention must to be drawn to the binomial tests of *SCI* impact factor. What is particularly striking is the relatively low Lexis Ratios of the impact factors in comparison to those of the other measures—particularly, *SCI* total citations—for the full set, biochemistry subset, and remaining subsets. The Lexis Ratios of the impact factors range from 1.6 to 1.8, and, in Lexian terms, being so close to one indicates that the “ordinary or unessential” binomial component or error component of the variance is rather large in respect to the “physical” component. This finding fits in with Garfield’s view noted above (p. 10-11) on the imprecision of the impact factor due to such reasons as the difficulty of classifying sources into “citable” and “non-citable” as well his own consideration of the impact factor as accurate to only one decimal place instead of the three decimal places published by ISI.

A glance at Table 4a and Figure 4 above reveals a distributional characteristic of the impact factor that compounds the error inherent in this measure. In all cases, the bulk of the journals are crowded into the extremely short range of the first two deciles 0.000 to 3.200, making quantitative distinctions at this level rather meaningless. The underlying cause of this situation becomes quite clear once one translates Garfield’s Constant into binomial terms. Garfield (1976, Feb. 9) defined his constant as the ratio of the number of references processed in a given year to the number of different items cited by those references. The 1993 *SCI* annual guide refers to this constant as “Average Number of Citations to Author Cited Items” (ISI, 1994b, p. 61). In 1993 this constant was 2.15, and it was calculated as follows:

Citations to Authored Cited Items = 13,359,470
Unique Authored Items Cited = 6,222,837
Average Number of Citations to Authored Cited Items = 2.15

To convert Garfield’s Constant to the binomial p, we will posit that the highest number of citations an unique authored item could have received in 1993 was 7,000, which Garfield (1996, Sept. 2, p. 13) stated was the number of citations which Oliver Lowry’s classic 1951 protein determination paper received in 1994. Lowry’s paper was the most highly cited paper up until that time. This makes each unique authored item a sample of 7,000 citations and the number of samples 6,222,837. Calculation of the binomial p and the resulting arithmetic mean is then as follows.

Size of Sample (s) = 7,000 citations
Number of Samples (n) = 6,222,837 unique authored items cited
Total Number of Possible Citations (tp) = s*n = 7,000*6,222,837 = 43,559,859,000
Total Number of Observed Citations (to) = 13,359,470
Binomial p = to/tp = 13,359,470/43,559,859,000 = 0.0003
Average Number of Citations to Authored Cited Items = s*p = 7,000*0.0003 = 2.15

As it is possible to see, the binomial arithmetic mean is equal to Garfield’s Constant, and the p of 0.0003 is the probability of each unique author item being cited under the conditions of probabilistic homogeneity required by the binomial. However, as low as the p of 0.0003 is, it is certainly an overestimate of the actual probability, because Garfield’s Constant is based on a distribution that is truncated on the left, i.e.: there is not taken into account an almost certainly huge zero class consisting of items that could have been cited but were not. A method to gauge the possible effect of this zero class upon p and the resulting Garfield’s Constant is presented by Price (1976, pp. 300-303) in his testing of the applicability of his Cumulative Advantage Distribution (CAD) to journal citation data. Interestingly enough, in his test Price also used Lowry’s classic paper as an estimator of the maximum number of citations an article could possibly receive. According Price’s calculations (p. 303), Garfield’s Constant—the number of citations per item actually cited—equals one more than twice the mean citation rate to the total number of items in the citable corpus. Using this equation, it was estimated that adding the zero class reduces Garfield’s Constant from 2.15 to 0.58. Moreover, given the heterogeneity of scientific literature and the skewed distributions resulting from it, most articles’ citations, whose arithmetic mean is being estimated for each journal by the impact factor, have to fall below Garfield's Constant or the overall set mean. Of the 4,541 journals ranked by the 1993 *SCI JCR*, only 553 (12.2%) had an impact factor above Garfield's Constant, and 2.15 demarcates too small of a range into which to squeeze some 4,000 journals and expect meaningful quantitative distinctions among them. Here we find the probabilistic bases for what Garfield (1973, Sept. 26) termed one of the "most surprising" (p. 5) discoveries of the 1971 ISI citation analysis of the scientific journal system: the relatively low impact of articles published in most journals, including journals that seem almost universally accepted as preeminent. Inspection of Tables 4a-4b above reveals that the mean impact factor of the journals comprising the full set, biochemistry subset, and remaining subsets were fairly close to Garfield's Constant for that year and followed the same type of distribution as the overall *SCI* journal set with the bulk of the journals clustered at or below Garfield's Constant. It should be pointed out that in general these are higher quality journals that had to pass three tests for inclusion into the set: 1) selected by the LSU chemistry faculty; 2) held by the UI Chemistry Library; and 3) covered by the *SCI*.

The Lexis distribution is a pioneering version of the compound distribution, which is a probability distribution that results when the parameter has its own distribution sometimes termed the "mixing distribution." Properly conceived, the Lexis distribution is a mixture of binomial distributions, and it is considered the forerunner of the compound binomial distribution. Moran (1968, p. 76) as well as Johnson & Kotz (1969, p. 79) describe the beta distribution as the "natural" mixing distribution for the parameter p in the compound binomial distribution. This form of the compound binomial is sometimes named the beta binomial distribution (BBD). It is interesting to note that the beta distribution is based upon the beta function, which Price (1976) utilized as the mathematical model for his Cumulative Advantage Distribution (CAD). In particular, Price advanced his CAD as "an appropriate underlying probabilistic theory for the Bradford Law, the Lotka Law, the Pareto and Zipf Distributions, and for all the empirical results of citation frequency analysis" (p. 292). Compound distributions are ideal vehicles for conceptualizing the structure of the probabilistically heterogeneous sets inherent in the analysis of scientific literature. This heterogeneity can take a number of different forms. First, there is the probabilistic heterogeneity of the different subject subsets and elements arising from Bradford's and Garfield's laws. Second, even within supposedly homogeneous subject sets, there can be probabilistically heterogeneous elements and subsets arising from such causes as differing size, nationality, language, types of publishers, functionality, etc. And, finally, probabilistic heterogeneity can arise from complex combinations of all of these causes.

Poisson Tests

R.A. Fisher (1970, p. 54) in his textbook noted that, whereas the normal distribution is the most important of the continuous distributions, the Poisson distribution is of the first importance among the discrete distributions. This certainly holds true in respect to information science, for which the Poisson's characteristics are ideally suited. First, the Poisson arises as a limit to the binomial as sample size s tends to infinity and probability p tends to zero in such a way that AVERAGE = s*p is a constant. When p is extremely small, the binomial and the Poisson are equivalent. As we have seen above, we are dealing with extremely low levels of p. The conditions necessary for the Poisson are dictated by its being a special case of the binomial, when p is very small. Elliott (1977, p. 22) has stipulated these conditions as follows: 1) p must be small and constant; 2) the number of occurrences per sample must be well below the maximum number of occurrences possible; 3) the occurrence of an event must not affect the probability of another occurrence of the event; and 4) the sample size s must be small relative to the population.

A second reason for the appropriateness of the Poisson for information science is the process from which it arises. The Poisson distribution is based upon counts of the occurrences of isolated events over some continuum—time, length, area, or volume—and, in contrast to the binomial, it does not require knowledge of the number of non-occurrences. This makes the process, on which the Poisson is based, ideally suited for information science for the same reason it is so suited for the other social sciences. Coleman (1964) summed up the advantages of the Poisson process for the social sciences thus:

> The appropriateness of the Poisson process for social phenomena lies not in its empirical fit to social data. It lies instead in the assumptions on which the distribution is based. In the first place, it deals with *numbers* of elements, or proportions, and with numbers of events. Therefore, continuous-variable measurements, which are extremely rare in social science, are unnecessary. Second, the Poisson process occurs continuously over time, rather than at discrete "trials," like the binomial distribution. Thus, for naturally occurring events, in contrast to controlled experiments, something akin to the Poisson process is often appropriate.
>
> Finally, the Poisson process is appropriate to social phenomena because it constitutes a rational model whose assumptions can mirror our assumptions about actual phenomena. Thus, it need not be simply an empirical frequency distribution like the normal curve, applied because it fits the data. The normal curve provides no such rational model, though it does stand as an approximation to both the binomial and the Poisson…. (p. 291)

The Poisson distribution has one parameter, lambda (λ), which can be defined as the expected number of occurrences over the continuum under analysis. Under the conditions of the Poisson, lambda equals both the arithmetic mean and variance, and the Poisson distribution is, thus, characterized by the following identity:

$$\lambda = \text{AVERAGE} = \text{VAR}$$

Taking all the above together, it is now possible to define Garfield's Constant as an estimate of the Poisson lambda or expected citation rate of those articles actually cited by the source items processed for the *SCI* during a given year on the hypothesis that the conditions of the Poisson are being met.

Violations of the two main conditions of the Poisson distribution—probabilistic homogeneity and independence of occurrences—is generally modeled by the negative binomial distribution (NBD), which takes two forms depending upon the type of violation. The first type of violation is modeled by the form of the NBD called the gamma Poisson, which is a compound Poisson distribution where the lambdas of the component simple Poisson distributions follow the gamma distribution. This distribution was developed by Greenwood and Yule (1920) on the basis of industrial accidents among British female munitions workers during World War I. According to the Greenwood and Yule model, each female worker was considered as having a mean accident rate over a given period of time or her own lambda. Thus, the accident rate of each female worker was represented by a simple Poisson distribution. However, the various female workers had different underlying probabilities of having an accident and therefore different lambdas. Greenwood and Yule posited that these different lambdas were distributed in a skewed fashion described by the gamma distribution. Therefore, certain workers were more accident prone, had much higher accident rates than the others, and accounted for the bulk of the accidents.

The violation of the Poisson assumption of independence of occurrences is modeled by the contagious form of the NBD, which was formulated by Eggenberger and Pólya (1984) in a 1923 paper that analyzed the number of deaths from smallpox in Switzerland in the period 1877-1900. They derived their model off an urn scheme that involved drawing balls of two different colors from an urn and not only replacing a ball that was drawn but also adding to the urn a new ball of the same color. In this way numerous drawings of a given color increased the probability of that color being drawn and decreased the chance of the other color being drawn. In a key paper Feller (1943) proved that the Greenwood-Yule model of probabilistic heterogeneity and the Eggenberger-Pólya model of contagion both result in the NBD. As result of this, Feller pointed out that one does not know which of these stochastic processes is operative—one, the other, or both interactively—when one finds the NBD, and he pointed out that this conundrum also applies to other types of contagious distributions.

It was the Eggenberger-Pólya contagious form of the NBD that Price (1976) posited as the statistical model of the "double-edged" Matthew Effect advanced by Merton and his students as the cumulative advantage process underlying the social stratification of science. Price based his Cumulative Advantage Distribution (CAD) model of the "single-edged" Matthew Effect on the work of Simon (1955), who rejected contagious distributions of the NBD type as not providing a satisfactory fit to a wide range of empirical data—particularly data descriptive of sociological, biological, and economic phenomena. Instead, Simon proposed a model based on the beta function, which he then proved to derive from stochastic processes closely similar to those yielding the NBD but to provide a better fit to the data of interest to him. He called his model the "Yule distribution" after G. Udny Yule, who had pioneered it in an analysis of the distribution of biological species. Price (1976; 1978) argued that the Yule distribution is more appropriate for citations and other bibliometric data than the NBD. To demonstrate the stochastic processes underlying the Yule distribution, he developed an urn model for "lifetime scores of a series of games played tournament style, gradually reducing the large field of players to a small elite of highly successful champions" (1978, p. 204). In Price's model there are no discrete failure events but only intervals separating successes, and each success increases the transition probabilities to further success in a process of cumulating advantage. This paper will not discuss the relative merits of the "double-edged" vs. "single-edged" Matthew Effect. I will simply note that Simon (1955, pp. 432-433; 1960, pp. 85-86) himself derived the Poisson distribution as one limit of the Yule distribution and Fisher's logarithmic series as another limit. In its turn the latter distribution arises as a limiting form of the NBD.

In his textbook R. A. Fisher (1970, pp.57-61) sets forth a chi-squared index of dispersion test for the Poisson distribution that derives from the same principle as his chi-squared index of dispersion test for the binomial. This test is based upon the variance-to-mean ratio, which should

Table 6. Poisson Tests of the Distributions of the Titles Comprising Full Set, Biochemistry Subset, and Remaining Subsets over the Measures of Scientific Value

Statistical Measures	***Full Set***	***Biochemistry Subset***	***Remaining Subsets***
LSU Faculty Score			
Variance	12639.9	4380.8	14718.2
Arithmetic Mean	103.4	81.3	109.3
Variance-to-Mean Ratio	122.2	53.9	134.7
Chi-Squared p (One-Tailed) (a)	0.00	0.00	0.00
University of Illinois (UI) Chemistry Library Use			
Variance	101458.3	14197.0	123502.5
Arithmetic Mean	148.3	85.7	164.7
Variance-to-Mean Ratio	684.3	165.6	749.7
Chi-Squared p (One-Tailed) (a)	0.00	0.00	0.00
Science Citation Index (SCI) Total Citations			
Variance	863556465.9	2327903129.8	456727712.3
Arithmetic Mean	14627.6	25827.4	11680.3
Variance-to-Mean Ratio	59036.2	90133.1	39102.6
Chi-Squared p (One-Tailed) (a)	0.00	0.00	0.00
Science Citation Index (SCI) Impact Factor			
Variance	6.547	9.576	5.632
Arithmetic Mean	2.384	3.177	2.175
Variance-to-Mean Ratio	2.8	3.0	2.6
Chi-Squared p (One-Tailed) (a)	0.00	0.00	0.00

(a) At the standard 0.05 level of significance the chi-squared test rejects the null hypothesis of the Poisson distribution in favor of the alternative hypothesis of a compound Poisson, contagious distribution at a p below this level.

[Above : TABLE 6]

equal or approximate one, if the distribution fits the Poisson, due to this distribution's identity of the lambda, arithmetic mean, and variance all equaling each other. However, given this identity, it is easily seen that this test is actually a comparison of the actual variance of a dataset to its theoretical Poisson variance. Fisher's Poisson index-of dispersion test was also placed by Cochran (1954) within the standard system of null and alternative hypotheses. The hypotheses for Fisher's Poisson index of dispersion test have been defined by Elliott (1977, pp. 40-44). In the system presented by him, the null hypothesis is the Poisson distribution. If the variance is significantly less than the mean, Elliott defines the alternative hypothesis as "a regular distribution"; if the variance is significantly greater than the mean, he states the alternative hypothesis as "a contagious distribution." According to Elliott (1977, 46 and 50-51), the positive binomial distribution is the approximate mathematical model for a regular distribution, whereas the NBD is the most useful mathematical model for the diverse patterns of contagious distributions. However, since the NBD and other types of contagious distributions also arise from probabilistic heterogeneity as well as contagion, Snedecor and Cochran (1989, p. 200) also regard the Poisson variance test as one for a compound Poisson distribution consisting of elements and subsets with differing means.

Table 6 above presents the results of the Poisson index of dispersion tests of the distributions of the titles comprising the full set, biochemistry subset, and remaining subsets over the measures scientific value. Inspection of the Poisson results confirms the findings of the binomial index of dispersion tests of these same distributions. Thus, in all cases, the variance-to-mean ratios, which compare actual variance to theoretical Poisson variance, are all significantly greater than one. Therefore, the actual variance of the distributions is greater than the hypothesized Poisson variance. This not only indicates the operation of a contagious stochastic process but suggests that these distributions are compound Poisson distributions comprised of elements and subsets with different lambdas. Proof of the latter hypothesis is provided by a comparison of the arithmetic means or lambdas of the biochemistry subset to the arithmetic means or lambdas of the remaining subsets after the removal of the biochemistry subset. With both LSU faculty score and UI library use, the arithmetic means of the biochemistry subset are much lower than the corresponding arithmetic means of the remaining subsets, whereas with the two citation measures the opposite holds true: the arithmetic means of the biochemistry subset are much higher than the corresponding arithmetic means of the remaining subsets. This indicates three things: 1) the probabilities of the biochemistry subset differed from the probabilities of the remaining subsets; 2) the titles of the biochemistry subset had a much higher probability of being cited than being highly rated by the LSU chemistry faculty or used at the UI Chemistry Library; and 3) many of the citations to the biochemistry titles are exogenous to the analysis and result from Garfield's Law of Concentration. And, finally, the variance-to-mean ratios of the *SCI* impact factor are all much lower than these ratios for the other measures of scientific value. The impact factor ratios approach fairly close to one, indicating a closer fit to the Poisson distribution and a higher component of random error in the variance.

Summary and Conclusions

This section has validated Garfield's Law of Concentration, exploring its causation and implications. Garfield formulated this law on the basis of his analysis of total citations, deriving it off of Bradford's Law of Scattering on the distribution of articles on a given scientific subject over scientific journals. According Garfield's law, total citations to the journals of a given discipline concentrate on a relatively few titles, which form the citation core of the discipline, and then spread out in decreasing numbers over a long tail of other titles that exponentially increase in number as the total citations to them decrease. This law stipulates that, given the interdisciplinary nature of science, the journals comprising the citation tail of one discipline are those forming the citation cores of other disciplines, so that a small, interdisciplinary core of journals dominates the entire scientific journal system in terms of total citations. Garfield utilized total citations to identify the titles of this interdisciplinary core, which for the most part are large research journals.

The statistical tests of this section have corroborated Garfield's findings in respect to total citations. It was found that the frequency distribution of the sample's 120 journals across total citations is highly and positively skewed. A glance of Figure 3 above shows that the two titles dominant in total citations are, in descending order, the *Journal of Biological Chemistry* and the *Journal of the American Chemical Society*, which are both large research journals. To explore the causation underlying this frequency distribution, it was tested against the normal, binomial, and Poisson probability distributions. It was found that the variance of the distribution is far too high to be explained by random error. Statistical tests also revealed that the frequency distribution of total citations closely resembles the frequency distributions of LSU faculty score and UI library use in this respect.

Such high variance is indicative of powerful causes at work. One of these powerful causes is the stochastic process of probabilistic heterogeneity, which in itself is a function of Garfield's Law of Concentration. Taken together, Bradford's and Garfield's laws have three major implications for probabilistic and statistical citation analyses of scientific journals. First, scientific journal sets are fuzzy, and it is very difficult, if not impossible, to define a crisp set of journals for a given scientific discipline. This was already evident in the difficulties involved in creating the journal sample from those titles indicated as important by the LSU chemistry faculty. Second, the fuzziness of scientific journal sets results in citations exogenous to the purposes of the analysis. The most important source of such citations was the anomalous relationship of biochemistry to chemistry. This is evident from the fact that, whereas the *Journal of the American Chemical Society* was highest on LSU faculty score and UI library use, the *Journal of Biological Chemistry* was highest in total citations. And, third, Bradford's and Garfield's laws dictate that the distribution of the journals in a given scientific subject set is a compound one due to the differing probabilities of the component subject subsets. This aspect of scientific journal distributions was demonstrated by showing that the biochemistry subset had differing probabilities from the remaining subsets. The action of exogenous citations was visible in the biochemistry subset having a higher probability of being cited than being highly rated by the LSU chemistry faculty or used at the UI Chemistry Library.

The other powerful cause underlying Garfield's findings in respect to total citations is the stochastic process technically known in statistics as "contagion." This is a term that was first suggested by the study of the probability distributions of epidemics such as that done by Eggenberger and Pólya, but contagion later became more broadly used to designate situations where trials are not independent, because the occurrence of an event affects the probability of its further occurrence. Contagion has been shown to be the stochastic process, off which both Merton's Matthew Effect and Price's Cumulative Advantage Distribution derive. It has been noted that heterogeneity and contagion lead to identical types of probability distributions.

This section has made the case that the best probabilistic model for Garfield's Law of Concentration and other such bibliometric phenomena is the compound Poisson distribution. The reasons for this are the following. First, the Poisson process is aptly suited for describing such information phenomena as the occurrence of citations, library uses, publication of articles, etc. It allows one to base one's calculations on visible data without having to make estimates of invisible zero events. Second, the Poisson distribution is the model for events having a low probability of occurrence, and it was proven that the probability of an article or a journal being cited is extremely low. Third, due to Bradford's and Garfield's laws, most scientific journal sets consists of subject subsets governed by differing probabilities. And, fourth, compound Poisson distributions model both stochastic processes of heterogeneity and contagion. It is readily admitted that the ideas and conclusions of this section are not original. Thus, Brookes (1977) explored Bradford's Law as a very mixed Poisson model ultimately indefinable in conventional frequency terms. Of the standard probability distributions, Brookes (1980, pp.219-220) considered the negative binomial as the one best describing Bradford's Law. For his part, Bookstein (1997, p. 10) advanced the family of compound Poisson distributions as uniquely suited for modeling the ambiguous phenomena found in information science.

Garfield's findings in respect to the impact factor as well as his utilization of this measure to select journals for coverage by ISI indexes have also been validated by the statistical tests presented in this section. The major finding resulting from his research on the nature of the impact factor was the low "impact" or citation rate of most papers. To describe this phenomenon, he formulated his "Garfield's Constant," which is the ratio of the number of references processed in a given year to the number of different items cited by those references. It is, thus, the arithmetic mean of the number of citations to those papers actually cited in given year. Using standard statistical theory and techniques, this section connected this arithmetic mean to the low overall probability of scientific papers being cited and demonstrated that Garfield's Constant is actually an estimate of the Poisson lambda of those papers actually cited in a given year on the assumption of their probabilistic homogeneity. Garfield's Constant was shown as setting a limit on journal impact factors, which are estimates of the mean citation rates of the articles of individual journals over a given year. Since articles differ vastly in their probabilities of being cited, arithmetic means are skewed upward by the relatively few articles causing high variance by being much more frequently cited than the others, and this was reflected in 1993 by only 12.2% of the journals having impact factors higher than Garfield's Constant for that year. The impact factors of the 120 sample journals followed the same type of distribution, and it is to be noted that in all cases the arithmetic means of these impact factors were significantly higher than the other measures of central tendency such as the median and geometric mean.

