

Where to Spend our E-journal Money? Defining a University Library's Core Collection Through Citation Analysis

Philip M. Davis

abstract: This paper identifies core journals in the life sciences for Cornell University researchers by analyzing the frequency of Cornell-authored citations in Biosis Previews between 1996 and 2001. The distribution frequency of journals confirms Bradford's Law of Scatter or the 80/20 Rule. The top 240 journals, providing 80 percent of the citations, were analyzed by publisher type and institutional subscription price. In general, journals from society and associations received the highest number of citations and were priced considerably lower than commercial journals. The methodology described is a fast, low-cost, and scalable procedure that can be adapted to various subject databases, and may be used to provide guidance on which titles to purchase for electronic access.

With the escalating price of scientific journals, few libraries have money to maintain their current print subscriptions, let alone afford electronic access to all titles within their collections. This means that librarians need to be discriminating selectors, focusing e-journal money on titles that are considered core to the collection. This article is about deciding what is "core."

Many large commercial publishers, however, do not make it easy to pick and choose individual journals. Some offer all-or-nothing deals, which Ken Frazier calls "The Big Deal." In his influential paper, *The Librarians' Dilemma*, Frazier argues that these deals "bundle the strongest with the weakest publisher titles, the essential with the non-essential" and warns that these packages "weaken our collection with journals we neither need nor want."¹ Suzan McGinnis refers to this type of journal selection as a "Selling of our Collecting Souls."²

There are several methods for determining the usefulness of journals for a user community including circulation statistics, faculty surveys, evaluation by subject experts,

portal: Libraries and the Academy, Vol. 2, No. 1 (2002), pp. 155–166. Copyright © 2002 by The Johns Hopkins University Press, Baltimore, MD 21218.

and citation analysis. This study analyses the importance of life science journals for Cornell University by analyzing the frequency of publications authored by our researchers. The top journals are then examined by publisher type and price. While not supplanting other rigorous study methods, the bibliographic analysis presented in this paper provides an easy, fast, and low-cost study that can be adapted to various bibliographic databases and scaled as needed to be used for studying distributed universities, individual campuses or departments.

Review of the Literature

Defining a Core Collection: Bradford's Law and the 80/20 Rule

In 1948, Samuel Bradford dedicated a chapter in his book *Documentation* to his principle of journal scatter.³ His theory states that if scientific journals are arranged in order of decreasing productivity of articles on a given subject, they can be divided into a "nucleus" of journals devoted to the subject with radiating zones of journals contributing fewer and fewer articles. In information science, this is often referred to as journal "dispersion" or "scatter." In economic terms, this principle is often discussed as the Law of Diminishing Returns.⁴

Richard Trueswell first tested this law on library circulation data.⁵ He found that approximately 80 percent of the total number of circulation transactions accounted for only 20 percent of the collection. Trueswell and others have used this theory to posit a core library collection. Many empirical studies have been done to verify the 80/20 Rule, reporting slight variations of the ratio.⁶

To identify journals for a serial cancellation project at the University of Illinois at Urbana-Champaign Chemistry Library, researchers found that 84 percent of journal use was generated by 20 percent of the collection, and that 40 percent of the use was concentrated among only ten journals.⁷

In a citation study of faculty publications, dissertations, and preliminary qualifying briefs at the Biology Library at Temple University, 51 titles (or 15 percent of journals) contributed 80 percent of citations.⁸ This illustration of Bradford's Law of journal dispersion has also been illustrated for several medical topics indexed in Index Medicus.⁹

Lancaster et al used Bradford's "nucleus" and journal zones to determine the collection priorities for a hierarchal network of medical lending libraries involving departmental, university, state and country collections.¹⁰

While Samuel Bradford passed away shortly after *Documentation* was written, one of his most enthusiastic supporters has been Eugene Garfield, founder of the Science Citation Index, who used Bradford's Law as justification for the selectivity of journals included in the index.¹¹ Simply stated, "a surprisingly small number of journals generate the majority of both what is cited and what is published."¹²

Journal Price Studies

Since the mid-1980s there have been numerous journal price studies. The report, "Measuring Journal Cost-Effectiveness: Ten Years After Barschall" ¹³ provides an excellent summary of the legendary study of Physics journals by Henry Barschall ¹⁴ and his sub-



sequent work. It also summarizes the Cornell study on Core Agricultural and Biological Journals¹⁵ and the ongoing Journal Cost-Effectiveness study at the University of Wisconsin.¹⁶ Other use/

price studies are adequately annotated in Soete and Salaba's comprehensive bibliography. Whether measured as cost per title, cost per page, cost per character, or cost by impact factor,

Whether measured as cost per title, cost per page, cost per character, or cost by impact factor, all of these journal price studies point speculatively at the comparatively high prices charged by commercial publishers.

all of these journal price studies point speculatively at the comparatively high prices charged by commercial publishers.