It is from this perspective that one can evaluate Garfield's use of the impact factor in journal selection. Garfield used total citations to identify those large research journals posited by his Law of Concentration to comprise the relatively small interdisciplinary core dominating the entire scientific journal system. He used the impact factor to counter the size advantage of these core journals, so that he could identify those journals that had a propensity to publish significant articles. Since the impact factor is based upon the arithmetic mean, which is skewed upward by the often cited articles causing high variance, he was using a measure uniquely suited to do this. Since review articles have a higher probability of being cited than other types of articles, Garfield found review journals generally dominant in terms of the impact factor. This finding was corroborated in this section, where it was shown that the small review journal, *Chemical*

Table 7. Relative Percentages of Aggregate Journal Quality Measures Attributable to Subsets Defined by Categorical Variables Biochemistry, Nationality, and Association Affiliation

			LSU Faculty Score		*UI Library Use*		*SCI Total Citations*		*SCI Impact Factor*	
Subset Name	***No. Titles***	***% Titles***	***% Aggregate***	***Aggregate-Titles % Ratio***	***% Aggregate***	***Aggregate-Titles % Ratio***	***% Aggregate***	***Aggregate-Titles % Ratio***	***% Aggregate***	***Aggregate-Titles % Ratio***
Biochemistry										
Biochemistry Subset	25	20.8%	16.4%	0.8	12.0%	0.6	36.8%	1.8	27.8%	1.3
Non-Biochemistry Subsets	95	79.2%	83.6%	1.1	88.0%	1.1	63.2%	0.8	72.2%	0.9
Nationality										
US Subset	53	44.2%	58.0%	1.3	57.3%	1.3	61.6%	1.4	49.8%	1.1
Foreign Subset	67	55.8%	42.0%	0.8	42.7%	0.8	38.4%	0.7	50.2%	0.9
Association Affiliation										
Association Subset	52	43.3%	62.1%	1.4	68.8%	1.6	65.1%	1.5	53.0%	1.2
Non-Association Subset	68	56.7%	37.9%	0.7	31.2%	0.6	34.9%	0.6	47.0%	0.8
Nationality and Association Affiliation										
US Association	27	22.5%	44.3%	2.0	50.8%	2.3	49.4%	2.2	32.3%	1.4
Foreign Association	25	20.8%	17.8%	0.9	18.1%	0.9	15.7%	0.8	20.7%	1.0
US Non-Association	26	21.7%	13.6%	0.6	6.6%	0.3	12.1%	0.6	17.5%	0.8
Foreign Non-Association	42	35.0%	24.3%	0.7	24.6%	0.7	22.8%	0.7	29.4%	0.8

Table 8. Comparison of Mean and Median as Journal Quality Central Tendency Measures for Subsets Defined by Categorical Variables Biochemistry, Nationality, and Association Affiliation

		LSU Faculty Score		*UI Library Use*		*SCI Total Citations*		*SCI Impact Factor*	
Set and Subset Names	***Central Tendency Measure***	***Measure***	***Ratio to Full Set Measure***	***Measure***	***Ratio to Full Set Measure***	***Measure***	***Ratio to Full Set Measure***	***Measure***	***Ratio to Full Set Measure***
Full Set	Mean	103.4	NA	148.3	NA	14627.6	NA	2.384	NA
	Median	55.5	NA	53.5	NA	5100.5	NA	1.599	NA
Biochemistry									
Biochemistry Subset	Mean	81.3	0.8	85.7	0.6	25827.4	1.8	3.177	1.3
	Median	53	1.0	51	1.0	6436	1.3	2.307	1.4
Non-Biochemistry Subsets	Mean	109.3	1.1	164.7	1.1	11680.3	0.8	2.175	0.9
	Median	56	1.0	56	1.0	5062	1.0	1.519	0.9
Nationality									
US Subset	Mean	135.7	1.3	192.4	1.3	20386.9	1.4	2.690	1.1
	Median	78	1.4	59	1.1	6380	1.3	1.746	1.1
Foreign Subset	Mean	77.9	0.6	113.3	0.8	10071.7	0.7	2.141	0.9
	Median	51	0.7	51	1.0	5062	1.0	1.399	0.9
Association Affiliation									
Association Subset	Mean	148.2	1.4	235.5	1.6	21987.8	1.5	2.917	1.2
	Median	101.5	1.8	82.5	1.5	8500	1.7	1.819	1.1
Non-Association Subset	Mean	69.2	0.7	81.6	0.6	8999.1	0.6	1.976	0.8
	Median	46.5	0.8	40	0.7	3014	0.6	1.487	0.9
Nationality and Association Affiliation									
US Association	Mean	203.8	2.0	334.5	2.3	32147.6	2.2	3.423	1.4
	Median	194	3.5	179	3.3	10228	2.0	2.768	1.7
Foreign Association	Mean	88.2	0.9	128.6	0.9	11015.3	0.8	2.370	1.0
	Median	69	1.2	59	1.1	5832	1.1	1.323	0.8
US Non-Association	Mean	65.0	0.6	44.9	0.3	8173.9	0.6	1.929	0.8
	Median	43.5	0.8	43	0.8	2420.5	0.5	1.446	0.9
Foreign Non-Association	Mean	71.8	0.7	104.2	0.7	9510.0	0.7	2.005	0.8
	Median	46.5	0.8	40	0.7	3269	0.6	1.525	1.0

[Above: TABLES 7-8]

Reviews, is highest on this measure. For various reasons—among them the difficulty of classifying articles into "citable" and "non-citable" for the denominator—Garfield did not consider the impact factor a very accurate measure, restricting himself to one decimal place in his employment of it instead of using the three decimal places published in the *JCR*s. Such a view of the impact factor was validated by statistical tests showing that the impact factor frequency distributions conformed much more closely than those of the other measures to the binomial and Poisson distributions. This is indicative that a larger component of the variance in the impact factor distributions resulted from random error.

4. SOURCES OF VARIANCE

This section continues the distributional analysis of the preceding section. A key hallmark of compound Poisson distributions is an excess of variance over that expected as a result of random error. Therefore, the main focus of this section is to pinpoint more precisely the sources of variance in the frequency distributions of the journals across all four measures of journal importance. The method of doing so is based upon the logic of the compound Poisson distribution. It is to separate out from the sample important subsets of journals and estimate through measures of central tendency how their underlying probabilities compare to those of other subsets. For this purpose, the sample of journals under analysis is broken down into subsets defined by the following categorical variables: biochemistry vs. non-biochemistry titles; US vs. foreign titles; association vs. non-association titles; as well as US association, foreign association, US non-association, and foreign non-association titles. In addition to these categorical variables, the size, age, and social status of the journals as well as their proportion of review articles are also examined in respect to their role in variance causation.

Subset Analysis

The first step in the analysis was to determine the role of the subsets defined by the categorical variables of interest in variance causation. Tables 7 and 8 above summarize the results of this step of the analysis. Both these tables are structured in the same way and consist of four basic divisions. The first division is disciplinary, and its main purpose to elucidate the effect of the subset of biochemistry titles upon the variance of the four measures of journal importance. This facet of variance was analyzed in the preceding distributional section in order to demonstrate the effect of Bradford's Law of Scattering and Garfield's Law of Concentration on the structure of scientific journal sets and their citation patterns. Taken together, these two laws ensure that scientific journal subject sets are both compound and fuzzy, being composed of different disciplinary subsets with many citations to their component journals being exogenous to the set's subject. The next three divisions of Tables 7 and 8 comprise a logical unit, which is designed to elucidate the national and institutional causation of variance. These divisions have the following aims: the second division compares the role of US vs. foreign journals in variance causation; the third division analyzes the amount of variance attributable to association vs. non-association journals; and the fourth division combines the national and institutional categorical variables to investigate the effect of US association, foreign association, US non-association, and foreign non-association journals on variance. The nationality of the journal was determined by country of publication, and, therefore, multinational publishers could have titles categorized as either US or foreign. A journal was defined as "association" by being published either by a scientific association or by another type of publisher for a scientific association.

Table 7 presents the results of comparing a given subset's percentage of the sample's aggregate amount of a scientific journal value measure to its percentage of the sample's total number of titles. The key measure in this table is the aggregate-titles % ratio. Using the biochemistry subset and LSU faculty score as an example, the ratio can be explained in the following way. The 25 titles in the biochemistry subset, which represent 20.8% of the sample's 120 titles, accounted for 16.4% of the aggregate amount of LSU faculty score for these 120 titles. Dividing 16.4% by 20.8% yields an aggregate-titles % ratio of 0.8, which indicates that the biochemistry subset accounted for a smaller percentage of LSU faculty score than its percentage of titles. Comparing these ratios across the four measures of journal importance shows that the biochemistry subset accounted for a smaller percentage of LSU faculty score and UI library use than its percentage of titles but a higher percentage of *SCI* total citations and *SCI* impact factor than its percentage titles. It is exactly the opposite for the non-biochemistry subsets. These results corroborate the findings of preceding distributional section on the biochemistry subset and indicate that many of the citations to the biochemistry journals increasing the variance of the citation measures were exogenous to the interests of the LSU chemistry faculty and the patrons of the UI Chemistry Library.

In contrast to the biochemistry subset, the results for the subsets defined by nationality and association affiliation were consistent across all four measures scientific journal value. Thus, the aggregate-titles % ratios were all higher than one for US journals and below one for foreign journals. The pattern is the same the subset defined by association affiliation, where the ratios are all above one for the association journals and below one for the non-association titles. However, the full implications of these patterns do not become clear until the nationality and association variables are combined to form more precise sets in the fourth division of the table. Here it is seen that only the aggregate-titles % ratios of the US association journals are consistently above one for all four measures of scientific value. It is also interesting to note that these ratios for the foreign association titles are very close to one and consistently above the equivalent ratios of both the US and foreign non-association journals. From this analysis it is possible to conclude that much of the variance in all four measures of journal importance is attributable to journals affiliated with US scientific associations.

These findings were corroborated by a series of tests, whose results are summarized in Table 8. The tests involve comparing two measures of subset central tendency—the arithmetic mean and the median—to the equivalent measures of central tendency for the entire sample of 120 journals. The comparison is made by means of a ratio calculated by dividing the subset measure of central tendency by the equivalent full set measure of central tendency. The arithmetic mean was chosen because of its close theoretical association with the concept of probability. Rietz (1927, pp. 16-17) established this connection through what he termed "the *mathematical expectation* of the experimenter or the *expected value* of the variable," stating that "the mathematical expectation of a variable *x* and its mean value from the appropriate theoretical distribution are identical" (p. 16). However, as was demonstrated in the preceding distributional section, the arithmetic mean is affected in positively skewed distributions by observations high in the value of the variable. These dominant observations cause the arithmetic mean to be higher than the modal value of the distribution, around which the bulk of the observations are concentrated. This modal value is better estimated by the median, and an arithmetic mean significantly higher than the median is indicative of the type of compound Poisson distributions with which we are dealing. Examination of Table 8 reveals that this condition is met for all the distributions except the use of US non-association titles at the UI Chemistry Library.

From the perspective of the previous findings, the two key findings in Table 8 are the following. First, the ratios of the biochemistry mean and median to their full sample equivalents are above one for both *SCI* total citations and *SCI* impact factor. This indicates that biochemistry titles had a higher probability of being cited than the other titles. Second, of the subsets defined by both nationality and association affiliation, only the US association journals have the ratios of both their mean and median to the full sample mean and median higher than one for all four measures of journal importance. This serves as proof that the US association journals had a greater probability of being highly rated by the LSU chemistry faculty, used at the UI Chemistry Library, and cited than the other types of titles. It is interesting to note that the foreign association titles tended to perform better in terms of these ratios than either the US or foreign non-association journals for all measures of journal importance except the impact factor particularly in respect to the ratios of the subset median to full sample median.

Causation of Variance

This subsection will explore the underlying reasons why certain categories of journals have a greater probability of scoring higher on measures of journal importance than others. Two types of reasons will be considered: journal size and sociological factors.

The Role of Journal Size

Garfield had a conflicted attitude toward the relationship of journal size to journal significance. It has been pointed out above (p. 8-9) that in their seminal paper on the impact factor Garfield and Sher (1963) regarded total citations to journals as largely a function of the number of their articles, dismissed ranking journals by total citations as not much more sophisticated than ranking them by number of their articles, and stated that the first step in obtaining a meaningful measure of journal importance is to control for size by calculating the mean number of citations per article. Yet Garfield (1970, May 6) stated that journals ranked highest by number of source articles invariably prove to be journals considered significant by most readers, and Garfield (1973, Sept. 26) found that heavily cited articles or "citation superstars" tend to concentrate in 15 large research journals such as the *Journal of Biological Chemistry*. Garfield (1970, May 6) recommended that the less significant journals merge with other small journals to form larger journals that "tend to acquire a special significance, due possibly to greater exposure" (p.6). This policy was urged by Garfield (1997) upon Third World countries, which, he thought, suffer from publishing dozens of marginal journals and should combine the best material into larger regional journals to achieve "a critical mass" (p. 640).

Journal size has two facets: physical and temporal. The physical facet is the number of physical units—pages, source items, etc.—published by the journal per year, whereas the temporal facet is the number of years the journal has been publishing these units. Both these facets of size are controlled by the impact factor. The temporal facet is controlled by restricting the backfile to the two years preceding the impact factor year under consideration, whereas the physical facet is controlled by estimating the mean citation rate of the citable source items published during these two years. Therefore, it is necessary to analyze the effect of both these facets of journal size on the measures of journal importance. In this section, the measure of the physical facet is the number of source items published in the sample journals in 1993, and the measure of the temporal facet is the number of backfile years prior to 1993.

Table 9. Pearson r Correlation Matrix of Number of Source Items and Years of Backfile with LSU Faculty Score, UI Library Use, SCI Total Citations, and SCI Impact Factor

		LSU Faculty Score	*UI Library Use*	*SCI Total Citations*	*SCI Impact Factor*
Number of Source Items	*Correlation*	0.58	0.73	0.83	0.02
	R-Squared	0.33	0.53	0.68	0.00
Years of Backfile	*Correlation*	0.27	0.41	0.47	-0.09
	R-Squared	0.07	0.17	0.23	0.01

Correlations above 0.254 are significant at the 0.01 level.

Table 10. Comparison of Means and Medians of Number of Source Items and Years of Backfile for Journals Classed within Subsets Defined by Categorical Variables Biochemistry, Nationality, and Association Affiliation

		Number of Source Items		*Years of Backfile*	
Set and Subset Names	*Central Tendency Measure*	*Measure*	*Ratio to Full Set Measure*	*Measure*	*Ratio to Full Set Measure*
Full Set	Mean	507.3	NA	37.9	NA
	Median	290	NA	29.5	NA
Biochemistry					
Biochemistry Subset	Mean	632.8	1.2	37.3	1.0
	Median	199	0.7	31	1.1
Non-Biochemistry Subsets	Mean	474.3	0.9	38.0	1.0
	Median	303	1.0	29	0.9
Nationality					
US Subset	Mean	575.4	1.1	35.8	0.9
	Median	317	1.1	30	1.0
Foreign Subset	Mean	453.4	0.9	39.5	1.0
	Median	271	0.9	28	0.9
Association Affiliation					
Association Subset	Mean	649.7	1.3	51.1	1.3
	Median	357	1.2	35	1.2
Non-Association Subset	Mean	398.4	0.8	27.7	0.7
	Median	272.5	0.9	26	0.9
Nationality and Association Affiliation					
US Association	Mean	825.2	1.6	45.2	1.2
	Median	503	1.7	34	1.2
Foreign Association	Mean	460.2	0.9	57.5	1.5
	Median	244	0.8	40	1.4
US Non-Association	Mean	316.0	0.6	26.1	0.7
	Median	185	0.6	25.5	0.9
Foreign Non-Association	Mean	449.5	0.9	28.8	0.8
	Median	279.5	1.0	27	0.9

[Above : TABLES 9-10]

Table 9 above is a Pearson correlation matrix showing the relationship of both facets of journal size to the four measures of journal importance under consideration. The first thing to be noted is that the impact factor has successfully controlled for both facets of journal size, because the correlation coefficients of the number of source items and the years of backfile with this measure are both close to zero and insignificant. In respect to the other three measures of journal importance, it is the physical aspect of journal size that appears to have the greater effect, for the correlations and coefficients of determination are all much higher for number of source items than for years of backfile. While no statistical explanation of this phenomenon can be provided for LSU faculty score, the fact that this measure is positively and significantly correlated with both facets of journal size validates Garfield's anecdotal observations on the relationship of the size of journals to scientists' evaluation of their significance. A statistical explanation can be provided as to why the physical facet of size is more highly correlated with both UI Chemistry Library use and *SCI* total citations than the temporal facet. The journal backfiles ranged from 3 years to 161 years, and the distribution of the journals over these years was positively skewed. Thus, the mean number of years was 37.9, whereas the median number of years was 29.5. However, the bulk of library use and citations in science concentrate on the most recent years. For example, post-1979 use accounted for 73.8% of the total UI Chemistry Library use of the sample titles (Bensman and Wilder, 1998, p. 2002), and post-1979 citations accounted for 76.8% of *SCI* total citations in 1993 (Institute for Scientific Information, 1994a, p. 62). This 14-year period is considerably shorter than even the median number of years. It is probably this concentration of use and citations on the later years that reduced the correlations of years of backfile with UI Chemistry Library use and *SCI* total citations. However, there is also the possibility that the physical and temporal facets of journal size can operate interactively to create a total effect of size.

The two facets of journal size are related to the subsets defined by the categorical variables under consideration in Table 10 above. Once again, the method of analysis is to compare the subset means and medians to the mean and median of the total sample of 120 journals through a ratio calculated by dividing the subset mean and median by the equivalent sample measure. The key findings can be summarized under two rubrics. First, in respect to the biochemistry subset, it is seen that the mean number of source items is greater than the sample mean number of source items, whereas the median number of source items is smaller than sample median number of source items. However, in respect to years of backfile, both the mean and median ratios indicate virtually no difference between the biochemistry subset and the total sample on this facet of journal size. The conclusion to be drawn is the higher number of citations to biochemistry titles is due to a few journals large in terms of the physical facet of size. Second, concerning the nationality and association categorical variables, the ratios reveal that only the US association journals are larger on both facets of journal size than the sample as a whole. Thus, the higher LSU faculty ratings, UI Chemistry Library use, and *SCI* citations of the US association titles is partly a function of their larger size. The foreign association titles have longer backfiles, but it has been seen that the temporal facet of journal size has less effect on faculty ratings, library use, and citations than the physical aspect of journal size.

Sociological Factors

Over the years studies have revealed that US academic research institutions form a highly stratified and stable social system. These studies entailed the evaluation of the "quality" of the faculty of university graduate programs. The traditional method of evaluating such quality was peer ratings. This method was pioneered by the noted psychologist Cattell (1906; 1910), who statistically constructed a list of the 1,000 most-eminent American scientists through a survey of the leading representatives of 12 scientific disciplines. He first ranked universities and then academic departments by the number of these scientists at them. Cattell's work was further developed by Hughes (1925), president of Miami University in Ohio. In 1924 Hughes had Miami University faculty members in 20 disciplines draw up a list of major doctorate institutions, select 40-60 professors in each field throughout the US to serve as raters, and on the basis of the responses construct a statistical ranking of the institutions offering the doctorate in the 20 disciplines.

The method of obtaining and presenting peer ratings of the quality of university faculty was standardized in the evaluation of US graduate programs directed by Cartter (1966) under the auspices of the American Council on Education in 1964. In his discussion of the evaluation, Cartter defended peer ratings against charges that they were mere surveys of impressionistic opinions heavily influenced by institutional, sectional, historical, and social biases. As one respondent put it, quoting Dr. Johnson. "a compendium of gossip is still gossip" (p. 8). To counter such a view of peer ratings, Cartter argued that quality is an elusive attribute not amenable to being quantified by such "objective" measures such as size of endowment, books in the library, publication record of faculty, number of Nobel laureates, etc. In his opinion, such "objective" measures of quality are for the most part "subjective" measures once removed. In the 1964 questionnaire (pp. 126-127) the raters were asked to judge "<u>the quality of the graduate faculty</u>," considering only their "scholarly competence and achievements," and assign them grades ranging from 1 ("Distinguished") to 6 ("Not sufficient to provide acceptable doctoral training"). The respondents were asked to limit the number of "Distinguished" ratings to no more than 5. In addition, the raters were given the option of not evaluating the programs by marking their questionnaire "Insufficient information" in the appropriate box. The grades were then assigned the following numerical weights (p. 15): Distinguished—5; Strong—4; Good—3; Adequate—2; Marginal—1; Not sufficient to provide adequate doctoral training—0. These numerical weights were then averaged to calculate a program's score. The 1964 method of peer rating was essentially implemented by all the subsequent major evaluations of the scholarly quality of US graduate program faculty, which included the following: the 1969 American Council on Education evaluation of graduate education directed by Roose and Andersen (1970); the 1981 assessment of US research-doctorate programs done under the auspices of the American Council of Learned Societies, American Council on Education, National Research Council, and Social Science Research Council (Jones, Lindzey, and Coggeshall, 1982); and the 1993 evaluation of US research-doctorate programs sponsored by the National Research Council (Goldberger, Maher, and Flattau, 1995).

A notable feature of these peer ratings is their remarkable stability over time particularly at the top. This is certainly the case in chemistry. A historical analysis done by this author revealed that of the top 15 chemistry programs in the 1924 ratings, 11—California at Berkeley, Cal Tech, Chicago, Columbia, Cornell, Harvard, MIT, Stanford, Illinois, Wisconsin, and Yale—remained consistently in the top 15 by peer ratings of scholarly quality in 1964, 1969, 1981, and 1993. Of these 11 programs, 8—Chicago, Columbia, Cornell, Harvard, MIT, Illinois, Wisconsin, and Yale—were listed among the 9 top chemistry departments by Cattell (1910, p.685). The continued dominance of these programs is all the more remarkable, given the continuous increase in the number of chemistry programs being rated: 96 in 1964; 125 in 1969; 145 in 1981; and 168 in 1993. Moreover, this stability was not restricted to the top but manifested itself throughout the rankings. As part of the 1993 evaluation, the National Research Council (NRC) assembled a huge electronic database, which was given to this author for testing. This database contained not only the data collected for the 1993 evaluation but also for the 1964, 1969, and 1981 evaluations. Using this database, a correlation matrix was constructed from all four peer ratings of the scholarly quality of program faculty, and the correlations ran from a low of 0.78 between 1964 and 1981 to a high of 0.93 between 1981 and 1993. In general, the closer the rating years were together, the higher the correlation, showing a slow change over time (Bensman and Wilder, 1998, p. 174).

A major advance in the evaluation of US academic research programs was marked by the 1981 assessment (Jones, Lindzey, and Coggeshall, 1982), which for the first time published together with the subjective ratings of the scholarly quality of program faculty such objective measures as number of faculty, number of graduates, number of graduate students, size of library, amount of research support, and the publication record of the program faculty. The relationship of these objective measures to the subjective ratings was analyzed. Of particular interest here is the finding that these subjective ratings are positively associated with program size as measured by the number of program faculty. In chemistry the correlation was 0.64 (p. 166).

Two publication measures of program faculty were presented in the 1981 assessment (Jones, Lindzey, and Coggeshall, 1982, pp. 15, 27-29, and 220-237). Both these publication measures were provided by a subcontractor, Computer Horizons, Inc. (CHI), which developed them from data provided by Garfield's Institute for Scientific Information (ISI). The first was the number of articles published in 1978-1979 in journals covered by the appropriate ISI citation index and attributable to the program. It should be emphasized that this was not a direct count of articles written by the faculty of a given program but of articles published in journals classified in the same discipline as the program by faculty affiliated with the same university as the program but not necessarily in the program itself. Articles published in multidisciplinary journals such as *Science* and *Nature* were apportioned to programs in accordance with the characteristic mix of subject matters in these journals.

The other publication measure of the 1981 assessment was the estimated "overall influence" of the articles attributed to the program. CHI (Narin, 1976, pp. 183-219) developed its "influence" method in a report prepared for the National Science Foundation. In this report it criticized Garfield's impact factor as suffering from three basic faults (p. 184). First, although the impact factor corrects for journal size, it does not correct for average length of articles, and this caused journals, which published longer articles such as review journals, to have higher impact factors. As a result, CHI concluded, the impact factor could not be used to establish a "pecking order" for journal prestige. The second limitation of the impact factor was that all the citations are considered equal, whereas it seems more logical to assign citations from prestigious journals a higher weight than citations from less prestigious ones. And, finally, there is no normalization for the different referencing characteristics of the different disciplines. To correct for the impact factor's faults, CHI developed its "influence" method, whose key measure is "influence per paper" or the weighted number of times an average paper in a given journal is cited, where the influence is based on the influence—or citation rate—of the citing journal. CHI

(Anderson, Narin, and McAllister, 1978) tested their influence method against the 1969 Roose-Andersen ratings of US academic research programs in 10 disciplines and found that these ratings correlated most highly with "total influence" or the total number of papers multiplied by the influence per paper. The analysis indicated that the Roose-Andersen ratings had two additive components: bibliometric size and bibliometric quality. Moreover, CHI also found that departmental ratings within a university were not independent but associated with the overall biblimetric size of the university. The 1981 assessment utilized "total influence" as its second publication measure, and the results in chemistry demonstrated that the citation weighting captured an additional component of subjective quality. Thus, whereas the correlation of the subjective rating of the scholarly quality of chemistry program faculty with number of 1978-1979 *SCI* articles attributed to this faculty was 0.80, this correlation rose to 0.86, once the total influence of these articles was taken into account.

A unique opportunity to test the relationship of the 1981 assessment's peer ratings of the scholarly quality of program faculty to the actual number of citations to programs presented itself, when Davis and Papanek (1984) published a ranking by total citations of 122 Ph.D.-granting departments in economics. To avoid a bias resulting from the selection of a single year, Davis and Papanek averaged the total number of citations these departments received in the *Social Sciences Citation Index* (*SSCI*) in 1978 and 1981, noting, however, that "there is remarkably little change in rankings between these years" (p. 226). Bensman (1985, pp. 22-23) found that 88 of these departments had also been evaluated by the 1981 assessment, and he analyzed the citation patterns to these 88 departments. He found the usual highly skewed distribution, with the top 11 of these departments or 12.5% accounting for 53.3% of the total citations received by all 88 departments. Moreover, Bensman also found that 7 of these departments were located at 7 universities, which studies had found to be consistently among the top 10 universities by peer ratings since 1924 every time the departmental ratings had been aggregated into institutional scores. The correlation of the 1981 peer ratings of the scholarly quality of the program faculty in economics with the total citations to their publications was a stunning 0.92, causing Bensman to conclude that "citations and peer ratings appear to be virtually the same measurement, and the concentration of citations on given departments may well be just as much a function of the overall prestige of their respective universities as high ratings in peer evaluations" (p. 23). It thus seems that citations may have all the same historical, institutional, and social biases as peer ratings.

Tests using chemistry data from the 1993 evaluation of US research-doctorate programs by the National Research Council (NRC) (Goldberger, Maher, and Flattau, 1995) corroborated the above findings on the relationship of peer ratings of scholarly quality of program faculty to bibliometric size and citations. From the perspective of the focus of this paper, it is important to point out that Jonathan Cole and Harriet Zuckerman, two students of Garfield's mentor, Robert K. Merton, putative founder of the sociology of science, played leading roles in the planning and implementation of the NRC evaluation. The NRC evaluation once again found fairly strong positive relationships between the peer ratings of scholarly quality and program size as measured by number of program faculty. For chemistry the reported correlation was 0.71 (p. 436). The NRC evaluation constructed two bibliometric measures of interest from data supplied by the Institute for Scientific information: 1) the ratio of the total number of program publications in the period 1988-1992 to the number of program faculty; and 2) the ratio of the total number of program citations in the period 1988-1992 to the number of program faculty. For purposes of simplification, the first measure will be called "mean publications per faculty," and the second will be designated "mean citations per faculty." The NRC data is downloadable off the Web into spreadsheets, and this capability was utilized to convert the two reported measures into ones capturing bibliometric size by multiplying them by the number of program faculty. These two new measures will be designated "total program publications" and "total program citations." Statistical tests of the five objective measures—number of program faculty, mean publications per faculty, total program publications, mean citations per faculty, and total program citations—found that their distributions had variances significantly higher than their means, indicating compound Poisson distributions modeling probabilistic heterogeneity and contagion. In both cases, introducing bibliometric size increased the correlations of the bibliometric measures with the peer ratings of the scholarly quality of the program faculty. Thus, the reported correlation of these peer ratings with mean publications per faculty was 0.82, and this increased to 0.87, when this measure was converted to total program publications. The same happened when mean citations per faculty was converted to total program citations. Here the correlation rose from a reported 0.81 to 0.91. Once again, as with the 1981 peer ratings of economics departments, the correlation of 0.91 is so high that it indicates that peer ratings and total citations are equivalent measures and have all the same biases.

To test for probabilistic heterogeneity, the NRC sample of 168 chemistry programs was split into two subsets: 1) the 11 elite chemistry programs that had consistently appeared in the top 15 chemistry programs by peer ratings ever since 1924; and 2) the other 157 programs. In 1993 the 11 elite programs ranked from 1 to 10 and then 12 by peer ratings of the scholarly quality of program faculty. Their median peer rating score was 4.55 compared to the median of 2.35 for the other 157 programs. The elite programs tended to be much bigger with a median of 30 faculty, compared to a median of 20 faculty for the 157 programs. Moreover, they had much higher mean publication rates, with a median of 18.8 on this measure, which was about double the median of 9.1 for the other 157 programs. Their greater size and higher mean publication rates caused them to be dominant in terms of total program publications. Their median total program publications was 529 in contrast to 176 for the other 157 programs. Although the 11 elite programs comprised only 6.5% of the chemistry programs evaluated in 1993, they accounted for 14.0% of the total publications. The same pattern held true in respect to citations. The median of the mean citations per faculty of the 11 elite departments was 165.5, compared to 36.6 for the other 157 programs. This led to their dominance in total program citations, with a median of 4,099 compared to one of 735 for the other 157 programs. By themselves the 11 elite programs accounted for 20.3% of the total citations received by the 168 chemistry programs evaluated by the NRC in 1993.

Noting the high concentration of citations on both the traditionally elite chemistry programs and US association journals, Bensman and Wilder (1998, p. 176) deduced that citations were concentrating on US association journals, because the faculty of the elite chemistry programs were publishing in these journals. They also stated that this was a major reason why US association journals were so highly rated by the LSU chemistry faculty, who were responding to the dictates of their social stratification system in rating journals. As a result, they concluded that the scientific journal system was in many respects an external manifestation of the underlying social stratification system of science. This conclusion has been corroborated here, as it has been shown that US association journals are a major source of variance in the distributions of all four measures of scientific journal value under consideration—LSU chemistry faculty ratings, UI Chemistry Library use, *SCI* total citations, and *SCI* impact factor. The analysis also revealed that foreign association journals resemble US association journals in that they perform better on these measures than either US or foreign journals without association affiliations. Relating these findings to the results of Garfield's research set forth in the first section (pp. 6-8 above), it is significant that the social stratification system of chemistry bears a marked resemblance to the structure of the scientific journal system discovered by Garfield with total citations, which measure scientific journals as sociological entities. In both cases, the system is dominated by a small, stable core of physically large sociological entities. This structure caused Garfield to formulate his Law of Concentration, and the similarity of the structures suggests that both systems result from similar stochastic processes. Merton interpreted these processes as the operation of the principle of cumulative advantage, which he called the "Matthew Effect." For his part, Price identified Merton's Matthew Effect with the contagious form of the negative binomial distribution (NBD), which is a compound Poisson distribution modeling not only contagion but also probabilistic heterogeneity. Describing this distribution as "double-edged" in that success is rewarded and failure punished, Price counter-posed a "single-edged" principle of cumulative advantage described by the beta function and modeled by a compound Poisson process in which success is rewarded but failure is not punished. Statistical tests of the NRC data indicated some form of a compound Poisson distribution. Moreover, a careful reading of the seminal article by Merton (1968) on the Matthew Effect indicates that he discussed the principle of cumulative advantage from two perspectives: 1) the misallocation of peer recognition of the contributions of scientists due to prior achievements; and 2) the mechanism underlying the rise of the highly stratified social system of US academic science. Both these processes appear operative in the data analyzed in this paper, and the extraordinarily high correlations of 0.9 of peer ratings with total citations indicate that the two measures are virtually equivalent measures of the same phenomena. Therefore, citations cannot be considered an objective gauge of scientific importance but must be regarded as a sociometric measure having the same historical, institutional, national, and social biases as subjective peer ratings.

Concentration and Variance

One consequence of excess variance is that the phenomenon being measured concentrates on a few members of a given set or subset. This is a characteristic feature of bibliometric distributions, and it is best epitomized by Trueswell's 80/20 Rule, whereby 80% of the use of the items in a library collection is satisfied by 20% of these items. These dominant members are the ones farthest above the mean and therefore account for a large proportion of the variance. Therefore, an analysis of their characteristics is very helpful in understanding the causation of the variance.

To do this analysis, the sample of 120 journals was split into two subsets for each measure of journal importance under consideration for comparative purposes: 1) the 12 (10%) journals ranking highest on the measure; and 2) the other 108 (90%) journals. The results of the analysis are summarized in Tables 11-13 below. Table 11 shows the level of concentration of each measure of journal importance on the upper 10% of titles. It also compares the top subset to the bottom subset on two measures of central tendency—the arithmetic mean and median. In each case the top 10% of the titles accounted for higher percentage of the aggregated measure of journal importance than their proportion of the titles of the sample—36.1% of LSU faculty score, 57.4% of UI library use, 56.3% of SCI total citations, and 36.4% of SCI impact factor. This was accomplished by the top 10% of the titles scoring higher on both measures of central tendency than the bottom 90%.

Table 11. Comparative Concentration and Central Tendencies Structure of the Top 12 (10%) vs. Bottom 108 (90%) Journals Ranked by LSU Faculty Score, UI Library Use, SCI Total Citations, and SCI Impact Factor				
Quality Measure	***Stratum***	***% Aggregated Measure***	***Arithmetic Mean***	***Median***
LSU Faculty Score	***Top 12 (10%) Titles***	36.1%	373.4	355.5
	Bottom 108 (90%) Titles	63.9%	73.4	49
UI Library Use	***Top 12 (10%) Titles***	57.4%	850.6	555.5
	Bottom 108 (90%) Titles	42.6%	70.2	50
SCI Total Citations	***Top 12 (10%) Titles***	56.3%	82390.6	60774
	Bottom 108 (90%) Titles	43.7%	7098.4	3730
SCI Impact Factor	***Top 12 (10%) Titles***	36.4%	8.686	6.481
	Bottom 108 (90%) Titles	63.6%	1.684	1.487

Table 12. Comparative Structure of the Top 12 (10%) vs. Bottom 108 (90%) Journals Ranked by LSU Faculty Score, UI Library Use, SCI Total Citations, and SCI Impact Factor in Terms of Biochemistry, United States, and Association Affiliated Categorical Variables							
Quality Measure	***Stratum***	***Biochemistry Titles***		***United States Titles***		***Association Affiliated Titles***	
		Number Titles	***Percent Titles***	***Number Titles***	***Percent Titles***	***Number Titles***	***Percent Titles***
LSU Faculty Score	***Top 12 (10%) Titles***	1	8.3%	10	83.3%	11 (a)	91.7%
	Bottom 108 (90%) Titles	24	22.2%	43	39.8%	41	38.0%
UI Library Use	***Top 12 (10%) Titles***	2	16.7%	8	66.7%	9 (b)	75.0%
	Bottom 108 (90%) Titles	23	21.3%	45	41.7%	43	39.8%
SCI Total Citations	***Top 12 (10%) Titles***	6	50.0%	8	66.7%	7 (c)	58.3%
	Bottom 108 (90%) Titles	19	17.6%	45	31.0%	45	41.7%
SCI Impact Factor	***Top 12 (10%) Titles***	4	33.3%	6	50.0%	9 (d)	75.0%
	Bottom 108 (90%) Titles	21	19.4%	47	43.5%	43	39.8%
(a) Of these 11 journals, 10 were US association titles. (b) Of these 9 journals, 8 were US association titles.							
(c) Of these 7 journals, 6 were US association titles. (d) Of these 9 journals, 5 were US association titles.							

Table 13. Comparative Structure of the Top 12 (10%) vs. Bottom 108 (90%) Journals Ranked by LSU Faculty Score, UI Library Use, SCI Total Citations, and SCI Impact Factor in Terms of Number of Source Items and Years of Backfile

Quality Measure	Stratum	Number of Source Items		Years of Backfile	
		Arithmetic Mean	Median	Arithmetic Mean	Median
LSU Faculty Score	***Top 12 (10%) Titles***	1208.8	1208	56.8	58.5
	Bottom 108 (90%) Titles	429.4	272.5	35.8	27
UI Library Use	***Top 12 (10%) Titles***	1863.1	1884.5	65.7	58.5
	Bottom 108 (90%) Titles	356.7	267	34.8	27
SCI Total Citations	***Top 12 (10%) Titles***	1992.4	1944.5	68.9	58.5
	Bottom 108 (90%) Titles	342.3	267	34.4	27
SCI Impact Factor	***Top 12 (10%) Titles***	738.0	96.5	38.8	26
	Bottom 108 (90%) Titles	481.7	309.5	37.8	30

[Above : TABLES 11-13]

Table 12 compares the compositions of the top 10% and bottom 90% of the titles in terms of the categorical variables under consideration in this paper. In respect to biochemistry titles, the findings corroborate previous findings in this paper. The percentage of biochemistry titles is higher in the top 10% of the titles than in the bottom 90% on both citation measures; the opposite is the case for LSU faculty score and UI library use. However, in respect to nationality, some interesting differences begin to emerge between the impact factor and the other measures of journal importance. For LSU faculty score, UI library use, and *SCI* total citations, the percentage of US titles is considerably higher in the top 10% of the journals than in the bottom 90%, but the percentages of US titles in the top 10% and bottom 90% approximate each other in respect to the impact factor. The full implications of this difference emerge in the comparison of the two subsets in terms of association titles. For all four measures, the percentage of association titles is considerably higher in the top 10% than in the bottom 90%. However, for LSU faculty score, UI library use, and *SCI* citations, all but one of the association titles in the top 10% are US journals, whereas with the impact factor the association journals in the top 10% are almost evenly split between 5 US association titles and 4 foreign association titles. In light of other findings concerning foreign association titles made in this paper, it thus seems that the impact factor is doing what Garfield designed it to do—pick out significant titles that would be otherwise overwhelmed by bibliometric size.

The differences of the impact factor from the other measures of journal importance become more pronounced, when the top 10% of the journals are compared against the bottom 90% on the physical and temporal facets of journal size. Table 13 above summarizes the results of this comparison by giving the arithmetic means and medians of the top and bottom strata in terms of number of source items and years of backfile. For LSU faculty score, UI library use, and *SCI* total citations, the top 10% is considerably larger in terms of number of source items and considerably older in years of backfiles on both measures of central tendency. The situation is much more complex with the impact factor. Here it is seen that the 10% of the journals highest on the impact factor have a higher mean number but a much lower median number of source items than the bottom 90%. The reasons for this apparent anomaly becomes clear upon investigation of the nature of the titles comprising the top 10% on the four measures of journal importance. Whereas the titles of the top 10% in terms of LSU faculty score, UI library use, and *SCI* total citations were for the most part large research journals like the *Journal of the American Chemical Society* (*JACS*), the top 10% of the journals in terms of the impact factor were divided into one-third large research journals like *JACS* and two-thirds small review journals like *Chemical Reviews*. In respect to years of backfile, Table 13 reveals that the top 10% of the journals on impact factor also differed from the top 10% on the other measures in that both measures of central tendency were approximately equal to the bottom 90% instead of being considerably higher. Moreover, the table also shows that, on both measures of central tendency, the top 10% of the journals highest in impact factor had considerably less years of backfile than the top 10% of the journals on the other three measures of journal importance, suggesting that the impact factor was correcting for age and capturing the importance of the newer titles.

Table 14 below compares the top 10% of the journals on all four measures of journal importance against each other in terms of the percentage of their articles being review articles. The table gives both the arithmetic mean and median of this percentage for all four measures. Inspection of the two measures of central tendency reveals some interesting characteristics of the

Table 14. Arithmetic Mean and Median % Review Articles in Top 12 (10%) Journals Ranked by LSU Faculty Score, UI Library Use, SCI Total Citations, and SCI Impact Factor		
Quality Measure	***Arithmetic Mean***	***Median***
LSU Faculty Score	17.9%	1.2%
UI Library Use	1.1%	0.9%
SCI Total Citations	0.9%	0.6%
SCI Impact Factor	56.5%	68.1%

[Above : TABLE 14]

compositions of the 10% of the titles highest on the four measures of journal importance. First, in respect to LSU faculty score, the higher mean than median is the result of two small review journals, *Accounts of Chemical Research* and *Chemical Reviews*, whose appearance here, as shown in Table 13, also made the top 10% on faculty score have a smaller mean and median in terms of number of source items than those of the top 10% of the titles by UI library use and *SCI* total citations. It was shown in the first section of this paper (pp. 5-6 above) that Garfield developed the citation index off the concept of the review article and highly valued this form of publication. The appearance of these two review journals among the top 10% of the titles by faculty score indicates that the LSU chemistry professors concurred with Garfield on this matter. Moreover, both *Accounts of Chemical Research* and *Chemical Reviews* are published by the American Chemical Society, showing the influence of the sociological factors in their appearance among the top 10% of the titles by faculty score. Second, the two measures of central tendency are very small and approximately equal for both UI library use and *SCI* total citations. This is a result of the top 10% of the titles on these two measures of journal importance being large research journals containing only a few review articles. And, finally, in respect to *SCI* impact factor, the lower mean than median in terms percentage of review articles results from the mean being reduced by one-third of the titles being large research journals with few review articles and two-thirds being small review journals. However, both measures of central tendency of the percentage of review articles is much greater for the top 10% of the titles highest on the impact factor than for the top 10% highest on the other three measures of journal importance. From this it is possible to conclude that a small part of the variance in faculty score is due to the importance assigned by the LSU chemistry professors to review literature and that a large part of the variance in the impact factor is due to the propensity of review articles to be more highly cited than the other forms of scientific literature.

Summary and Conclusions

The analysis of this section has once again corroborated the findings made by Garfield in his analyses of total citations. These findings caused him to formulate his Law of Concentration, which posits that the scientific journal system is dominated by a small interdisciplinary core of large research journals. In terms of *SCI* total citations, the sample under analysis was dominated by 12 large research journals, which comprised 10% of the titles but accounted for 56.3% of the total citations. These journals were both physically large in terms of number of source items and temporally large in terms of years of backfile. Large research journals both in the physical and temporal sense also dominated University of Illinois (UI) Chemistry Library use and, to a lesser extent, the subjective ratings of journal importance by the Louisiana State University (LSU) chemistry faculty. Among the top 10% of the titles by LSU faculty score were not only two small review journals, *Accounts of Chemical Research* and *Chemical Reviews*, but also the instructional title, *Journal of Chemical Education*, which is important in the teaching of chemistry and, like the two review journals, published by the American Chemical Society. The latter journal ranked rather high on UI library use (21^{st}) but extremely low on both *SCI* total citations (77^{th}) and *SCI* impact factor (118^{th}), indicating the limitations of citations in measuring journal importance.

Garfield's concept of scientific journals as sociological entities was also validated by the research discussed in this section. The proof was provided by the analysis of the journal subsets defined by nationality and association affiliation together with the analysis of the relationship of citations to peer ratings in the evaluation of US research-doctorate programs. Here it was seen that journals affiliated with US scientific associations had a greater probability of being highly rated by the LSU chemistry faculty, used at the UI Chemistry Library, as well as scoring higher on both *SCI* total citations and impact factor. This was particularly true of the journals published by the American Chemical Society. Moreover, a large proportion of the 10% of the journals scoring highest on all four measures of journal importance were US association journals. At the same time it was seen that there was an extraordinarily high correlation of citations to the peer ratings of US research-doctorate programs in chemistry and that bibliometric size played an important role in these peer ratings. Thus, the correlation of 0.81 between the peer ratings and mean citations per faculty rose to 0.91, when the latter measure was converted to total program citations. Bibliometric size also played a role in the LSU chemistry faculty's journal ratings, which had a correlation of 0.58 with number of source items and 0.27 with years of backfile. Moreover, it was proven that citations concentrated on the eleven elite programs, which were consistently among the top 15 chemistry programs by peer ratings from 1924 through 1993. From these findings a number of conclusions can be drawn. First, there is a sociological interlock between the faculty of the elite US chemistry programs and US association journals. Second, the correlation of peer ratings with citations is so high that the two appear to be equivalent measures of the same phenomenon. The extraordinarily high correlation of peer ratings with citations and the concentration of citations on the eleven traditionally elite chemistry programs by peer ratings lead to the conclusion that citations have all the historical, social, institutional, national, and other biases of peer ratings. This conclusion combined with the concentration of citations on the eleven traditionally elite chemistry programs leads to the final third and final conclusion. In his seminal paper on the Matthew Effect, Merton (1968) discussed cumulative advantage from two perspectives: 1) the misallocation of scientific credit due to prior achievements; and 2) the stochastic process underlying the formation of the highly stratified social system of US academic science. The above findings indicate that both factors are operative in the data under analysis, and citations must therefore be considered not as an objective measure of scientific significance but as a sociometric measure influenced by factors unrelated to scientific quality. Of great interest was the evidence that the link between the scientific journal system and scientists is more sociological than national. This can be deduced from the fact that foreign association journals appear to play the same role as US association journals and performed better in general on all four measures of journal importance than did the US non-association journals.

The analysis in this section clearly demonstrates that the impact factor accomplishes the objectives for which Garfield created it. Garfield designed the impact factor to counteract the effects of size in both its temporal and physical facets that so affect total citations. He did this by restricting the backfile to the two most recent years and then estimating the mean citation rate to the articles published in those two years. His objectives in doing this were the following: 1) to select which journals should be indexed by the *SCI* from those outside the small interdisciplinary core posited by his Law of Concentration to be dominating the scientific journal system; and 2) to measure the current significance of journals as against their historical significance. Analysis of the data shows that he succeeded in these objectives. Thus, the correlations of both the number of source items and years of backfile with the impact factor were both low and insignificant, showing that the effects of size in both its physical and temporal aspects had been neutralized. Moreover, the mix of journals comprising the 12 journals or 10% of the sample titles highest on the impact factor was extremely heterogeneous and differed markedly from the mix of this same stratum highest on the other measures of journal importance. First, in terms of size, the 12 top journals on the impact factor had fewer numbers of source items and years of backfile than the 12 top journals on the other measures of journal importance. Furthermore, unlike these other measures, the top 12 journals on the impact factor more closely approximated the bottom 108 titles on the impact factor on both facets of journal size. Second, of the 12 top impact factor titles, 8 can be classified as small review journals—a result consistent with those of Garfield in his rankings of journals by the impact factor. Thus, the impact factor captures the importance of the review article. This importance was not caught by either UI library use or *SCI* total citations, whose top titles were all large research journals, but it was captured by LSU faculty score, whose top 12 titles included two small review journals, thus confirming Garfield's opinion on review literature's significant role in science. However, 4 of the top 12 impact factor titles were large research journals, of which 2 were chemistry journals and 2 were biochemistry titles. Of these 4 large research journals, 3 were also among the top 12 total citations journals, indicating that they were both historically and currently significant. These titles included both the *Journal of the American Chemical Society* and the *Journal of Biological Chemistry*. Finally—and most interestingly—the 12 top impact factor journals had a much different composition of association journals than the 12 top titles on the other measures. A large proportion of the 12 top titles on each of the four measures of journal importance were association journals, but, whereas with the other measures, all the association journals except one were US association titles, the association journals of the 12 top impact factor titles were almost evenly split between 5 US association journals and 4 foreign association journals. One of these foreign association journals was *Angewandte Chemie* (International English Edition), which is published under the auspices of the Gesellschaft Deutscher Chemiker. This was the fourth large research journal. *Angewandte Chemie* was not among the top 12 total citations journals, but it did rank 17th on this measure. Braun and Glänzel (1995) proved that the citation counts to this journal were inflated due to the double counting of citations by German-speaking authors to both the German and English editions of this journal. Nevertheless, the importance of *Angewandte Chemie* was confirmed by its being the only foreign association journal among the top 12 titles by LSU faculty score and its ranking 13th on UI library use.