The principal goal of journal price studies is to support local decision-making, chiefly for journal cancellation decisions.¹³ Many of these studies use Impact Factor as a measure of journal importance or prestigiousness. Impact factors are one of several variables reported each year in ISI's Journal Citation Report. They measure the frequency by which the average article gets cited in a given year. Impact Factors provide a "gross approximation of the prestige" of a journal, and according to Garfield, one should not depend entirely on impact factors in making journal comparisons.^{17, 18}

Despite their wide use and appeal, Impact Factors tell you nothing about the local use of journals.¹⁹ In essence, they report the citation patterns of hundreds of thousands of articles published annually by the entire scientific community. While studying the importance of biomedical journals at six libraries, Elizabeth Pan discovered no significant correlation between impact factor rankings and their local use.²⁰

While impact factors are an excellent guide for identifying key journals in particular fields, no campus is typical and as such, impact factors (if used at all) should be supplemented by local data. At the Georgia Institute of Technology, Amy Dykeman analyzed one year of faculty authored citations in the engineering index, INSPEC.²¹ In a study of journals in molecular and cellular biology at Penn State, Janet Hughes reported using a combination of faculty publications and faculty citations along with ISI citation rankings and impact factors. According to Hughes, publication in a journal implies that the journal is not only read but also highly respected by the author.²²

Our study analyses citation frequency as a measure of journal importance. It identifies journals in which Cornell researchers frequently publish, and by extension, represent the scientific communities to which they belong.

Methods

Biosis Previews is widely known as the most comprehensive bibliographic index in the Life Sciences, indexing over 5,000 international serial sources.²³ Citations from Biosis Previews (SilverPlatter interface) were searched for "Cornell" AND "Ithaca or Geneva" in the author affiliation field for the years 1996–2001. Records coming from the Cornell Medical College in New York City were excluded from the search, since they represent a distinct and geographically isolated campus.

Five thousand six hundred thirty-three citations were downloaded, specifying only the Source Field (SO), a compound field including the journal title, volume, issue, pages, etc. The citations were brought into EndNote, a bibliographic management database, so that the source file could be parsed down to leave just the journal title. The records were exported from EndNote into a text file, and imported into MS Excel. A second exported file was run through a UNIX program to remove all duplicates. This second file was used to count the frequency of journal titles in the first export file using the Excel COUNTIF function.

Of the 5633 citations, 852 were unique titles. This is expected since a single journal title may receive multiple citations over the five-year period. Proceedings of conferences, patent gazettes and other non-obvious journal titles were removed leaving 5292 citations comprising 841 unique journals.

The top 240 journals providing 80 percent of the citations in this study were analyzed and placed into one of four publisher categories 1) society/association, 2) commercial, 3) university, and 4) government. Societies or associations publishing with a commercial publishing house were put in the first category since copyright is usually owned by the society/association.¹⁵ Institutional print prices were found for all of the top 240 journals. Prices were found on publisher's web sites or by consulting Ulrich's Periodical Directory. Because of non-standard pricing models for electronic journals, electronic pricing was not investigated.

Five years of citation data were used in the analysis to make sure that journals, which publish infrequent articles by Cornell researchers, were not missed in the study. Several limitations to the methodology should be noted: 1) Journals that publish more frequently (i.e. weekly) will receive more citations than monthly or quarterly titles; 2) new journals (started after 1996) will be underrepresented; 3) journals that have gone through a name change, or are not fully indexed in Biosis Previews will also be underrepresented; and 4) since author affiliation is provided only for the first author, articles that include Cornell researchers as second or third authors will not be counted.

Observations

Distribution of Journals

Between 1996 and 2001, Cornell first-authors contributed 5292 citations published in 841 unique journals. Of these 841 journals:

- 334 journals (40 percent) contributed only one citation each to the study.
- The top 10 journals together (1.2 percent) contributed 1307 (25 percent) of citations
- The top 50 journals (6 percent) contributed 2640 (50 percent) of citations
- The top 240 journals (29 percent) contributed 4233 (80 percent) of citations
- The top 409 journals (49 percent) contributed 4762 (90 percent) of citations

The analysis does not include journals that received no articles from Cornell authors. If we consider that Biosis indexes 5,000 serial sources, then Cornell first-authors contributed to only 17 percent of these journals over the five-year study period.

When the cumulative effect of unique journals is plotted against frequency, it takes on a typical Bradford distribution (figure 1). Additional journals contribute fewer and



fewer citations to the study illustrating diminishing returns. If we were to extend this line to all 5,000 Biosis journals, the slope of the line would become flat and extend five more graph-widths to the right. When plotted on a semi-log graph (figure 2), the first upward curved part of the line describes Bradford's "nucleus" or core group of journals, followed by a straight line. In summary, a small number of journals reflect the general publishing practices of Cornell researchers.