In the first *SCI JCR* the impact factor was given the following definition: "A measure of the frequency with which the 'average cited article' in a journal has been cited in a particular year" (Garfield, 1976, p. 6). The most recent *SCI JCR* defines it thus: "The journal impact factor is the average number of times articles from the journal published in the past two years have been cited in the *JCR* year" (Institute for Scientific Information, 2005). These definitions are misleading. The first one is even technically dubious, because the impact factor is not an estimate of the "average cited article" but of the "average citation rate" to the articles. Moreover, both definitions neglect the bases for the validity and importance of the arithmetic mean in statistics. The arithmetic mean assumed its theoretical importance in the 18th century as the best point estimate in a law of error developed for astronomical and geodetic measurements. Therefore, it is an accurate representation of the population only on the condition that the variance of the observations surrounding it is solely due to random error. This is rarely the case in bibliometric phenomena, where powerful causes are operative. Seglen (1997) made this fact one of the central issues in his rejection of the impact factor as a valid measure for evaluating scientific research. Thus, he wrote: "For the journal's impact factor to be reasonably representative of its articles, the citation rate of individual articles in the journal should show a narrow distribution, preferably a Gaussian [i.e., normal—SJB] distribution, around the mean value (the journal's impact factor)" (p. 497). Seglen then presented a graph showing the 1986 or 1987 citation rates to 1983 or 1984 articles published in three biochemical journals included in the sample under analysis here. One of these journals was the *Journal of Biological Chemistry*. The frequency distributions of the articles manifested all the characteristics of compound Poisson distributions, being highly and positively skewed with the bulk of the articles concentrated in the lowest citation ranges and a long tail to the right. Seglen noted that 15% of the articles accounted for 50% of the citations and that only a few articles were anywhere near the population mean. In an earlier paper Seglen (1992) spelled out implications of these findings thus:

> Citational heterogeneity is thus a fundamental irreducible property of the articles in a journal (as well as of other units of science…). Very few articles will actually have a citedness close to the journal mean, thus the journal impact factor cannot be used as a representative indicator for individual journal articles. The overall journal impact can be heavily determined by a few very highly cited articles…. (p. 145)

However, in this criticism, Seglen is ignoring one crucial point. With such distributions, the arithmetic mean is far above the bulk of the observations closely packed around the mode located in the lowest part of the range due to the few highly cited articles causing most of the variance. Scientific significance tends to be concentrated in these few highly cited articles and is virtually absent from the papers packed around the mode. Therefore, by estimating the arithmetic mean instead of the other measures of central tendency yielding results more representative of the bulk of the articles, the impact factor is better able to identify those journals with a propensity to publish articles that are highly cited.
Such articles are articles reporting important research results or review articles performing an important function.

The full ramifications of Seglen's considerations only become clear when taken together with perhaps the most important discovery made by Garfield (pp. 8-11 above) by analyses utilizing the impact factor as his measure, i.e., the low citation rate of articles published in most journals, including journals that seem almost universally accepted as preeminent. This low citation rate is the result of their low probability of being cited. S. Cole (2000) came to same conclusion as Seglen on the relationship of journal importance to article importance, explaining the reason for this in the following manner:

> Journals, acting as organizational authorities, are part of the evaluation system. The large majority of important work ends up being published in only a few high-prestige and high-visibility journals…. The effectiveness of journal evaluation is, however, limited by the inherent difficulty of judging new contributions at the research frontier. The concentration of high-quality articles is in part a result of the ability of referees to make quality distinctions and probably to a somewhat greater extent the result of the self-selection of authors. Because of the difficulty

> that both referees and authors have in predicting quality, the leading journals publish large numbers of articles of relatively low quality. This limits the effectiveness of the authority of journals. Since most articles, even in the most prestigious journals, turn out to be of minor significance, readers cannot use the prestige of the journal to judge the quality of an individual article. (p. 132)

Even Joshua Lederberg (1977, p. xiv), winner of the Nobel Prize in 1958, stated that his highly cited work on replica plating was not the item he would have chosen from his own bibliography for note-worthy impact. Thus, scientific research must be considered a crapshoot with the odds stacked against you, and large numbers of insignificant papers are a necessary cost of doing science.

5. RELATIONSHIP AMONG THE MEASURES OF JOURNAL IMPORTANCE

Graphic Analysis

The analysis of the relationship of the measures of journal importance was begun by plotting the measures against each other. These plots are shown in Figures 5-10 below. Two things were added to the plots to assist in the analysis. First, the positions of the three journals scoring highest on the four measures of journal importance are shown on the plots. These journals were the *Journal of the American Chemical Society* (*JACS*), a large chemistry research journal, which was highest on both LSU faculty score and UI library use; the *Journal of Biological Chemistry*, a large biochemistry research journal, which was highest on *SCI* total citations; and *Chemical Reviews*, a small chemistry review journal, which was highest on *SCI* impact factor. Second, to illustrate the perspective of the LSU chemistry faculty and the UI Chemistry Library patrons, hypothetical regression lines were drawn from the origins of the plots to *JACS*. The logic behind these regression lines was that the more the observations conformed to these regression lines, the more they conformed to the perspective of the LSU chemistry faculty and UI Chemistry Library patrons.

Inspection of the plots makes a number of things immediately clear. First, the plots of LSU faculty score, UI library use, and *SCI* total citations against each other as well as their plots against *SCI* impact factor manifest a great deal of similarity. In the former case the observations conform much more closely to the hypothetical regression line than do the observations in the latter case. This demonstrates that there is a greater commonality among LSU faculty score, UI library use, and *SCI* total citations, than there is a commonality of all these measures with *SCI* impact factor. Second, there is much greater scatter of the observations on all the impact factor plots than on the other plots. This greater scatter is due not only to the impact factor capturing simultaneously the importance of both the research and review functions but also is graphic evidence of what both the Poisson and binomial tests had proven—that the impact factor has a larger proportion of random error in its variance than do the other measures. It thus serves as a validation of Garfield's finding of the ability of the impact factor to identify review journals as well as his opinion about the measure's lack of precision. Third, the position of the *Journal of Biological Chemistry* as an extreme outlier on all the plots of the citation measures is a graphic representation of the effects of Bradford's Law of Scattering and Garfield's Law of Concentration. On the plots of LSU faculty score and UI library use against the citation measures, this journal is much too high on the citation measure relative to its position on the

Figure 5. Plot of LSU Faculty Score against University of Illinios (UI) Chemistry Library Use

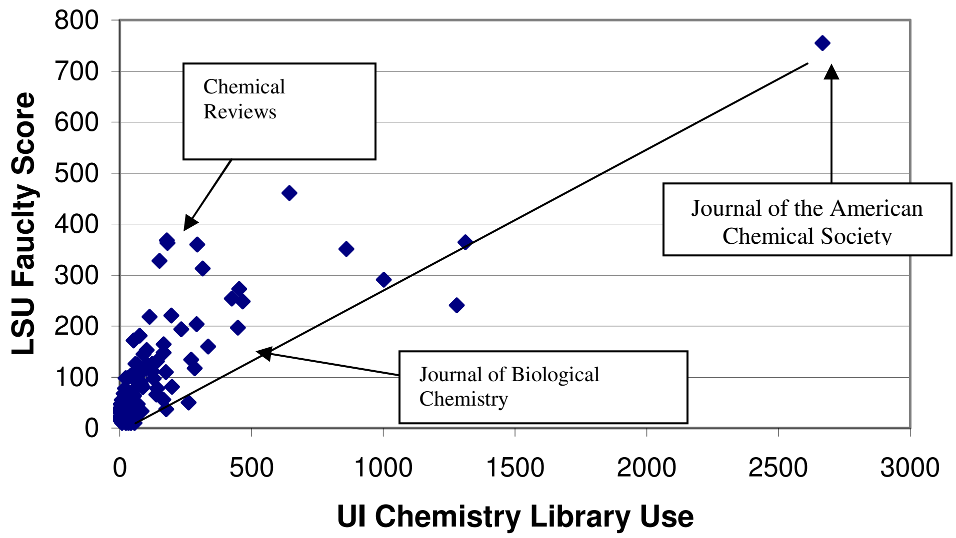


Figure 6. Plot of LSU Faculty Score against Science Citation Index (SCI) Total Citations

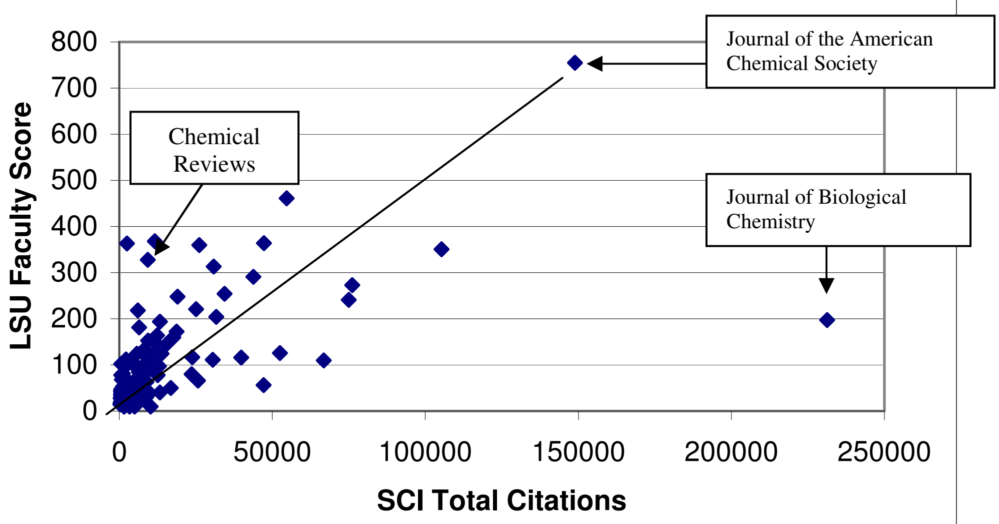


Figure 7. Plot of LSU Faculty Score again Science Citation Index (SCI) Impact Factor

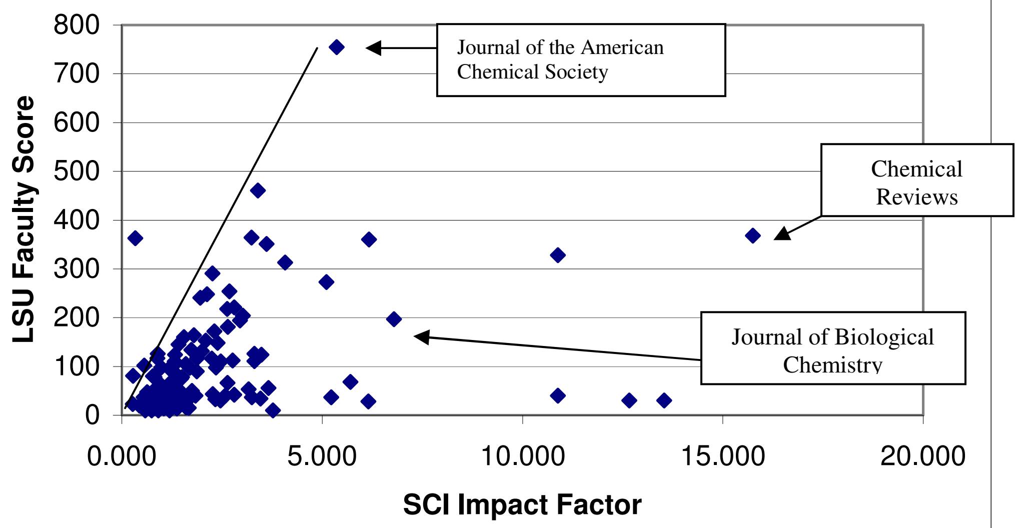


Figure 8. Plot of University of Illinois (UI) Chemistry Library Use against Science Citation Index (SCI) Total Citations

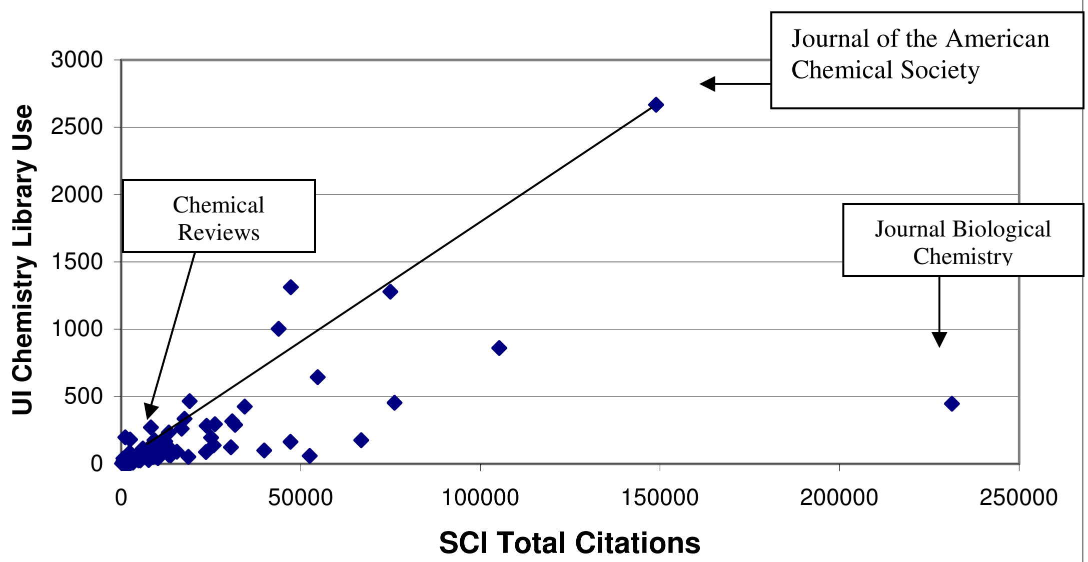


Figure 9. Plot of University of Illinois (UI) Chemistry Library Use against Science Citation Index (SCI) Impact Factor

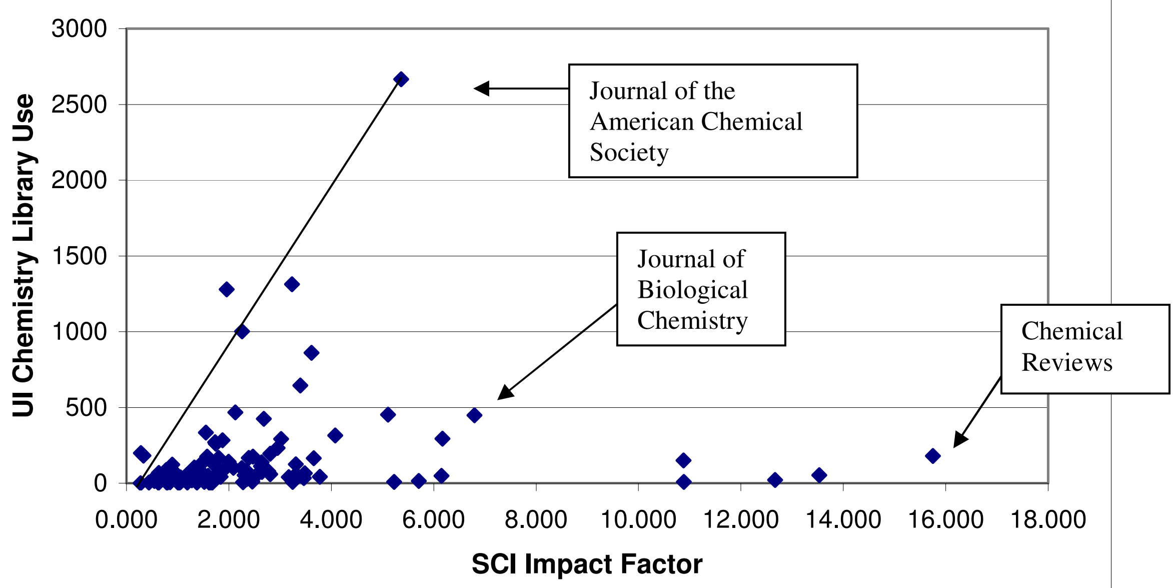
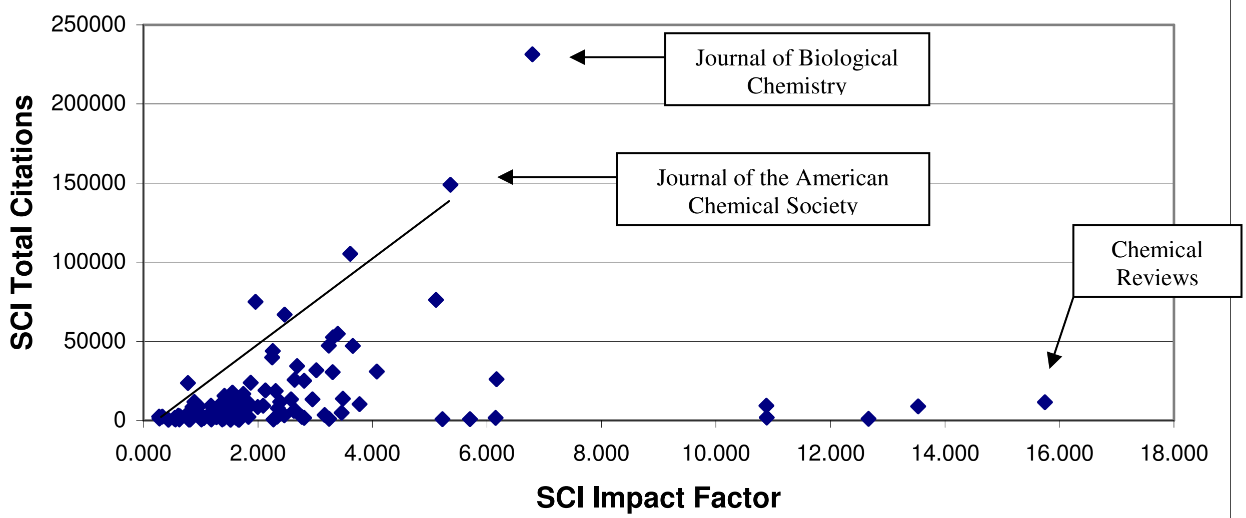


Figure 10. Plot of Science Citation Index (SCI) Total Citations against Science Citation Index (SCI) Impact Factor



[Above : FIGURES 5-10]

other measure. The plot of *SCI* total citations against *SCI* impact factor reveals that the *Journal of Biological Chemistry* is so far above the guidepost *JACS* on the total citations as to seem to belong to another system. All these serve as graphic representations of the *Journal of Biological Chemistry* receiving citations exogenous to the subject interests of the LSU chemistry faculty and the UI Chemistry Library patrons. And, finally, the position of *Chemical Reviews* on the plots is most revealing. This is particularly true in respect to the plots involving LSU faculty score. On the plots of LSU faculty score against UI library use and *SCI* total citations, this small review journal is extremely high on LSU faculty score relative to its position on the other two measures. Since the amount of UI library use and *SCI* total citations were much more a function of size than was LSU faculty score, this is graphic proof that the LSU chemistry faculty valued this journal much more than warranted by its size and were in agreement with Garfield on the important function of review articles in scientific literature. However, on the plot of LSU faculty score against *SCI* impact factor, *Chemical Reviews* is an extreme outlier with an impact factor far too high relative to its faculty score. This is also its position on the plots of UI library use and *SCI* total citations against *SCI* impact factor. These plots serve as graphic evidence that the impact factor is not a holistic measure of journal importance but is capturing a facet of journal importance that should perhaps be considered separately from the other facets of journal importance.

The Pearson Product-Moment Correlation Coefficient r

The Pearson product-moment correlation coefficient, whose standard symbol is *r*, is a measure of the closeness of a linear relationship between two variables. Its calculation may be divided into three parts: 1) the determination of the form of the relationship—the regression line; 2) the measure of the variation of the observations about the established form of the relationship; and 3) the reduction of the measurement of association to a relative basis—the correlation coefficient, which ranges from -1 to 1. The slope of the regression line, which can be positive or negative, indicates the nature of the relationship, and the more closely the correlation coefficient approaches -1 or 1, the more closely the observations fit the regression line. Pearson's *r* is a parametric statistical technique in that correct inferences from it require both variables to be normally distributed. With the type distributions, with which we are dealing, this is done by converting them to lognormal distributions through the logarithmic transformation (Elliott, 1977, pp. 33 and 102). This was accomplished here with the natural logarithm. The logarithmic transformation altered the distributional characteristics of the measures, eliminating the extreme concentration of titles in the lower deciles. However, it did not eliminate errors resulting from exogenous citations or the inability to determine precisely a "citable" item. As will be seen, such errors resulted in outliers—or observations deviating much further than the others from the regression line—which, if influential, can alter the slope and position of the regression line, thereby affecting the correlation coefficient.

An important method of determining of the nature of the relationship between two variables is to analyze the outliers. For this purpose, the logarithmic measures were plotted against each other; the regression line through the observations was plotted; and then the plots were visually inspected to determine which observations represented extreme outliers. The correlation coefficients were calculated first with the outliers included and then with the outliers

Table 15. Pearson r Correlation Matrix for LSU Faculty Score, University of Illinois (UI) Chemistry Library Use, Science Citation Index (SCI) Total Citations, and Science Citation Index (SCI) Impact Factor

	Outliers	UI Chemistry Library Use	SCI Total Citations	SCI Impact Factor
LSU Faculty Score	Included	0.73	0.67	0.37
	Excluded	0.77	0.74	0.56
UI Chemistry Library Use	Included		0.82	0.36
	Excluded		0.89	0.42
SCI Total Citations	Included			0.43
	Excluded			0.63

All correlations significant at the 0.01 level.

[Above : TABLE 15]

excluded to gauge how much the outliers were affecting the relationship between the two measures.

Table 15 above summarizes the results of the analysis. Here it can be seen that LSU faculty score, UI library use, and *SCI* total citations are more linearly related to each other than any of them are linearly related to *SCI* impact factor. With the outliers included, the Pearson *r* of LSU faculty score is 0.73 with UI library use, 0.67 with *SCI* total citations, but only 0.37 with *SCI* impact factor; the Pearson *r* of UI library use with *SCI* total citations is 0.82 but only 0.36 with *SCI* impact factor; and the Pearson *r* of *SCI* total citations with SCI impact factor is only 0.43—much less than its correlations with LSU faculty score and UI library use. The main reason for this phenomenon is that, to a greater or lesser extent, LSU faculty score, UI library use, and *SCI* total citations are functions of what may be defined as "bibliometric size," which the impact factor was specifically designed to control. Two other conclusions may be drawn from these correlations. First, the high correlation of LSU faculty score with *SCI* total citations corroborates Garfield's opinion that total citations identifies those journals which scientists intuitively regard as the important journals of science. Second, given the high correlation of citations with the peer ratings of the scholarly quality of US research-doctorate programs in chemistry, the high inter-correlations among LSU faculty score, UI library use, and *SCI* total citations are proof that the LSU chemistry faculty and the patrons of the UI Chemistry Library were functioning as part of the same social stratification system and therefore regarded the same journals as important.

Inspection of the logarithmic plot of LSU faculty score against UI library use revealed five journals as outliers in the sense having a faculty score too low in respect to their library use. Their titles were: *Bulletin de la Société chimique de France*, *International Journal of Chemical Kinetics*, *Journal of Lipid Research*, *Monatshefte fűr Chemie*, and *Zeitschrift fűr anorganische und allgemeine Chemie*. These journals had the lowest possible LSU faculty score and ranked near the bottom in terms of UI library use. An examination of their characteristics revealed the reasons for this. Two titles—*Bulletin de la Société chimique de France* and *Monatshefte fűr Chemie*—were foreign association titles but published in a foreign language. In addition, the former journal classed in biochemistry. One title—*Journal of Lipid Research*—was a US association journal, but it, too, classed in biochemistry. The last two titles—*International Journal of Chemical Kinetics* and *Zeitschrift fűr anorganische und allgemeine Chemie*—were both non-association journals. Of the two, the first was a US title, whereas the second was foreign and suffered from the additional handicap of being published in a foreign language. Exclusion of these titles outliers raised the correlation of LSU faculty score with UI library use from 0.73 to 0.77. These correlations can be considered very high, showing the ratings of the small, nonrandom sample of 25 LSU chemists not only as predictive of library use but as predictive of library use at another institution. This serves as an indication of a powerful universality of interests and norms among scientists of a given discipline.

The same five journals also appeared as the same type of outlier on the logarithmic plot of LSU faculty score against *SCI* total citations. Here, again, their LSU faculty score was too low relative to their total citations, and these titles were located in the lowest decile of the total citations range. The *Journal of Biological Chemistry* appeared as an outlier with more total citations than warranted by its faculty score. As an outlier, it differed from the previous five titles by being at the very apex of the total citations ranking. It is a US association title classed in biochemistry. One other journal appeared as an outlier but of an entirely different type in that it had a faculty score far too high for its total citations. It was the *Journal of Chemical Education*, an instructional title published by the American Chemical Society. This title ranked in the top 10% of the journals in terms of faculty score, and it is significant that it did not appear as an outlier in the logarithmic plot of LSU faculty score against UI library use, thereby indicating the limitation of citations in measuring journal importance. Removal of the above seven journals as outliers increased the initial correlation of LSU faculty score with *SCI* total citations from 0.67 to 0.74. This is a correlation high enough to indicate that the judgments of the 25 LSU chemists were in synch with those of thousands of publishing scientists worldwide.

Visual inspection of the logarithmic plot of LSU faculty score against *SCI* impact factor revealed two groups of outliers. One group was located below the regression line, indicating a faculty scores too low in relationship to their impact factors. This group can be divided into two subgroups: those having low impact factors; and those having high impact factors. The titles of the first subgroup were *Bulletin de la Société chimique de France*, *International Journal of Chemical Kinetics*, *Monatshefte fűr Chemie*, and *Zeitschrift fűr anorganische und allgemeine Chemie*. These titles had appeared as outliers in the logarithmic plots of LSU faculty score against both UI library use and *SCI* total citations, where they also were characterized as being extremely low on both measures. The subgroup with high impact factors consisted of the following journals: *Chemical Society Reviews*, *Critical Reviews in Biochemistry and Molecular Biology*, *Journal of Lipid Research*, *Surface Science Reports*, and *Trends in Biochemical Sciences*. Of these titles, the *Journal of Lipid Research*, a US association journal in biochemistry, is of special interest, because it had appeared as an outlier with the titles of the first subgroup on the logarithmic plots of faculty score against library use and total citations, where it, too, had also been low on both measures. The reason for its transition from being low on total citations to being high on the impact factor can be found in its size. Its number of 1993 source items was 206, which was below the entire set's median number of 1993 source items of 290, and, from this, it is possible to see that the impact factor had captured a facet of importance missed by faculty score, library use, and total citations due to the effect of bibliometric size upon them. The *Journal of Lipid Research* can be categorized as a small, narrowly focused research journal in that only 9 of its 206 source items were review articles, but the other outliers high on the impact factor can be categorized as small review journals. Their other characteristics can be described as follows: *Chemical Society Reviews*—foreign association journal in chemistry; *Critical Reviews in Biochemistry and Molecular Biology*—US non-association journal in biochemistry; *Surface Science Reports*—foreign non-association journal in chemistry; and *Trends in Biochemical Sciences*—foreign association journal in biochemistry.

The other group of outliers visible on the logarithmic plot of LSU faculty score against *SCI* impact factor was above the regression line, indicating that they were more valued by the LSU chemists than their impact factors warranted. This group consisted of three titles, all of which were published by the American Chemical Society. Two of these titles were low on the impact factor, and they were the informational *Chemical & Engineering News* and the instructional *Journal of Chemical Education*. The latter title also had a faculty score too high in respect to its total citations. Their appearance as outliers is further evidence of the limitations of citations in measuring the importance of journals. The third title was high on the impact factor, and it was the *Journal of the American Chemical Society* (*JACS*). *JACS* was the journal highest on LSU faculty score and UI library use, and *SCI* total citations consistently identified it as a component title of the interdisciplinary core posited by Garfield's Law of Concentration as dominating the entire scientific journal system. The underestimation of its importance by the impact factor is evidence of problems with this measure. Exclusion of the outliers raised the correlation of LSU faculty score with *SCI* impact factor from 0.37 to 0.56, but the latter correlation was still considerably below the correlations of LSU faculty score with UI Chemistry library use and *SCI* total citations even with the outliers included, indicating that the latter citation measure is a better surrogate for the opinion of the LSU chemists than the impact factor.

Analysis of the logarithmic plots of UI Chemistry Library use against *SCI* total citations and *SCI* impact factor revealed that this measure interacted with the two citation measures in much the same fashion as LSU faculty score. This is not surprising, given the relatively high correlations—0.73 with outliers, 0.77 without outliers—between LSU faculty score and UI library use. Visual inspection of the logarithmic plot of UI library use against *SCI* total citations once again uncovered two groups of outliers—one below the regression line and one above the regression line. The position of the group below the regression line indicated that its titles had low library use relative to their total citations. This group could also be subdivided into a subgroup low on both measures and a subgroup low in library use but high in total citations. The first subgroup consisted of the following three titles: *Nucleosides and Nucleotides*, *Optics and Spectroscopy*, and *Zeolites*. Two of these journals—*Nucleosides and Nucleotides* as well as *Zeolites*—were US non-association titles, whereas *Optics and Spectroscopy* was a US association journal published by the American Institute of Physics for the Optical Society of America. However, it was not a US association journal in the usual sense but a translation of the Russian title *Optika i spektroskopiya*. The absence of true association status on the part of these titles was probably causal in their low library use. ISI classed *Nucleosides and Nucleotides* in biochemistry and *Optics and Spectroscopy* in optics and spectroscopy, so that both these titles may be considered as only partially within the chemistry set—another reason for their low library use. *Zeolites* classed in physical chemistry, a true subset of chemistry, but its narrow focus on molecular sieves was probably a reason for its low library use. The other subgroup below the regression line may be defined as having fairly high library use and extremely high total citations. It consists of the following two titles: *Biochemical and Biophysical Research Communications* and the *Journal of Biological Chemistry*. The first was a US non-association title, whereas the second was a US association journal. Both titles classed in biochemistry, and this may be considered the reason for their library use being low in relationship to their total citations.

The group of titles above the regression line on the logarithmic plot of UI library use against *SCI* total citations was much too high on the former measure in relationship to the latter measure. It consisted of three titles published by the American Chemical Society. These titles do not serve a research purpose and were extremely low in total citations. They were the informational *Chemical & Engineering News,* the applied journal *Chemtech*, and the instructional *Journal of Chemical Education*. Two of these journals—*Chemical & Engineering News* and *Journal of Chemical Education*—had appeared as the same type of outliers on the logarithmic plot of LSU faculty score against *SCI* impact factor. This group of outliers is once again proof that citations are basically a research measure and poor gauges of reading done for other purposes.

Exclusion of the outliers raised the correlation of UI library use with *SCI* total citations from 0.82 to 0.89. These are extraordinarily high correlations, and they demonstrate that the *SCI* total citations measure captures many of the variables that determine why scientists choose to read certain journals over others.

The same cannot be said for *SCI* impact factor, whose initial correlation with UI library use was merely 0.36. Two reasons can be adduced for this dichotomy. First, UI library use and *SCI* total citations were highly size-dependent, and the impact factor specifically controls for size by a method that is virtually impossible to apply to library use. Second, the impact factor has a high proportion of random error due to its method of calculation. Visual inspection of the logarithmic plot of UI library use against *SCI* impact factor again revealed one group of outliers below the regression line and another group of outliers above the regression line. The position below the regression line indicates library use too low in respect to the impact factor, whereas the position above the regression line suggests a library use too high for the impact factor. Each of these groups of outliers could in turn be divided into one subgroup low on the impact factor and a second subgroup high on the impact factor.

Turning to the two subgroups beneath the regression line, the subgroup low on both library use and impact factor consisted of the following six titles: *Applied Spectroscopy Reviews*, *Journal of Coordination Chemistry*, *Microchemical Journal*, *Nucleosides and Nucleotides*, *Optics and Spectroscopy*, and *Zeolites*. A number of reasons can be deduced for either their low library use or low citation rates. First, most did not rank high in the social system of chemistry. Four were non-association titles, of which 3 were US publications, and one was foreign. Two were US association journals, but, of these, one was a Russian translation journal published by a US association. Second, all had a relatively narrow subject focus, making them of major interest to only small subsets of scientists. Moreover, three were not fully in the chemistry set, one being a biochemistry journal and two being spectroscopy titles. The subgroup beneath the regression line low on library use but high on the impact factor comprised two titles: *Critical Reviews in Biochemistry and Molecular Biology* and *Natural Products Reports.* Both were small review journals that ranked among the 12 titles highest on the impact factor. *Critical Reviews in Biochemistry and Molecular Biology* had 14 source items in 1993, of which 13 were review articles, and *Natural Products Reports* had 29 source items in 1993, of which all were review articles. In addition to being a review journal, the citation rate of *Critical Reviews in Biochemistry and Molecular Biology* was helped by its being a biochemistry title, whereas that of *Natural Products Reports* was probably raised by its issuance by the Royal Society of Chemistry.

Concerning the two subgroups above the regression line on the logarithmic plot of UI library use against *SCI* impact factor, the subgroup categorized by high library use and low impact factor consisted of the American Chemical Society's informational *Chemical & Engineering News* and its instructional *Journal of Chemical Education*. These journals were in the same position on the logarithmic plot of library use against total citations and for the same reason—the inability of citation measures to capture the full importance of journals not primarily dedicated to research. The outlier subgroup high on both library use and the impact factor comprise the following four titles: *Journal of Organic Chemistry*, *Journal of the American Chemical Society, Journal of the Chemical Society,* and *Tetrahedron Letters*. It should be noted that all four of these titles were among the twelve titles highest in total citations, and, given the high correlation of UI library use with *SCI* total citations, their high library use was a function of the same variables causing their high total citations—a combination of bibliometric size, better quality, and high social status. Two of these titles were American Chemical Society publications; one was issued by the Royal Society of Chemistry; and one was a foreign non-association title, demonstrating once again the dominance of associations in scientific publishing. None of these titles classed in biochemistry.

Exclusion of all these outliers raised the initial correlation of UI library use with *SCI* impact factor from 0.36 to 0.42. These correlations are considerably lower than the equivalent correlations of UI library use with both LSU faculty score and *SCI* total citations, indicating that the impact factor is not a good predictor of library use unlike the other two measures. The main reason for this is that the impact factor deliberately controls for bibliometric size, which is one of the key variables affecting library use. It should be kept in mind that bibliometric size consists of two facets—a physical facet in terms of number of published source items and a temporal facet in terms of years of backfile. The physical facet of size was found to be a measure of social significance in the evaluations of the quality of US research-doctorate programs, whereas size in its temporal facet can be interpreted as a measure of historical significance. Moreover, it is extremely difficult, if not impossible, to control library use for bibliometric size by the same method as the impact factor. Not only are there the logical problems involved in defining "use" and "usable item," but the number of "uses" at a given library are so much lower than the number of citations, with which ISI works, that the result of the calculations would be numbers so low and compacted as to be worthless for comparative purposes. Indeed, this is the case for the vast bulk of even citation impact factors.

The final step in the correlation analysis was to explore the relationship of *SCI* total citations to *SCI* impact factor in order to determine whether this relationship was similar to those of LSU faculty score and UI Chemical Library use to this measure. Visual inspection of the logarithmic plot of *SCI* total citations against *SCI* impact factor disclosed a group of outliers below the regression line and a group of outliers above the regression. An extreme position below the regression line indicated totals citations too low in respect to the impact factor, and an extreme position above the regression line suggested total citations too high in respect to the impact factor. The group below the regression line contained the following nine titles: *Applied Spectroscopy Reviews*, *Bioconjugate Chemistry*, *Chemical Society Reviews*, *Critical Reviews in Analytical Chemistry*, *Critical Reviews in Biochemistry and Molecular Biology*, *Natural Products Reports*, *Progress in Lipid Research*, *Progress in Nuclear Magnetic Resonance Spectroscopy*, and *Surface Science Reports*. Five of these titles were among the twelve titles highest on impact factor. The nine titles represent a complex mix of journals by nationality, association, and subject. Thus, one was a US association journal, two were foreign association titles, three were US non-association titles, and three were foreign non-association titles. In terms of subject, four titles classed in chemistry, three were biochemistry titles, and two journals classed in spectroscopy. Their most common uniting characteristic was that they were small review titles. Thus, eight of the journals had number of 1993 articles ranging from 6 to 48, and the distribution of their percentage of review articles was as follows: one title - 55.0%; one title - 81.25%; one title – 92.9%; and five titles – 100.0%. The outlier title that was not a small review journal, *Bioconjugate Chemistry*, was a small biochemistry title containing only 87 articles. Of the eight review journals, three had appeared as outliers in the logarithmic plot of LSU faculty score against *SCI* impact factor, and three has appeared as outliers in the logarithmic plot of UI Chemistry Library use against *SCI* impact factor.

There were six outliers above the regression line on the logarithmic plot of *SCI* total citations against *SCI* impact factor. Of these outliers, one could be described as high on total citations but low on the impact factor. This title was *Acta Crystallographica*, which was a foreign association title classed in biochemistry. An explanation of its combination of high total citations and low impact factor may be that it was a fairly old title with a backfile dating back to 1948 containing articles that had a slow rate of citation obsolescence. The five other outliers above the regression could be categorized as high on both total citations and the impact factor. These titles are: *Biochimica et Biophysica Acta*, *Journal of Biological Chemistry*, *Journal of Chemical Physics*, *Journal of the American Chemical Society*, and *Journal of the Chemical Society*. All five of these titles were among the twelve highest in *SCI* total citations. The *Journal of the American Chemical Society* had appeared as a similar outlier in the logarithmic plots of both LSU faculty score and UI Chemistry Library use against *SCI* impact factor, whereas the *Journal of the Chemical Society* has appeared as such an outlier in the logarithmic plot of UI Chemistry Library use against *SCI* impact factor. These journals represented the familiar mix of association and biochemistry titles. Three were US association titles; one was the main publication of the Royal Society of Chemistry, and one was a foreign non-association journal. Two journals classed in biochemistry, of which one was the foreign non-association title.

Exclusion of the outliers raised the correlation of *SCI* total citations with *SCI* impact factor from 0.43 to a respectable 0.63. This is a very interesting result, because it shows that, when the set is more carefully defined by restricting it to research journals through exclusion of the review journals, total citations and the impact factor begin to approximate each other as measures of journal importance. It has been seen in the first section above (pp. 12-13) that Garfield achieved this effect by restricting his set to those journals publishing at least 100 articles in a deliberate attempt to exclude the review journals.

Nonparametric Ranking Comparisons: The Spearman Rank Correlation Coefficient rho and the Chi-Squared Test of Independence

Most journal evaluations involve ordinal rank comparisons. These ordinal ranks may be based upon either subjective human judgments such as LSU chemistry faculty ratings or interval/ratio data such as UI Chemistry Library uses or *SCI* citations. Siegel (1956) points out that measures like LSU faculty score are "precisely numerical in appearance only" (p. 3). Ordinal ranking comparisons are the domain of nonparametric or "distribution-free" statistics. Nonparametric statistical techniques differ from parametric ones in that the former do not require truly numerical data but can be applied to observations measured not only on the ordinal or ranking scale but even on the nominal or classificatory scale. Moreover, inferences from nonparametric techniques are not based upon assumptions about the underlying population distributions such as normality or homogeneity in variance. As a result, nonparametric statistical methods are valid for a wide range of different distributions and can be utilized to analyze highly skewed data by converting the measurements from the interval or ratio scale to the nominal or ordinal scale. This section will utilize two nonparametric methods to analyze the relationships among the measures of journal importance: the Spearman rank correlation coefficient rho and the chi-squared test of independence.

Ordinal Structure vs. Ratio Structure

The nonparametric analysis of the relationships among the measures of journal importance was begun by collating the ordinal structure of these measures with their ratio structure. This was done to clarify the ratio distances between the ordinal ranks at various levels of the measures. To do this, the 120 journals were first ranked in ascending order on each measure of journal importance. Then, the following ordinal categories were defined: Low—the lowest 90 or 75.0% of the titles; Medium—the next highest 12 or 15.0% of the titles; and High—the top 12 or 10.0% of the titles. For various investigative purposes, the measures were dichotomized by collapsing the Medium and High categories into a Medium/High category

Table 16. Rank Structure of Journals by LSU Faculty Score, University of Illinois (UI) Chemistry Library Use, Science Citation Index (SCI) Total Citations, and Science Citation Index (SCI) Impact Factor in Terms of Ordinal Categories in Respect to These Measures on the Ratio Scale					
	No Titles	***% Titles***	***Ratio Range***	***Ratio Decile Range***	***% Category Aggregate of Total Aggregate of Measure***
LSU Faculty Score					
Low	90	75.0%	10 - 124	1st - 2nd	38.4%
Medium	18	15.0%	126 - 248	2nd - 4th	25.5%
High	12	10.0%	254 - 755	4th - 10th	36.1%
Medium/High	30	25.0%	126 - 755	2nd -10th	61.6%
UI Chemistry Library Use					
Low	90	75.0%	1 - 138	1st	22.0%
Medium	18	15.0%	141 - 294	1st - 2nd	20.6%
High	12	10.0%	315 - 2667	2nd - 10th	57.4%
Medium/High	30	25.0%	141 - 2667	1st - 10th	78.0%
SCI Total Citations					
Low	90	75.0%	255 - 12509	1st	21.3%
Medium	18	15.0%	12999 - 34414	1st - 2nd	22.4%
High	12	10.0%	39801 - 231324	2nd - 10th	56.3%
Medium/High	30	25.0%	12999 - 231324	1st - 10th	78.7%
SCI Impact Factor					
Low	90	75.0%	0.273 - 2.646	1st - 2nd	43.0%
Medium	18	15.0%	2.648 - 4.075	2nd -3rd	20.5%
High	12	10.0%	5.109 - 15.748	4th - 10th	36.4%
Medium/High	30	25.0%	2.648 – 15.748	2nd - 10th	57.0%

[Above : TABLE 16]

encompassing the top 30 or 25.0% of the titles highest on the measure. The results are shown in Table 16 above, which gives for each category the following information: 1) number and percentage of titles in each category; 2) the ratio range covered the journals in the category; 3) the ratio decile range, within which the titles of the category are located; and 4) the percentage of the total aggregate of the measure for all 120 titles accounted for by the aggregate of the measure for the titles of each category. In general, Table 16 shows that the bottom 90 Low titles on all measures are compressed into an extremely short ratio range and account for very small percentages of the total aggregate of the measure. As one transcends to the Medium and High categories, the ratio ranges and distances between titles exponentially increases as do the percentages of the total aggregate of the measure accounted for by the titles. Closer examination of Table 16 reveals two special characteristics of the measures most affected by size—UI library use and *SCI* total citations—that distinguishes them from the other measures. First, the Low category titles of UI library use and *SCI* total citations are all squeezed into the first decile of the ratio range, whereas the Low category titles of LSU faculty score and SCI impact factor extend from the first into the second decile. Second, the Medium and High categories of UI library use and *SCI* total citations account for a larger percentage of the total aggregate of the measure—respectively, 78.0% and 78.7%, when combined—than do LSU faculty score and *SCI* impact factor—respectively, 61.6% and 57.0%, when combined. Despite these differences, these figures indicate that the titles of the Low category on each measure can all be considered as lacking significance.

One way to uncover the ordinal structure of the four measures of journal importance is to determine the number of titles tied at given ranks at the different ordinal levels. An investigation of the distribution of such tied titles revealed a sharp dichotomy between LSU faculty score and UI library use, on the one hand, and *SCI* total citations and *SCI* impact factor, on the other. Both LSU faculty score and UI library use had numerous tied titles concentrated at the lower frequency levels. For LSU faculty score the distribution of the tied titles over the ordinal categories was as follows: Low—57 tied titles, 63.3% of the category titles, 96.6% of the total tied titles; and Medium—2 tied titles, 11.1% of the category titles, 3.4% of the total tied titles. Overall 59 or 49.2% of the 120 titles were tied ones for LSU faculty score. UI library use had a total of 51 tied titles or 42.5% of the 120 titles, and all these titles were concentrated in the Low category, accounting for 56.7% of the Low category titles. Neither of the two citation measures had any tied titles. For *SCI* total citations this fact can be explained by the extremely long range, running from 255 to 231,324 total citations. However, *SCI* impact factor had an extremely short range of only 0.273 to 15.748, and the lack of tied titles can be considered an artifact of the deliberate ISI policy of calculating the measure to three decimal places precisely to avoid such ties. When Garfield's method of reducing the measure to one decimal place was implemented, 102 or 85.0% of the 120 titles emerged as tied. Unlike LSU faculty score and UI library use, these ties were not concentrated at the lower frequency levels but were more evenly distributed across the ordinal categories in the following manner: Low—86 tied titles, 95.6% of the category titles, 84.3% of the total tied titles; Medium—12 tied titles, 66.7% of the category titles, 11.8% of the total tied titles; and High—4 tied titles, 33.3% of the category titles, 3.9% of the total tied titles.

The Spearman Rank Correlation Coefficient rho

According to Siegel (1956, p. 202), of all the statistics based on ranks, the Spearman rank correlation coefficient—also known as rho—was the earliest to be developed and became the best known. However, its use to compare ranked journal lists based on citations to ranked journal lists based on other measures was subjected to a devastating critique by Brookes (1976). Brookes pointed out that Spearman was an experimental psychologist, who devised his correlation coefficient for comparing ranks derived from human judgment rather than from frequencies. Spearman then took what Brookes (p. 320) described as "the very dubious step of equating the ordinal 1st with the cardinal number 1, the ordinal 2nd with the cardinal number 2 and so on." Brookes stated that the Spearman rho rests upon implicit assumptions about the underlying probability distribution of the entities being ranked that appear never to have been adequately explored. He cited as an example that rho appears to require the probability distribution to have a variance which is proportional to the square of the mean because it assumes that a shift of rank from r to r+1 is equivalent for all ranks of r. Brookes then focused his attention on the lower citation frequencies presented in the ISI *JCR*s. In his opinion, these citation frequencies were too low for calculations based upon ranks. According to Brookes, citation counts are random events, which in this simple case conform to the binomial distribution. He then calculated the size of the sample needed before any credence could be given to the assertion that one journal is cited more than another at the lower frequencies. According to his estimate, the size of the sample would have to be huge even by ISI standards of activity. Brookes (p. 321) then stated, "Such lists as those published by ISI are interesting in themselves, as *samples* of the current scene, but *calculations* on them should be restricted to serials whose ranks are reasonably well-established by the frequencies available." In his opinion, the effect of using ranked samples based on low frequencies is to introduce into the correlation of such samples a scatter which arises solely from sampling variance.

One way to gauge the effect of sampling variance on ordinal ranks is through the confidence interval. Briefly defined, a confidence interval is a range of values within lower and upper limits calculated from the sample observations that are believed, with a particular probability, to contain the true parameter value. For example, a 95% confidence interval implies that if the estimation process were frequently repeated, 95% of the calculated intervals would be expected to contain the true parameter value. We are essentially dealing with Poisson processes, and tables and equations for calculating the confidence limits of an observed Poisson lambda are contained in E.S. Pearson and Hartley (1966, 80-83, 136-137, and 227) as well as Beyer (1968, pp. 238-239). To estimate the potential of sampling variance on the ordinal rankings of the four journal measures under consideration, the 120 journals were first ranked in ascending order on each measure and classified into the above ordinal categories of Low, Medium, and High. Then the journal at the mid rank of each category was selected, and the 95% confidence interval around its observed lambda was calculated to determine how many other journals fell within this confidence interval. The results are summarized in Table 17 below. To understand these results, it is necessary to know that the ratio of the confidence interval to the lambda descends as the lambdas increase. For example, this ratio is 5.5 for an observed lambda of 1, 0.6 for an observed lambda of 40, 0.2 for an observed lambda of 467, and 0.1 for an observed lambda 2,553. With highly skewed distributions such as these, this characteristic creates a fairly large potential for rank churning at the lower frequencies, and the consequences are visible in a comparison of the results for LSU faculty score (range 10-755) and UI library use (range 1-2,667) to those for *SCI* total citations (range 255-231,324). The titles selected for the former two measures have a much

Table 17. Effect of 95% Poisson Confidence Intervals on Rankings of Journals by LSU Faculty Score, University of Illinois (UI) Chemistry Library Use, Science Citation Index (SCI) Total Citations, and Science Citation Index (SCI) Impact Factor at Different Ordinal Levels

LSU Faculty Score								
Ordinal Category	**No. Category Titles**	**Title**	**Faculty Score**	**95% Poisson Confidence Limits**		**95% Confidence Interval**	**No. Titles within Interval**	**Percent Category Titles**
				Lower (a)	**Upper (a)**			
Low	90	Tetrahedron-Asymmetry	42	30	57	27	45	50.0%
Medium	18	Chemische Berichte	164	140	191	51	7	38.9%
High	12	Journal of Chemical Physics	351	315	390	75	6	50.0%
UI Chemistry Library Use								
Ordinal Category	**No. Category Titles**	**Title**	**UIUC Use**	**95% Poisson Confidence Limits**		**95% Confidence Interval**	**No. Titles within Interval**	**Percent Category Titles**
				Lower (a)	**Upper (a)**			
Low	90	Coordination Chemistry Reviews	40	29	55	26	25	27.8%
Medium	18	Chemical Reviews	179	154	207	53	9	50.0%
High	12	Tetrahedron	467	425	511	86	4	33.3%
SCI Total Citations								
Ordinal Category	**No. Category Titles**	**Title**	**SCI Total Citations**	**95% Poisson Confidence Limits**		**95% Confidence Interval**	**No. Titles within Interval**	**Percent Category Titles**
				Lower (a)	**Upper (a)**			
Low	90	Chromatographia	2553	2455	2654	199	3	3.3%
Medium	18	Tetrahedron	19034	18764	19306	542	1	5.6%
High	12	Journal of Physical Chemistry	54721	54263	55181	918	1	8.3%
SCI Impact Factor								
Ordinal Category	**No. Category Titles**	**Title**	**SCI Impact Factor**	**Maximum 95% Poisson Confidence Limits (b)**		**Maximum 95% Confidence Interval**	**No. Titles within Interval**	**Percent Category Titles**
				Lower	**Upper**			
Low	90	Electrochimica Acta	1.307	1.136	1.505	0.369	20	22.2%
Medium	18	Progress in Lipid Research	3.244	1.999	5.353	3.354	35	194.4%
High	12	Angewandte Chemie (International English Edition)	6.168	5.629	6.767	1.138	3	25.0%

(a) Rounded to nearest whole number.

(b) The lower maximum 95% Poisson confidence limit was calculated by dividing the lower 95% Poisson confidence limit of two-year citations by the higher 95% Poisson confidence limit of two-year citable items, whereas the higher maximum 95% Poisson confidence was calculated by dividing the higher 95% Poisson confidence limit of two-year citations by lower 95% confidence limit of two-year citable items.

Table 18 Spearman rho Correlation Matrix for LSU Faculty Score, University of Illinois (UI) Chemistry Library Use, Science Citation Index (SCI) Total Citations, and Science Citation Index (SCI) Impact Factor

	UI Chemistry Library Use	SCI Total Citations	SCI Impact Factor
LSU Faculty Score	0.77	0.68	0.39
UI Chemistry Library Use		0.84	0.42
SCI Total Citations			0.49

All correlations significant at the 0.0005 level.

Table 19. Rank Differences per Ordinal Category in Spearman rho Correlations between LSU Faculty Score, University of Illinois (UI) Chemistry Library Use, Science Citation Index (SCI) Total Citations, and Science Citation Index (SCI) Impact Factor

Ordinal Category	Range of Ranking Measure		No Titles	% Total Titles	Squared Rank Differences	% Total Squared Rank Differences
	LSU FACULTY SCORE-UI CHEMISTRY LIBRARY USE					
	Titles Ranked in Ascending Order of UI Chemistry Library Use					
	Range Limits	*Range Length*				
Low	1 - 138	138	90	75.0%	56189.00	83.4%
Medium	141 - 294	154	18	15.0%	10748.00	15.9%
High	315 - 2667	2352	12	10.0%	464.00	0.7%
Totals	NA	NA	120	100.0%	67401.00	100.0%
	LSU FACULTY SCORE-SCI TOTAL CITATIONS					
	Titles Ranked in Ascending Order of SCI Total Citations					
Ordinal Category	*Range Limits*	*Range Length*	*No Titles*	*% Total Titles*	*Squared Rank Differences*	*% Total Squared Rank Differences*
Low	255 - 12509	12255	90	75.0%	79353.00	86.3%
Medium	12999 - 34414	21416	18	15.0%	7634.75	8.3%
High	39801 - 231324	191523	12	10.0%	4984.75	5.4%
Totals	NA	NA	120	100.0%	91972.50	100.0%
	LSU FACULTY SCORE-SCI Impact Factor					
	Titles Ranked in Ascending Order of SCI Impact Factor					
Ordinal Category	*Range Limits*	*Range Length*	*No Titles*	*% Total Titles*	*Squared Rank Differences*	*% Total Squared Rank Differences*
Low	0.273 - 2.646	2.374	90	75.0%	103517.50	59.1%
Medium	2.648 - 4.075	1.428	18	15.0%	28792.00	16.4%
High	5.109 - 15.748	10.64	12	10.0%	42791.00	24.4%
Totals	NA	NA	120	100.0%	175100.50	100.0%

Table 19 [Cont.]. Rank Differences per Ordinal Category in Spearman rho Correlations between LSU Faculty Score, University of Illinois (UI) Chemistry Library Use, Science Citation Index (SCI) Total Citations, and Science Citation Index (SCI) Impact Factor

UI CHEMISTRY LIBRARY USE-SCI TOTAL CITATIONS						
Titles Ranked in Ascending Order of SCI Total Citations						
Ordinal Category	*Range of Ranking Measure*		*No Titles*	*% Total Titles*	*Squared Rank Differences*	*% Total Squared Rank Differences*
	Range Limits	*Range Length*				
Low	255 - 12509	12255	90	75.0%	38083.25	82.6%
Medium	12999 - 34414	21416	18	15.0%	4330.75	9.4%
High	39801 - 231324	191523	12	10.0%	3686.00	8.0%
Totals	NA	NA	120	100.0%	46100.00	100.0%
UI CHEMISTRY LIBRARY USE-SCI IMPACT FACTOR						
Titles Ranked in Ascending Order of SCI Impact Factor						
Ordinal Category	*Range of Ranking Measure*		*No Titles*	*% Total Titles*	*Squared Rank Differences*	*% Total Squared Rank Differences*
	Range Limits	*Range Length*				
Low	0.273 - 2.646	2.374	90	75.0%	104077.00	62.0%
Medium	2.648 - 4.075	1.428	18	15.0%	21311.75	12.7%
High	5.109 - 15.748	10.64	12	10.0%	42508.25	25.3%
Totals	NA	NA	120	100.0%	167897.00	100.0%
SCI TOTAL CITATIONS-SCI IMPACT FACTOR						
Titles Ranked in Ascending Order of SCI Impact Factor						
Ordinal Category	*Range of Ranking Measure*		*No Titles*	*% Total Titles*	*Squared Rank Differences*	*% Total Squared Rank Differences*
	Range Limits	*Range Length*				
Low	0.273 - 2.646	2.374	90	75.0%	78755.00	53.6%
Medium	2.648 - 4.075	1.428	18	15.0%	19728.00	13.4%
High	5.109 - 15.748	10.64	12	10.0%	48553.00	33.0%
Totals	NA	NA	120	100.0%	147036.00	100.0%

[Above : TABLES 17-19]

larger number of other titles encompassing a much larger percentage of the titles in their respective ordinal categories within their 95% confidence intervals than the titles selected for the latter measure, where the titles selected for the two higher ordinal categories are the only ones within this interval. Confidence intervals for the impact factor can be considered a special case, because it is possible to hypothesize this measure as a function of two Poisson lambdas: 1) the number of citations to the two-year journal backfile; and 2) the number of items published in this backfile and judged to be citable. Therefore, it was deemed necessary to calculate what is termed in Table 17 the "Maximum 95% Confidence Interval." This interval was derived by dividing the lower confidence limit of two-year citations by the upper confidence limit of the two-year source items to calculate the lower maximum confidence limit of the impact factor and by dividing the upper confidence limit of the two-year citations by the lower confidence limit of the two-year source items to calculate the upper maximum confidence limit of the impact factor. Inspection of Table 17 reveals that there is considerable scope for ordinal rank churning with the impact factor particularly at the lower frequencies. The seemingly anomalous results for the journal *Progress in Lipid Research*—35 titles within its maximum confidence interval that extends far outside its ordinal category—is an artifact of its low observed lambdas, 133 citations and only 41 citable items, causing their interval-to-lambda ratios to exceed by far those of the other two journals. Thus, it appears that Brookes was correct in urging caution in using Spearman's rho for comparative purposes due to sampling variance particularly at the lower frequencies.

Table 18 above is the Spearman rho correlation matrix of LSU faculty score, UI library use, *SCI* total citations, and *SCI* impact factor. This table corroborates the findings of the Pearson r analysis, i.e., that LSU faculty score, UI library use, and *SCI* total citations are much more highly inter-correlated with each other than any of them is with *SCI* impact factor. Thus, whereas the range of Spearman rhos runs from 0.68 to 0.84 for the correlations among LSU faculty score, UI library use, and *SCI* total citations, the range of rhos is from 0.39 to 0.49 for the correlations of these measures with *SCI* impact factor. The Spearman rho is based upon squared rank differences. In calculating this measure, the journals were first sorted in ascending order by that variable either considered more "numerical"—UI library use, *SCI* total citations, and *SCI* impact factor instead of LSU faculty score—or by the variable considered of most interest—citations with preference given to the impact factor measure. This had the effect making 5 of the 6 rankings based by upon citations, of which three are sorts by the impact factor. The journals were then grouped into the above three ordinal categories—Low, Medium, and High—on the basis of the sorting measure. The squared rank differences were then distributed across these ordinal categories in order to analyze the factors affecting the Spearman correlations. Table 19 above summarizes the results of this process. What is the most interesting in this table is the dichotomy revealed between the correlations involving the impact factor and those between the other measures. This dichotomy matches the dichotomy in the Spearman rhos presented in Table 18. First, the distribution of squared rank differences across the ordinal categories is much different for the correlations among LSU faculty score, UI library use, and *SCI* total citations than for the correlations of these measures with *SCI* impact factor. For the first set of correlations, the ranges of percentages of total squared rank differences across the ordinal categories are the following: Low – 82.6% to 86.3%; Medium – 8.3% to 15.9%; High – 0.7% to 8.0%. In contrast, for the correlations with the impact factor, the percentage ranges are these:

Low - 53.6% to 62.0%; Medium – 12.7% to 16.4%; High – 24.4% to 33.0%. From these figures, it can be seen that, with the correlations among LSU faculty score, UI library use, and SCI total citations, the bulk of the squared rank differences is concentrated in the Low category containing 75% of the journals. This can be logically expected, because these are the 90 titles with the lowest frequencies, where the ratio distances between titles are either small or nonexistent. . However, with the correlations involving the impact factor, much of the squared rank differences shifts from the Low category to the High category of 12 titles (10%). To investigate this phenomenon, titles accounting for 4.0% or more of the total squared rank differences were identified and analyzed. This analysis revealed that the titles causing the greatest amount of squared rank differences in the correlations involving the impact factor were small journals in the High category with large proportions of review articles. Thus, one reason for the smaller Spearman correlations of the impact factor with the other measures is that this measure captures the important review function that is overwhelmed in the other measures by the effect of size. Other such functional discrepancies revealed by this technique were that the instructional *Journal of Chemical Education*, the applied title *Chemtech*, as well as the informational titles *Chemical and Engineering News* and *Chemistry in Britain* were more highly rated by the LSU chemistry faculty or more highly used by the UI Chemistry Library patrons than warranted by their citation measures.

Another facet of the dichotomy between the Spearman correlations among LSU faculty score, UI library use, and *SCI* total citations, on the one hand, the correlations of these measures with *SCI* impact factor, on the other, is that the correlations involving the impact factor have not only greater total amounts of squared rank differences but also greater amounts squared rank differences per ordinal category despite the different distributional pattern of the squared rank differences across these categories. For the correlations involving only LSU faculty score, UI library use, and *SCI* total citations, the ranges of squared rank differences are as follows: Low – 38,083.25 to 79,353.00; Medium – 4,330.75 to 10,748.00; High – 464.00 to 4,984.75; and total squared rank differences – 46,100.00 to 91,972.50. For the correlations involving the impact factor, these ranges are as follows: Low – 78,755.00 to 104,077.00; Medium – 19,728.00 to 28,792.00; High – 42,508.25 to 48,553.00; and total squared rank differences – 147,036.00 to 175,100.50. Of great significance here are the generally higher amounts of squared rank differences in the Low category of the correlations involving the impact factor despite the smaller percentages of the total squared rank differences attributable to this category in comparison with the correlations not involving the impact factor. It is difficult not to deduce from the above numbers that one reason for the lower Spearman rhos of the impact factor with the other measures is error resulting from a greater amount of sampling variance in the impact factor due to the interaction effects inherent in its being calculated by dividing one random Poisson variable into another.

The Chi-Squared Test of Independence

The Spearman correlation analysis above has revealed that the measures of journal importance under consideration in this paper contain too much error and are too differently influenced by journal functionalities for comparisons among them to be made on the basis of individual ranks. This is particularly true in respect to the impact factor. In this sense the Spearman analyses have corroborated the findings of the distributional and Pearson correlation investigations. The main sources of error were found to be exogenous citations arising from the subject fuzziness of scientific journal sets and sample variance. Concerning functionality, the

Table 20a. Chi-Squared Test of Independence on Relationship of Four Measures of Journal Importance to Each Other: 3X3 Contingency Table Summary

		UI Chemistry Library Use	SCI Total Citations	SCI Impact Factor
Low Category [Total Titles - 90; Expected No. Titles - 67.5 (75.0%)]				
LSU Faculty Score	Observed No. Titles	83	79	75
	Observed % Total Titles	92.2%	87.8%	83.3%
UI Chemistry Library Use	Observed No. Titles		80	75
	Observed % Total Titles		88.9%	83.3%
SCI Total Citations	Observed No. Titles			76
	Observed % Total Titles			84.4%
Medium Category [Total Titles - 18 Titles; Expected No. Titles - 2.7 (15.0%)]				
		UI Chemistry Library Use	**SCI Total Citations**	**SCI Impact Factor**
LSU Faculty Score	Observed No. Titles	7	7	4
	Observed % Total Titles	38.9%	38.9%	22.2%
UI Chemistry Library Use	Observed No. Titles		6	4
	Observed % Total Titles		33.3%	22.2%
SCI Total Citations	Observed No. Titles			7
	Observed % Total Titles			38.9%
High Category [Total Titles - 12; Expected No. Titles - 1.2 (10.0%)]				
		UI Chemistry Library Use	**SCI Total Citations**	**SCI Impact Factor**
LSU Faculty Score	Observed No. Titles	8	6	5
	Observed % Total Titles	66.7%	50.0%	41.7%
UI Chemistry Library Use	Observed No. Titles		8	3
	Observed % Total Titles		66.7%	25.0%
SCI Total Citations	Observed No. Titles			3
	Observed % Total Titles			25.0%

[Above : TABLE 20a]

two citation measures were found to be differently influenced by journal function. *SCI* total citations better capture the research function, where SCI impact factor better captures the review function. Both citation measures were proven to underestimate the importance of the instructional, informational, and applied journal functions. All these factors tended to reduce the correlation coefficients and were particularly influential in respect to the impact factor.

Given the nature of the data and the variables, a fair statistical technique should be robust against error and allow wide scope for the differing effects of journal functionality. One way to do this is to base the analysis upon amount of joint location within broadly defined ordinal categories instead of precision of fit to a regression line or matches in individual ranks. Such a method has the advantage of allowing one to define the amount of error one is willing to tolerate. For example, one can specify that a given title be located in the upper 10% of the journals on two measures, or one can specify that a given title be located in the upper 25% of the journals on two measures. It is possible to design such a test through the chi-squared test of independence, which is also known as the chi-squared test of association or homogeneity. The test requires only nominal or categorical variables, but the observations can be classified into ordinal categories specifying the amount of error. To investigate the results of using the chi-squared test of independence, this method was applied to the ordinal categories defined in Table 16 above.

The first step in applying the chi-squared test of independence to these ordinal categories was to construct 3X3 contingency tables to compare two measures of journal importance against each other by their title overlap in each ordinal category. For each ordinal category, there was calculated the number of overlapping titles that could be expected on the condition that the two measures were unrelated and the overlap was what could be expected as a result of the proportion of the titles of the category in the sample. These expected numbers of overlapping titles are then compared against the number of overlapping titles actually observed. The results of these tests are summarized in Table 20a above, which only shows how well each measure matched the other measures in the category of interest. Starting with the overlap of titles in the Low category, the expected number of overlapping titles is 67.5 or 75% of the total possible overlap of 90 titles. The observed numbers of overlapping titles were all far above that, ranging from a low of 75 or 83.3% of the total possible overlap of *SCI* impact factor with both LSU faculty score and UI library use to a high of 83 or 92.2% of the total possible overlap between LSU faculty score and UI library use. In other words, there was a fairly high level of agreement among all the measures as to which titles can be considered insignificant. However, it should be noted that the *SCI* impact factor overlap with the other measures was consistently lower than the overlap of the other measures among themselves.

Inspection of the results for the Medium and High categories leads to a number of interesting conclusions. First, once again, it was proven that the measures are not independent but strongly related to each other. For the Medium category, the expected overlap was 2.7 titles or 15.0% of the possible overlap. This expected overlap was exceeded in all cases by the observed overlap, which ranged from a low of 4 titles or 22.2% of the possible overlap of *SCI* impact factor with both LSU faculty score and UI library use to a high of 7 titles or 38.9% of the possible overlap of LSU faculty score with both UI library use and *SCI* total citations. For the High category, the expected overlap was 1.2 titles or 10.0% of the possible overlap. Once again, the expected overlap was exceeded in all cases by the observed overlap, which ranged from a low of 3 titles or 25.0% of the possible overlap of *SCI* impact factor with both UI library use and

Table 20b. Chi-Squared Test of Independence on Relationship of Four Measures of Journal Importance to Each Other: 2X2 Contingency Table Summary				
Low Category [Total Titles - 90; Expected No. Titles - 67.5 (75.0%)]				
		UI Chemistry Library Use	**SCI Total Citations**	**SCI Impact Factor**
LSU Faculty Score	Observed No. Titles	83	79	75
	Observed % Total Titles	92.2%	87.8%	83.3%
UI Chemistry Library Use	Observed No. Titles		80	75
	Observed % Total Titles		88.9%	83.3%
SCI Total Citations	Observed No. Titles			76
	Observed % Total Titles			84.4%
Medium/High Category [Total Titles - 30; Expected No. Titles - 7.5 (25.0%)]				
		UI Chemistry Library Use	**SCI Total Citations**	**SCI Impact Factor**
LSU Faculty Score	Observed No. Titles	23	19	15
	Observed % Total Titles	76.7%	63.3%	50.0%
UI Chemistry Library Use	Observed No. Titles		20	15
	Observed % Total Titles		66.7%	50.0%
SCI Total Citations	Observed No. Titles			16
	Observed % Total Titles			53.3%
The chi-squared tests of independence using 2X2 contingency tables were all significant at the 0.001 level.				

[Above : TABLE 20b]

SCI total citations to a high of 8 or 66.7% of the possible overlap between LSU faculty score and UI library use. Second, the percentages of the possible overlap in the High category, which ranged from 25.0% to 66.7%, tended to be greater than the percentages of possible overlap in the Medium category, which ranged from 22.2% to 38.9%. Taken together with the high percentages of possible overlap in the Low category, this indicates that the measures of journal importance were more consistent with each other at the extremes of the distributions. And, finally, once again, the percentages of possible overlap of *SCI* impact factor with the other measures were for the most part smaller in the Medium and High categories than the percentages of possible overlap of the other measures among each other. There are two very interesting anomalies in the overlap pattern of *SCI* impact factor with the other measures. First, for the Medium category, the percentage of possible overlap of *SCI* impact factor with *SCI* total citations is much higher than its percentages of possible overlap with the other two measures. This is a sign that at the Medium level *SCI* impact factor was picking out key research journals, whose rankings were being lowered by the propensity of the impact factor to rank review journals higher. Second, in the High category, *SCI* impact factor's percentage of possible overlap with LSU faculty score is much higher than its percentage of possible overlap with the other measures. The reason for this is that, unlike UI library use and *SCI* total citations, which were heavily influenced by size, there were two important review journals among the top 12 journals ranked by LSU faculty score. These tests indicate that the more precisely one defines the journal sets in terms of function and the more robust the test is against error, the more *SCI* total citations and *SCI* impact factor tend to be equivalent as measures of journal importance.

The standard rules for conducting the chi-squared test of independence are summarized by Cochran (1952; 1954) and Siegel (1956, pp. 94-116 and 174-179). According to these rules, the test requires that the expected frequencies not be too small, with less than 20% of the cells having an expected frequency less than 5. Otherwise the cells must be combined in such a way as to satisfaction this condition. The 3X3 contingency tables summarized in Table 20a did not meet this condition, but combining the Medium and High categories into the Medium/High category and constructing 2X2 contingency tables satisfied it. Combining these categories also provided an opportunity to analyze how increasing the scope for error affected the results of the test. The results of the tests based on the 2X2 contingency tables are summarized in Table 20b above. All these tests rejected the null hypothesis of no relationship among the measures of journal importance at the 0.001 level of significance.

Table 20b corroborates previous findings and conclusions. Despite the greater allowance for error through the combination of the Medium and High categories into the Medium/High category, the overlap among LSU faculty score, UI library use, and *SCI* total citations are consistently larger among each among themselves than the overlap of any of these measures with *SCI* impact factor. Thus, the percentages of total possible overlap among LSU faculty score, UI library use, and *SCI* total citations ranges from 63.3% to 76.7%, whereas the percentages of total possible overlap of SCI impact factor with these other measures range from 50.0% to 53.3%. Thus, *SCI* total citations appear to be a better holistic measure of journal importance than *SCI* impact factor. However, the performance of *SCI* impact factor is quite credible, and the analysis based on the Pearson *r* revealed that this performance would greatly improve, if the set were better defined in terms of journal function. Evidence of this was also found in the Spearman rho analysis.

Summary and Conclusions

This section demonstrated that total citations are a better holistic measure of journal importance than is the impact factor. This holds true even when allowance is made for the greater proportion of random error in the variance of the impact factor. The main reason for this is that LSU faculty score, UI library use, and *SCI* total citations are all largely functions of bibliometric size, which the impact factor was specifically designed to control. However, the impact factor's performance was a credible one, and its performance would have been greatly enhanced and approximated that of total citations, if the sets had been better defined in terms of the research vs. review functions of scientific journals. This section also showed that better functional set definition should not be limited to this. Both citation measures tended to underestimate the importance of journals dedicated to instruction, news, and practical techniques.

The need for better set definition is in accordance with the frequency theory of probability, upon which standard statistical methods are based. According to this theory, the probability of an event should be defined as the limit of its relative frequency in a large number of trials within a well-defined set. Failure to create such a set introduces error resulting from exogenous variables. Given Bradford's and Garfield's laws, it is extremely difficult if not impossible to delimit precisely a disciplinary set of scientific journals. The outlier analysis in the Pearson r part of this section confirmed that a large amount of the error was due to the anomalous relationship of chemistry to biochemistry as well as the decision to include spectroscopy journals within the sample due to the emphasis placed upon it by the LSU Department of Chemistry.

However, the fuzzy nature of scientific journal sets is not the only source of error. As stated above, Brookes (1976) defined ISI citation lists as samples of the current scene, and he proved them to be unreliable at the lower citation frequencies due to sample variance. Brookes' view on the role of sample variance was corroborated during the Spearman rho analysis through the use Poisson confidence intervals. Here it was also shown that the impact factor is more prone to errors resulting sampling variance due to its being calculated by dividing one Poisson lambda—number of citations to a two-year journal backfile—by another Poisson lambda—the number of items in this backfile judged to be "citable." The effect of this is compounded by the difficulty of defining precisely what a "citable" item is. Moreover, given the nature of compound Poisson distributions, most scientific journals are constricted to a short segment at the lowest end of the citation frequency range, where the ratio distances between them are extremely small or nonexistent. At this level quantitative distinctions and rankings based upon them are essentially meaningless due to random error alone. Therefore, instead of using correlation techniques based upon linear fits and individual rankings for comparative purposes, it seems much more rational to utilize a technique such as the chi-squared test of independence that forces one to recognize openly this basic fact and to define the amount of acceptable error.

6. THE STABILITY OF THE CITATION MEASURES ACROSS TIME

Hypothesis and Data

A hallmark of the type of distributions under discussion is the stability of the rankings based upon them over time. Examples of such stability have been pointed out and discussed above. Thus, Garfield found that the journals, which comprise the small, interdisciplinary core of titles posited by his Law of Concentration as dominating the scientific journal system in terms of total citations, tend to maintain their dominance for decades. Moreover, as measured by subjective peer ratings, eleven programs remained at the apex of US research-doctorate programs in chemistry for the period from 1924 through 1993. The concentration of total citations upon both these elite chemistry programs and the journals affiliated with US scientific associations as well as the high correlations, which have been found between scientists' subjective ratings and total citations, show a link between the stable social stratification system of US academic chemistry and the chemistry journal system. The issue of stability is of crucial importance in the evaluation of total citations as a measure of the historical significance of journals vs. the impact factor as a measure of their current significance. If the citation patterns are stable over time, then the journals, which are historically significant, should also be currently significant, and the two measures should have a tendency to pick out the same journals as significant, provided the journal set is more precisely defined in terms function. That this is the case was indicated when Garfield deliberately restricted his sample to those journals publishing 100 items or more in a given year to exclude review journals and 12 of the 25 journals highest on the impact factor also appeared among the 25 journals highest on total citations.

Such stability can be hypothesized to be the result of the combined effect of probabilistic heterogeneity and contagion, which are the two stochastic processes leading to the formation of compound Poisson distributions. Contagion is the statistical model of cumulative advantage or the Matthew Effect. These two stochastic processes can operate interactively. For example, if certain elements of a set are probabilistically advantaged at the start and success breeds success, whether or not failure leads to further failure, then contagion can reinforce the initial probabilistic differences and make them permanent. This, in turn, makes the results of a given sampling period dependent on the results of previous sampling periods, causing high inter-correlations between these periods. Under such conditions, citation measures taken at different points in time should reveal the dominant journals increasing or, at least, maintaining their dominance, and there should be high inter-correlations between the citation rankings of the different periods.

To test this hypothesis, the *SCI* total citations and *SCI* impact factor patterns of the chemistry journal sample were analyzed over the eleven-year period 1993-2003. The first step in this analysis was to construct two twenty-title sub-samples—one for each measure—from the full 120-title sample. It was decided to include in these sub-samples only bibliographically stable titles, i.e., only those that did not experience any title changes, separation into parts, merger of parts, absorptions of other titles, etc., during the period. The reasons for this were twofold. One—and not the least important one—was to avoid any complex data manipulations to standardize the titles that would be necessitated by the absence of bibliographic stability. The other was to obtain a clearer picture of any stochastic processes affecting the stability of the two citation measures over time without the interference of exogenous variables arising from bibliographic instability. Both sub-samples were constructed in the same manner. First, the full 120-title sample was ranked in descending order by the citation measure. Then, starting from the highest title fully in the chemistry set, titles were selected by counting down 6 titles. If the title was bibliographically stable, it was included in the sub-sample; if not, the titles immediately above and below it were investigated in a standardized fashion until one was found that was bibliographically stable. By this method it was hoped to obtain bibliographically stable titles representative of all levels of the citation measures.

The first stage of the analysis was to investigate the distributions and patterns of the two citation measures over every year of the period under study. To do this, the four titles at the upper quartile limits of the sub-samples of both measures were isolated for purposes of deeper

Table 21. Key Statistical Measures on the Stability of Total Citations of the Journals at Upper Quartile Limits of the Twenty-Title Sample Constructed to Test the Stability of This Measure over the Time Period 1993-2003

Changes in Total Citations over Time

Title	Slope	Total Citations 1993	Total Citations 1998	Total Citations 2003	% Change 1993-1998	% Change 1993-2003
Journal of the American Chemical Society	5616.85	148900	179036	212938	20.24%	43.01%
Journal of Organometallic Chemistry	274.07	17651	19746	20848	11.87%	18.11%
Langmuir	3134.66	6099	18038	38487	195.75%	531.04%
Chemistry and Physics of Lipids	30.66	1978	2576	2491	30.23%	25.94%

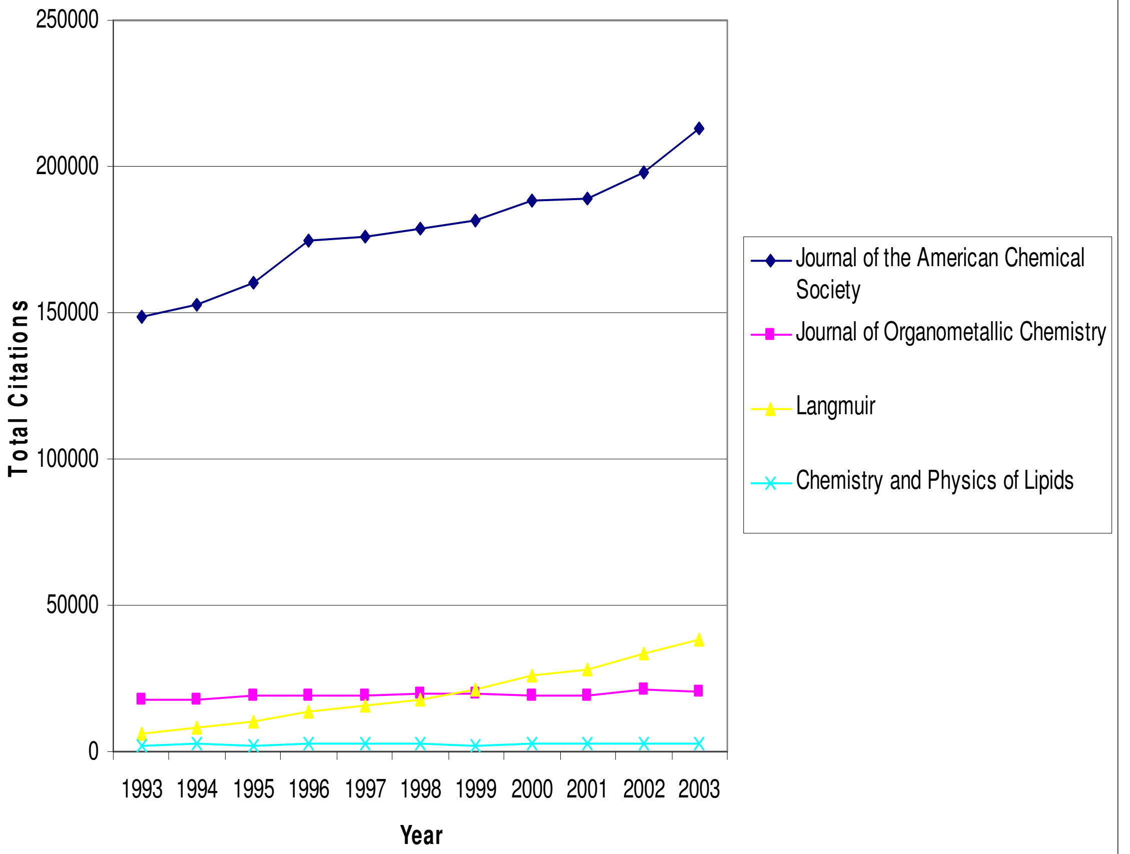
Linearity and Predictive Error

Title	Linearity R^2	Predictive Error: Total Citations Standard Error	Predictive Error: 95% Confidence Interval Width	Ratio 95% Confidence Interval Width to Total Citations 1993	Ratio 95% Confidence Interval Width to Total Citations 1998	Ratio 95% Confidence Interval Width to Total Citations 2003
Journal of the American Chemical Society	0.95	4431.81	17727.26	0.12	0.10	0.08
Journal of Organometallic Chemistry	0.70	622.21	2488.86	0.14	0.13	0.12
Langmuir	0.98	1643.36	6573.45	1.08	0.36	0.17
Chemistry and Physics of Lipids	0.29	168.56	674.25	0.34	0.26	0.27

Changes in Total Citations Sample Rank over Time

Title	Sample Rank 1993	Sample Rank 1998	Sample Rank 2003
Journal of the American Chemical Society	1	1	1
Journal of Organometallic Chemistry	5	5	6
Langmuir	10	6	5
Chemistry and Physics of Lipids	15	15	17

Figure 11. Temporal total citations patterns of the journals at the upper quartile limits of the twenty-journal sample constructed to test the stability of this measure over the time period 1993-2003



[Above : TABLE 21 AND FIGURE 11]

analysis. For total citations the four titles in descending order were: the *Journal of the American Chemical Society*; the *Journal of Organometallic Chemistry*; *Langmuir*; and *Chemistry and Physics of Lipids*. Two of these journals—the *Journal of the American Chemical Society* and *Langmuir*—were American Chemical Society titles; the other two were foreign non-association titles. In respect to the impact factor, the four titles in descending order were: *Chemical Reviews*; *Chemical Physics Letters*; *Synthesis-Stuttgart*; and *Talanta*. Here only the top title—*Chemical Reviews*—was an American Chemical Society periodical; the other three were again foreign, non-association serials.

Statistical Tests of the Hypothesis

Two statistical methods were utilized to test and compare the stability of total citations and the impact factor over time. The first was to do a linear regression of the citation measures on the 11 years of the observation period 1993-2003 for the four journals at the upper quartile limits of the sub-sample of 20 titles constructed for each citation measure. This is a standard method of time series analysis, and it yields a number statistics, of which the following were considered the most important for the purposes of the analysis. The first is the slope of the regression line, which here is the average annual change in the citation measure for the 11 years of the observation period. The second is the R^2 or the coefficient of determination. R^2 is a measure of the closeness of the fit of the data points to the regression line, and its possible values range from 0 to +1. Zero indicates that the points are widely and randomly scattered around the regression line, whereas +1 describes a situation, where all the data points fall precisely on the regression line. R^2 is a measure of both the accuracy of the regression prediction and the volatility of the data. The third statistic is the standard error of the citation measure, which is closely related to R^2. It is the estimated standard deviation of the actual value of the citation measure from the predicted value and hence is a sort of average error in the regression's prediction of the citation measure. The 95% confidence interval of the citation measure for any given year is defined as 2 standard errors on either side of the observed value of the citation measure for that year. To obtain some idea of the relative amount of predictive error, the range or width of the 95% confidence interval was calculated, and for three key years the 95% confidence interval width was divided by the observed value of the citation measure to calculate a ratio of predictive error to the observed value of the citation measure. The other statistical method for testing the stability of the citation measures over time was to construct a Spearman rank-order correlation matrix for the full sub-sample of 20 journals in the years 1993, 1998, and 2003.

The Stability of SCI Total Citations

The results of the time series regression analysis of the *SCI* total citations of the four journals at the upper quartile limits of the 20-title sub-sample constructed to test the stability of this measure over time are summarized in Table 21 and graphed in Figure 11 above. What is most striking about these results is the steep slope of the two American Chemical Society titles—the *Journal of the American Chemical Society* and *Langmuir*—in comparison to the two foreign non-association titles. The first journal is seen to be adding on the average 3.8% of its already high 1993 total citations per year, whereas the second title is seen to be adding on the average a stunning 51.4% of its 1993 total citations per year. The *Journal of the American Chemical*

Table 22. Spearman rho Correlation Matrix for the Full Sample of 20 Journals Constructed to Test the Stability of Total Citations over the Time Period 1993-2003

Year	**1998**	**2003**
1993	0.97	0.90
1998		0.96

All correlations significant at the 0.01 level

[Above : TABLE 22]

Society increased its total citations by 20.24% in the 1993-1998 period and by 43.01% in the 1993-2003 period. *Langmuir* had an even more spectacular rise, increasing its total citations by 195.75% in the 1993-1998 period and 531.04% in the 1993-2003 period. An explanation for *Langmuir's* performance is that it was a relatively new journal established in 1985. Figure 11 shows the *Journal of the American Chemical Society* maintaining and increasing its dominance over the observation period, while at the same time *Langmuir* is rising to prominence. This is confirmed by the rank order changes shown in Table 21 for the four journals in the 20-title sub-sample over the period. The *Journal of the American Chemical Society* maintained its top rank in 1993, 1998, and 2003, whereas *Langmuir* rose from the tenth position to the fifth position over the observation period. In terms of linearity both American Chemical Society titles had extraordinarily good fits to the regression line—R^2 equaling 0.95 and 0.98 respectively—manifesting strong causal processes and lack of randomness. This is confirmed for the *Journal of the American Chemical Society* by its low ratios of 95% confidence interval width to observed total citations in 1993, 1998, and 2003. *Langmuir*'s failure on this measure can be explained by the fact that the confidence interval is an average for the entire observation period, during which *Langmuir*'s total citations were rapidly increasing. In contrast to the American Chemical Society journals, the two foreign non-association titles manifested stagnation and more randomness. Their ranks in the 20-title sub-sample did not change much over the period. Of particular interest are the differences in the linearity and predictability between the two foreign, non-association titles. The *Journal of Organometallic Chemistry* had a rather high total citations rank, which it basically maintained over the period, a rather good fit to the regression line, and low ratios of 95% confidence interval width to observed total citations. In contrast, *Chemistry and Physics of Lipids* had a low total citations rank, poor fit to the regression line, and rather high ratios 95% confidence interval widths to observed total citations. All these are signs of its having a more random total citations pattern over time, which may hypothesized to be the result of the greater effect of sample variance at the lower citation frequencies due to the considerations presented by Brookes (1976).

The performance of the two American Chemical Society titles in comparison to the two foreign, non-association titles is consistent with the two stochastic processes of probabilistic heterogeneity and contagion. These stochastic processes are embodied in Merton's Matthew Effect, which theorizes that such skewed patterns of scientific recognition are a function of both the rise of a highly stratified scientific social system due to cumulative advantage and the misallocation of such recognition as a result of prior achievements. The Matthew Effect has been shown to be operative in the social structure of both scientists and their journals. As US association titles, the American Chemical Society titles rank high in the social stratification system of science, attracting the work of the better scientists at the more highly rated scientific institutions.

Table 22 above summarizes the results of the Spearman rank-order correlation test of the full twenty-title sub-sample constructed to analyze the stability of *SCI* total citations over time. This table reveals an extraordinarily high degree of stability in the rankings of journals with inter-correlations between 1993, 1998, and 2003 ranging from 0.90 to 0.97. As was to be expected, the five-year correlations—0.96 and 0.97—are higher than the ten-year correlation—0.90. A major reason for the lower ten-year correlation was due to the meteoric rise of *Langmuir* that displaced other journals in the rankings. Three basic factors appear to be causal in the high

Table 23. Key Statistical Measures on the Stability of the Impact Factor of the Journals at Upper Quartile Limits of the Twenty-Title Sample Constructed to Test the Stability of This measure over the Time Period 1993-2003

Changes in Impact Factor over Time

		Impact Factor			*% Change*	
Title	***Slope***	***1993***	***1998***	***2003***	***1993-1998***	***1993-2003***
Chemical Reviews	0.74	15.748	20.228	21.036	28.45%	33.58%
Chemical Physics Letters	-0.04	3.018	2.257	2.438	-25.22%	-19.22%
Synthesis-Stuttgart	0.02	1.730	2.150	2.074	24.28%	19.88%
Talanta	0.09	1.129	1.291	2.091	14.35%	85.21%

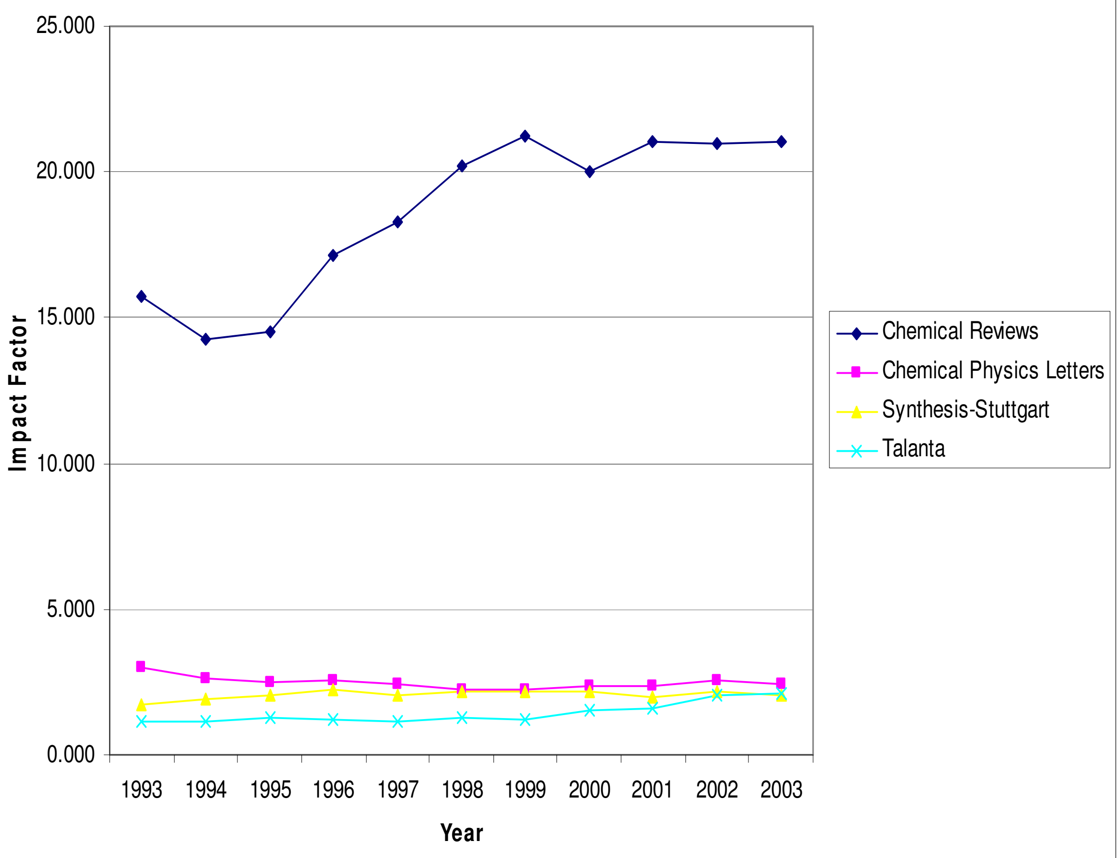
Linearity and Predictive Error

	Linearity	*Predictive Error*				
				Ratio 95% Confidence Interval Width to Impact Factor		
Title	***R^2***	***Impact Factor Standard Error***	***95% Confidence Interval Width***	***1993***	***1998***	***2003***
Chemical Reviews	0.81	1.27	5.09	0.32	0.25	0.24
Chemical Physics Letters	0.38	0.18	0.70	0.23	0.31	0.29
Synthesis-Stuttgart	0.30	0.13	0.53	0.31	0.25	0.26
Talanta	0.73	0.19	0.78	0.69	0.60	0.37

Changes in Impact Factor Sample Rank over Time

	Sample Rank		
Title	***1993***	***1998***	***2003***
Chemical Reviews	1	1	1
Chemical Physics Letters	5	6	6
Synthesis-Stuttgart	10	7	9
Talanta	15	16	8

Figure 12. Temporal impact factor patterns of the journals at the upper quartile limits of the twenty-journal sample constructed to test the stability of this measure over the time period 1993-2003



[Above : TABLE 23 AND FIGURE 12]

degree of stability of the total citations rankings. First, properly understood, the scientific journal system is an outward manifestation of the underlying social structure of science. It has been shown above that, as measured by peer ratings and citations, the social structure of US academic chemistry is highly stable over time. It was also shown that both peer ratings and citations tend to concentrate on both the elite US chemistry programs and US association journals, thereby linking the social structure of chemistry with its journal system. Due to this link, US association journals have a higher probability of being cited, and this advantage is constant over time. *Langmuir* was thus advantaged from the start, but it took some years of backfile buildup for this to register in its total citations. Statistical analyses above indicated that foreign scientific associations appear to play the same role as US associations. The second causal factor in stability lies in the interactive stochastic processes of probabilistic heterogeneity and contagion. As a result of these processes, highly cited journals have a continually greater probability of being highly cited than lower cited ones. The third causal process of the stability of total citations over time relates to Bradford's and Garfield's laws, which dictate that the sets of scientists citing journals must be compound ones composed of subsets of scientists in different disciplines with different probabilities of citing given journals. In two brilliant papers establishing the theoretical necessity of diversifying insurance portfolios, Bortkiewicz (1931;1941), proved that there is an inverse relationship of stability to homogeneity and that the more heterogeneous the set, the more stable is its probability over time. This is because any change in the probability of one subset over time will tend to be counterbalanced and negated by concomitant changes in the probabilities of other subsets.

The Stability of SCI Impact Factor

The results of the time series regression analysis of the impact factors of the four journals at the upper quartile limits of the 20-title sub-sample constructed to test the stability of this measure over time are summarized in Table 23 and graphed in Figure 12 above. Of the four titles under analysis, *Chemical Reviews* was the most highly ranked title on this measure in 1993 for two basic reasons: 1) it was the only American Chemical Society publication, the other three being foreign, non-association serials; and 2) it was the only review journal with 100.0% review articles, whereas the proportion of review articles in the other three titles ranged from 0.0% to 6.0%. The impact factor statistics are interesting both for their differences from and similarities to the total citations ones. On the average, the impact factor slopes are much less steep than the total citations slopes, and this has resulted in less proportional change over time. Such a result is to be expected, because impact factor controls for size by calculating the mean citation rate over a two-year backfile, whereas total citations records the buildup of citations as the backfile increases. However, the American Chemical Society title *Chemical Reviews* has the steepest slope, and Figure 12 shows this periodical to be maintaining and increasing its dominance over the observation period in a similar but more random pattern as the *Journal of the American Chemical Society* did. In terms of linearity, the R^2 measures of impact factor are less on the average than the total citations ones, indicating poorer fits to the regression line and more random scatter about this line. This can be hypothesized to result partly from the random error arising from the difficulty in defining the number of "citable" items for the denominator. However, two titles, *Chemical Reviews* and *Talanta*, have rather high coefficients of determination, 0.81 and 0.73, and interestingly, similar to the total citations results, these are the

Table 24. Spearman rho Correlation Matrix for the Full Sample of 20 Journals Constructed to Test the Stability of Impact Factor over the Time Period 1993-2003

Year	**1998**	**2003**
1993	0.97	0.91
1998		0.88

All correlations significant at the 0.01 level

[Above : TABLE 24]

two titles with the steepest slopes and most proportional change over time. This is indicative of a strong causal process in operation. Such a conclusion is justified by the statistics on the relative amount of predictive error and the changes of the four titles in rank over time. In terms of the relative amount of predictive error, the ratio of 95% confidence interval width to observed impact factor for the top three journals are fairly constant over time and similar to each other. Moreover, there is either no change or only small changes in rank over time. For *Chemical Reviews*, the strong causal process is revealed by its maintaining and increasing its hold on the top rank over time. However, the situation is much different with *Talanta*, which was the lowest ranked of those under analysis in 1993 impact factor. This journal has the highest ratios of 95% confidence interval width to observed impact factor but also the greatest change in rank. But this change took place only in the last half of the period. Close inspection of the data revealed that *Talanta*'s impact factor varied randomly around a mean of 1.202 from 1993 to 1999 but then experienced two stepwise jumps to a mean of 1.571 in 2001-2002 and then to a mean of 2.073 in 2002-2003. This caused the title to jump from being ranked 16th of the twenty-title sub-sample in 1998 to being ranked 8th in 2003. Something significant took place with *Talanta*: either an overall improvement in the title or the random publication of key articles temporarily affecting the two-year measure. Investigation of this problem is beyond the scope of this paper.

Table 24 above presents the results the Spearman rank-order correlation analysis of the stability of the impact factor ranks of the full twenty-title sub-sample constructed for this purpose. Two things are very striking in this table. First, the table indicates that there is a considerable amount of impact factor rank stability over time with correlations ranging from 0.88 to 0.97. Second, the two lower correlations both involve the 2003 ranks, and the correlation of the 1993 ranks with the 2003 ranks (0.91) is even higher than the correlation of 1998 ranks with the 2003 ranks (0.88). Inspection of the calculations revealed that much of the reduction in the correlations involving the 2003 ranks resulted from the sudden jump of *Talanta* in rank in the latter part of the observation period. This jump may or may not have been due to a random event, but it does indicate that impact factor ranks are more prone to sudden shifts than total citations ranks.

Overall it must be concluded that impact factor ranks are very stable over time, although more subject to random variation than total citations ranks. Given the way the impact factor sample was constructed by the random selection of titles at separated rank levels, the analysis indicates that, while titles may vary in impact factor over time, their variation is restricted for the most part within a fairly narrow range. It is possible to explain the stability of impact factor ranks over time by means of statistical theory. Defined in statistical terms, impact factor is the arithmetic mean of citations per article over a two-year period. As has been explained above, statistical theory closely connects the arithmetic mean with the probability governing a given set—in this case a set of articles in a given journal. From this theoretical stance, the high temporal stability of impact factor ranks serves as proof that the relative probability of journals being cited is fairly constant over time.

Summary and Conclusions

The statistical tests in this section revealed that journals rankings based upon both citation measures are highly stable over time. This is the result that was expected from the operation of the two stochastic processes—probabilistic heterogeneity and contagion—causing the shape of their distributions. The effect of contagion modeling cumulative advantage is seen in the continued and increasing dominance of the top journals on both measures. Garfield's observation on the stability of the interdisciplinary core of large research journals posited by his Law of Concentration as dominating the entire scientific journal system in terms of total citations is corroborated by the increasing ascendancy of the *Journal of the American Chemical Society* on the this measure. The stability of the rankings based upon total citations can be partly attributed to the continued buildup of backfiles, but the stability of the rankings based upon the impact factor, which is calculated on the basis of moving two-year samples, is proof that the relative probabilities of journals being cited are fairly stable across time. However, the statistical tests also showed that the impact factor is less stable across time due to the random errors involved in its calculation and being more prone to being effected by random events such as the publication of a major article particularly at the bottom of the citation range. The link of the chemistry journal system to the highly stratified and stable social system of US academic chemistry is revealed by the top journals on both citation measures being published by the American Chemical Society (ACS) and the meteoric rise of *Langmuir*, another ACS title, in the total citations rankings. It has been shown above that this stratification system is the result of the same two stochastic processes shaping the frequency distributions of the titles across the two citation measures.

This stability of the total citations and impact factor rankings across time ensures that established significant journals should be both historically and currently significant. Therefore, in terms of identifying significant titles, total citations and the impact factor should yield similar results, provided that the journal set is more precisely defined in terms of the research vs. review function and that the method of comparison is robust against random error. That still leaves a role for the impact factor in identifying which newly established journals are significant. For example, in 1993 *Langmuir* ranked higher on the impact factor (33rd) than it did on total citations (55th)—an indication of its future rise on the latter measure. However, it should be pointed out that at the same time it ranked 16th on LSU faculty score, showing that perhaps a consensus of expert opinion is the best predictor of future preeminence.

FINAL CONSIDERATIONS ON THE UTILIZATION OF CITATION MEASURES OF JOURNAL IMPORTANCE

Journal importance is multifaceted, and different measures capture different facets of this importance. For example, the functional composition of the top 12 (10%) journals on each measure of journal importance under consideration is in ascending order of the effect of size upon them: *SCI* impact factor—4 research journals and 8 review journals; LSU faculty score—9 research journals, 2 review journals, and 1 instructional journal; and both UI library use and *SCI* total citations—12 research journals. From this perspective, it appears that the best holistic measure of journal importance is the informed consensus of the expert opinion of the scientists, who read and write the articles in the journals being evaluated. Therefore, the more the citation measure conforms to this consensus of expert opinion, the better a holistic measure it is. Since expert opinion is influenced by size, total citations are the better holistic measure than the impact factor, which was specifically designed to control for size. However, total citations miss the important review function captured by the impact factor. Moreover, due to the powerful causation shaping the frequency distributions of the journals across the measures, the performance of the impact factor as a holistic measure is quite credible, and it approximates that of total citations on the condition of better set definition in terms of journal function and greater allowance for error in the method of comparison. It should be pointed out that both citation measures missed the instructional function captured by LSU faculty score.

Although Garfield considered the impact factor the more important measure and concentrated his attention upon it, he never used it as a holistic measure in determining the journal coverage of the ISI citation indexes. Both in his seminal research on the structure of the scientific journal system and his method of selecting journals for coverage, he always used the impact factor in conjunction with total citations and other measures. Total citations were utilized by Garfield to identify the large research journals comprising the small interdisciplinary core posited by his Law of Concentration as dominating the entire scientific journal system, and the impact factor was utilized by him to pinpoint for coverage those types of journals—review journals, newly established journals, etc.—missed by the former measure. The research presented above has fully validated Garfield's research in respect to these measures as well his method of utilizing them to determine the journal coverage of the ISI citation indexes.

This part will conclude by suggesting some considerations and rules in employing citation measures of journal importance for selection and evaluation purposes. First, one has to make a philosophical decision concerning the research vs. review function of journals as well as the relationship of size to significance. Second, it is necessary to define as precisely as possible the discipline of the journals and scientists being evaluated. Given Bradford's and Garfield's laws, subsets governed by differing probabilities are inevitable, and it may be desirable to perform the evaluation in terms of these subsets. Third, it is also necessary to define the journals bibliographically, deciding whether to combine or treat separately segments demarcated by title changes, different parts, continuations, supersessions, absorptions, etc. Fourth, given the shape of compound Poisson distributions, which make it possible to utilize citation measures to pick out the significant journals, it is not possible to make meaningful quantitative distinctions with these measures among the insignificant journals at the lowest part of the citation range, where the bulk of the titles are compressed. This is particularly true of the impact factor, which has an extremely constricted citation range and is more subject to random error. Fifth, since the distribution of articles over citations is of the same type as the distribution of journals over citations, the impact factor is able to identify significant journals only because it is an estimate of the arithmetic mean of citations per article in a journal. With this type of distribution, the arithmetic mean is skewed upward by the few highly cited items, and the impact factor thus identifies journals with a propensity to publish highly cited articles. If the impact factor were an estimate of a measure of central tendency more representative of the articles in the various journals, the impact factor probably would not be able to distinguish the significant journals from the insignificant ones. Therefore, the impact factor cannot be utilized as a surrogate for the citation rate of a specific article in a journal. And, finally, since journal importance is multifaceted and total citations and the impact factor capture different facets of this importance, neither can be utilized as a single holistic measure of journal importance. Moreover, there are some facets of journal importance that neither citation measure captures. Therefore, total citations and the impact factor should be utilized not only in conjunction with each other but also together with other measures to identify missing facets, error, distortions, etc. In particular, the two citation measures should be tested against the opinion of the scientists, whose interests are most affected by the journal evaluation.

References

Adair, W.C. (1955). Citation indexes for scientific literature? *American Documentation*, 6 (1), 31-32.

Aitchison, J., & Brown, J.A.C. (1957). *The lognormal distribution*. Cambridge: At the University Press.

Andersen, R.C., Narin, F., & McAllister, P. (1978). Publication ratings versus peer ratings of universities. *Journal of the American Society for Information Science*, 29 (2), 91-103.

Bensman, S.J. (1985). Journal collection management as a cumulative advantage process. *College & Research Libraries*, 46 (1), 13-29.

Bensman, S. J. (1996). The structure of the library market for scientific journals: The case of chemistry, *Library Resources & Technical Services,* 40 (2), 145-170.

Bensman, S. J. (2001). Bradford's Law and fuzzy sets: Statistical implications for library analyses. *IFLA Journal*, 27 (4), 238-246.

Bensman, S.J. (2007). Garfield and the impact factor. *Annual Review of Information Science and Technology*, 41, 93-155.

Bensman, S. J., & Wilder, S. J. (1998). Scientific and technical serials holdings optimization in an inefficient market: A LSU Serials Redesign Project exercise, *Library Resources & Technical Services*, 42 (3), 147-242.

Bernal, J.D. (1940). *The social function of science* (2nd corrected ed.). London: George Routledge.

Bernal, J.D. (1948). Preliminary analysis of pilot questionnaire on the use of scientific literature. In *The Royal Society Scientific Information Conference, 21 June-2 July 1948: Report and papers submitted* (pp. 589-637). London: The Royal Society.

Beyer, W.H. (1968). *CRC handbook of tables for probability and statistics* (2nd ed.). Cleveland, Ohio: Chemical Rubber Co.

Bookstein, A. (1997). Informetric distributions, Part III: Ambiguity and randomness. *Journal of the American Society for Information Science*, 48 (1), 2-10.

Bortkiewicz, L.v. (1931). The relations between stability and homogeneity, *Annals of Mathematical Statistics*, 2 (1), 1-22.

Bortkiewicz, L.v. (1941). *Homogeneitaet und Stabilitaet in der Statistik = Homogeneity and stability in statistics* (E. Altenschul, W.C. Waite, and R.Rosendale, Trans.). Work Projects Administration official project, no. 165-1-71-124. Minneapolis: University of Minnesota.

Bradford, S.C. (1934, Jan. 26). Sources of information on specific subjects. *Engineering*, 137, 85-86.

Bradford, S.C. (1948). Complete documentation. In *The Royal Society Empire Scientific Conference, June-July 1946: Report* (Vol. 1, pp. 729-748). London: The Royal Society.

Bradford, S.C. (1953). *Documentation* 2nd ed.). London: Crosby, Lockwood.

Braun, T., and Glänzel, W. (1995). The sweet and sour of journal citation rates: On a source of error in computing impact factors. *Chemical Intelligencer*, January, 31-32.

Brookes, B.C. (1976). Citation *v* usage of serials. *Journal of Documentation*, 32 (4), 320-321.

Brookes, B.C. (1977). Theory of the Bradford Law. *Journal of Documentation*, 33 (3), 180-209.

Brookes, B.C. (1980). The foundations of information science, Part II: Quantitative aspects: Classes of things and the challenge of human individuality. *Journal of Information Science*, 2 (5), 209-221.

Cartter, A.M. (1966). *An assessment of quality in graduate education*. Washington, DC: American Council on Education.

Cattell, J.M. (1906). A statistical study of American men of science. *Science*, 24 (621), 658-665; (622) 699-707; (623), 732-742.

Cattell, J. M. (1910), A further statistical study of American men of science, *Science*, 32 (No. 827), 633-648; (No. 828), 672-688.

Chrzastowski, T. E., (1993), *University of Illinois Chemistry Library 1993 journal-use study data.* Unpublished raw data.

Cochran, W.G. (1952). The X^2 test of goodness of fit. *Annals of Mathematical Statistics*, 23 (3), 315-345.

Cochran, W.G. (1954). Some methods for strengthening the common X^2 tests. *Biometrics*, 10 (4), 417-451.

Cole, J.R., & Cole, S. (1972). The Ortega hypothesis. *Science*, 178 (4059), 368-375.

Cole, S. (2000). The role of journals in the growth of scientific knowledge. In B. Cronin & H.B. Atkins (Eds.), *The web of knowledge: A festschrift in honor of Eugene Garfield* (pp. 109-142). ASIS monograph series. Medford, NJ: Information Today.

Coleman, J.S. (1964). *Introduction to mathematical sociology*. New York: Free Press of Glencoe.

Davis, P., & Papanek, G.F. (1984). Faculty ratings of major economics departments by citations. *American Economic Review*, 74 (1), 225-230.

Eggenberger, F., & Pólya, G. (1984). Űber die Statistik verketteter Vorgänge. In G. Pólya, *Probability; Combinatorics; Teaching and learning in mathematics* (G.-C. Rota, Ed.) (pp. 94-104). Collected papers / George Pólya, vol. 4. Mathematicians of our times, vol. 22. Cambridge, Mass.: MIT Press.

Eisenhart, C. (1983). Laws of error I: Development of the concept. In S. Kotz and N.L. Johnson (Eds.), Encyclopedia of statistical sciences (Vol. 4, pp. 530-547). New York: John Wiley.

Elliott, J.M. (1977). *Some methods for the statistical analysis of samples of benthic invertebrates* (2nd ed.). Freshwater Biological Association scientific publication, no. 25. Ambleside, England: Freshwater Biological Association.

Feller, W. (1943). On a general class of "contagious" distributions. *Annals of Mathematical Statistics*, 14 (4), 389-400.

Fisher, A. (1922). *The mathematical theory of probabilities and its applications to frequency curves and statistical methods*: Vol. 1. *Mathematical probabilities, frequency curves, homograde and heterograde statistics* (2d ed.; C. Dickson & W. Bonynge, Trans.). New York: Macmillan.

Fisher, R.A. (1970). *Statistical methods for research workers* (14th ed.). New York: Hafner.

Garfield, E. (1956). Citation indexes—new paths to scientific knowledge. *Chemical Bulletin*, 43 (4), 11.

Garfield, E. (1970, Apr. 22). Calling attention to Chauncey D. Leake—Renaissance scholar extraordinaire. *Current Contents*, No. 16, 5-6.

Garfield, E. (1970, May 6). What is a significant journal? *Current Contents*, No. 18, 5-6.

Garfield, E. (1971, Aug. 4). The mystery of transposed journal lists—wherein Bradford's Law of Scattering is generalized according to Garfield's Law of Concentration. *Current Contents*, No. 17, 5-6.

Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science*, 178 (4060), 471-479.

Garfield, E. (1972, Feb. 2). What is the "core" literature of biochemistry as compared to the "core" of chemistry? *Current Contents*, No. 4, 5-8.

Garfield, E. (1972, Feb. 23). Citations-to divided by items-published gives journal impact factor; ISI lists the top fifty high-impact journals of science. *Current Contents*, No. 7, 5-8.

Garfield, E. (1973, Sept. 26). Which journals attract the most frequently cited articles? Here's a list of the top fifteen. *Current Contents*, No. 39, 5-6.

Garfield, E. (1974, Oct. 30). So you wanted more review articles—ISI's new *Index to Scientific Reviews* (*ISR*) will help you find them. *Current Contents*, No. 44, 5-6.

Garfield, E. (Ed.). (1976a). *Journal citation reports: A bibliometric analysis of references processed for the 1974 Science Citation Index*. *Science Citation Index* 1975 annual, v. 9. Philadelphia: Institute for Scientific Information.

Garfield, E. (1976b). Significant journals of science. *Nature*, 264 (5587), 609-615.

Garfield, E. (1976, Feb. 9). Is the ratio between number of citations and publications cited a true constant? *Current Contents*, No. 6, 5-7.

Garfield, E. (1978, Feb. 13). To remember Chauncey D. Leake. *Current Contents*, No. 7, 5-15.

Garfield, E. (1983). *Citation indexing—its theory and application in science, technology, and the humanities.* Philadelphia: ISI Press.

Garfield, E. (1983, Nov. 21). What's in a name?: The eponymic route to immortality. *Current Contents*, No. 47, 5-16.

Garfield, E. (1987). Oral history: Transcript of an interview conducted by Arnold Thackray and Jeffrey Sturchio at the Institute for Scientific Information, Philadelphia, PA, on 16 November 1987. Philadelphia: Beckman Center for the History of Chemistry.

Garfield, E. (1987, May 4). Reviewing review literature: Part 1, Definitions and uses of reviews. *Current Contents*, No. 18, 5-8.

Garfield, E. (1990, May 28). How ISI selects journals for coverage: Quantitative and qualitative considerations. *Current Contents*, No. 22, 5-13.

Garfield, E. (1991, Sept. 2). In truth, the 'flood' of scientific literature is only a myth. *The Scientist*, 5 (17), 11.

Garfield, E. (1996, Sept. 2). The significant scientific literature appears in a small core of journals. *The Scientist*, 10 (17), 13.

Garfield, E. (1997). A statistically valid definition of bias is needed to determine whether the *Science Citation Index* discriminates against Third World journals. *Current Scien*ce, 73 (8), 639-641.

Garfield, E. (1998). Random thoughts on citationology: Its theory and practice. *Scientometrics*, 43 (1), 69-76.

Garfield, E. (1999, Nov.). A citation analyst's perspective on Japanese science. Paper presented at ISI Symposiums in Osaka and Tokyo. Retrieved Nov. 10, 2006 from http://www.garfield.library.upenn.edu/papers/japantalky1999.html.

Garfield, E. (2000). In-depth interview: Interview with Eugene Garfield, Ph.D., by Chungji Kim. *Medical Writing*, 8 (1). Retrieved July 20, 2005, from http://www.garfield.library.upenn.edu/papers/medicalwritingv8(1)1999.html.

Garfield, E. (2005). The agony and the ecstasy—the history and meaning of the journal impact factor: Presented at the International Conference on Peer Review and Biomedical Publication. Retrieved November 18, 2005 from http://garfield.library.upenn.edu/papers/jifchicago2005.pdf

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

Garfield, E., & Sher, I.H. (1963). New factors in the evaluation of scientific literature through citation indexing. *American Documentation*, 14 (3), 195-201.

Geary, R.C. (1947). Testing for normality. *Biometrika*, 34 (3/4), 209-242.

Goldberger, M. L., Maher, B. A., & Flattau, P. E. (Eds.). (1995). *Research-doctorate programs in the United States: Continuity and change*. Washington, DC: National Academy Press. The data tables in this book can be downloaded into spreadsheets at the following URL: http://www.nap.edu/readingroom/books/researchdoc/

Greenwood, M. & Yule, G.Y. (1920). An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *Journal of the Royal Statistical Society*, 83 (2), 255-279.

Grieg-Smith, P. (1983). *Quantitative plant ecology* (3d ed.). Studies in ecology, vol. 9. Berkeley: University of California Press.

Hughes, R.M. (1925). *A study of the graduate schools of America*. Oxford, Ohio: Miami University.

Institute for Scientific Information. (1994a). *Science citation index journal citation reports, 1993: A bibliometric analysis of science journals in the ISI database.* Philadelphia, PA: Institute for Scientific Information.

Institute for Scientific Information. (1994b). *Science citation index 1993 annual guide and lists of source publications.* Philadelphia, PA: Institute for Scientific Information.

Institute for Scientific Information. (2005) *ISI journal citation reports: JCR science edition.* Philadelphia, Pa.: Institute for Scientific Information. Available to subscribers at the following URL: http://portal.isiknowledge.com/portal.cgi/

Johnson, N. L., & Kotz, S. (1969). *Discrete distributions*. Distributions in statistics. Boston: Houghton Mifflin.

Jones, L.V., Lindzey, G., & Coggeshall, P.E. (Eds.). (1982). *An assessment of research-doctorate programs in the United States: Mathematical & physical sciences*. Washington, DC: National Academy Press.

Keynes, J.M. (1921) *A treatise on probability*. London: Macmillan.

Lederberg, J. (1977). Foreword. In E. Garfield, *Essays of an information scientist* (Vol. 1: 1962-1973, pp. xi-xv). Philadelphia: ISI Press.

Merton, R.K. (1968). The Matthew Effect in science. *Science*, 159 (3810), 56-63.

Moran, P.A.P. (1968). *An introduction to probability theory*. Oxford: Clarendon Press.

Moroney, M.J. (1956). *Facts from figures* (3rd ed.). Baltimore, Md.: Penguin Books.

Narin, F. (1976). *Evaluative bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity*. Cherry Hill, NJ: Computer Horizons, Inc.

Pearson, E.S., & Hartley, H.O. (Eds.). (1966). *Biometrika tables for statisticians* (3rd ed.). Vol. 1. Cambridge: At the University Press.

Pearson, K. (1956). Contributions to the mathematical theory of evolution II: Skew variation in homogeneous material. In *Karl Pearson's early statistical papers* (pp. 41-112). Cambridge: Cambridge University Press.

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

Price, D.J.d.S. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, 27 (5), 292-306.

Price, D.J.d.S. (1978). Cumulative advantage urn games explained: A reply to Kantor. *Journal of the American Society for Information Science*, 29 (4), 204-206.

Rietz, H.L. (1924) Bernoulli, Poisson, and Lexis distributions. In H.L. Rietz (Ed.), *Handbook of mathematical statistics* (pp., 82-91). Boston: Houghton Mifflin.

Rietz, H.L. (1927). *Mathematical statistics*. The Carus mathematical monographs, no. 3. Chicago: Published for the Mathematical Association of America by Open Court Publishing Co.

Roose, K.D., & Andersen, C.J. (1970). *A rating of graduate programs*. Washington, DC: American Council on Education.

Seglen, P.O. (1992). How representative is the journal impact factor. *Research Evaluation*, 2 (3), 143-149.

Seglen, P.O. (1997). Why the impact factor of journals should not be used for evaluating research. *BMJ: British Medical Journal*, 314 (7079), 497.

Shapiro, S.S. (1980). *How to test normality and other distributional assumptions*. The ASQC basic references in quality control : statistical techniques, v. 3. Milwaukee, Wis.: American Society for Quality Control.

Shapiro, S.S., & Wilk, M.B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52 (3/4), 591-611.

Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill.

Simon, H.A. (1955). On a class of skew distribution functions. *Biometrika*, 42 (3/4), 425-440.

Simon, H.A. (1960). Some further notes on a class of skew distributions functions. *Information and Control*, 3 (1), 80-88.

Snedecor, G.W., & Cochran, W.G. (1989). *Statistical methods* (8th ed.). Ames: Iowa State University Press.

Tempest, D. (2005). The effect of journal title changes on impact factors. Learned Publishing, 18 (1), 67-62.

Von Mises, R. (1957). *Probability, statistics and truth*. 2nd rev. English ed. prepared by H. Geiringer. New York: Macmillan. (Original work published 1951).

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8 (3), 338-353.

Zuckerman, H. (1977). *Scientific elite: Nobel laureates in the United States*. New York: Free Press.