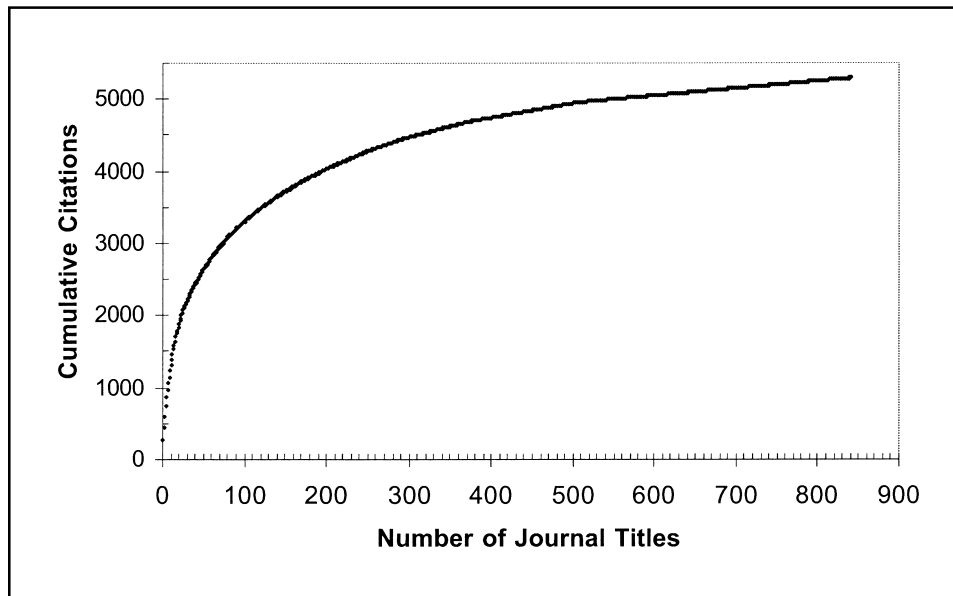


Figure 1. Cumulative Number of Cornell-Authored Publications in Biosis Previews 1996–2001

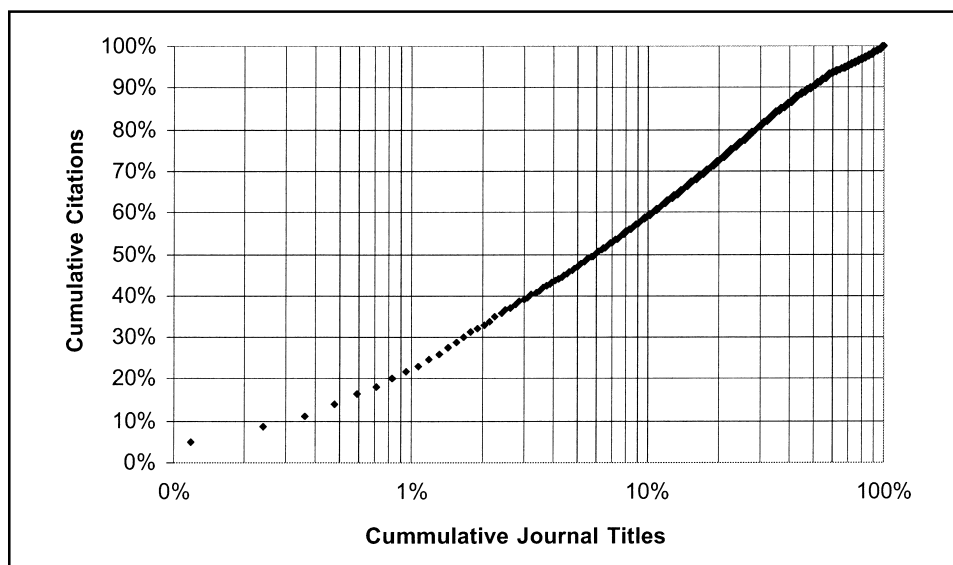


Figure 2. Cumulative Percentage of Cornell-Authored Publications in Biosis Previews 1996–2001



To see whether Cornell researchers exhibit unique publishing patterns that do not reflect other institutions that publish heavily in the Life Sciences, a similar analysis was done for the University of Wisconsin-Madison. The results were almost identical: 49 titles (5 percent) contributed 50 percent of citations and 245 titles (27 percent) contributed 80 percent.

Journal Price Analysis

The top 50 and 240 journals (contributing 50 percent and 80 percent of total citations respectively) were analyzed for their publisher type and institutional price (table 1). The results illustrate that society or association publications dominate the top ranking journals and are significantly less expensive than their commercial counterparts. Figure 3 shows a comparison of journal citation frequency versus print subscription price. The log-log plot illustrates that many commercial journals are more expensive by a factor of ten over their society/association counterparts.

The journal price analysis was almost identical for the University of Wisconsin-Madison. For their top forty-nine journals, thirty-eight (78 percent) were published by societies or associations, at an average price of \$506; eight (16 percent) were commercial journals at an average price of \$2,503.

Table 1

Journal Price Analysis

Journals that make up 50% of Citations (N=50)

Publisher Type	Count	Percent	Ave Price	Median Price
Society or Association	38	76%	662.11	410.50
Commercial	10	20%	2,640.30	2,438.50
University	2	4%	510.00	510.00
Government	0	0%	-	-
Total	50	100%	1,051.66	532.50

Journals that make up 80% of Citations (N=240)

Publisher Type	Count	Percent	Ave Price	Median Price
Society or Association	119	50%	542.41	350.00
Commercial	106	44%	2,415.75	1,549.00
University	10	4%	811.10	480.00
Government	5	2%	576.20	687.00
Total	240	100%	1,381.70	652.00

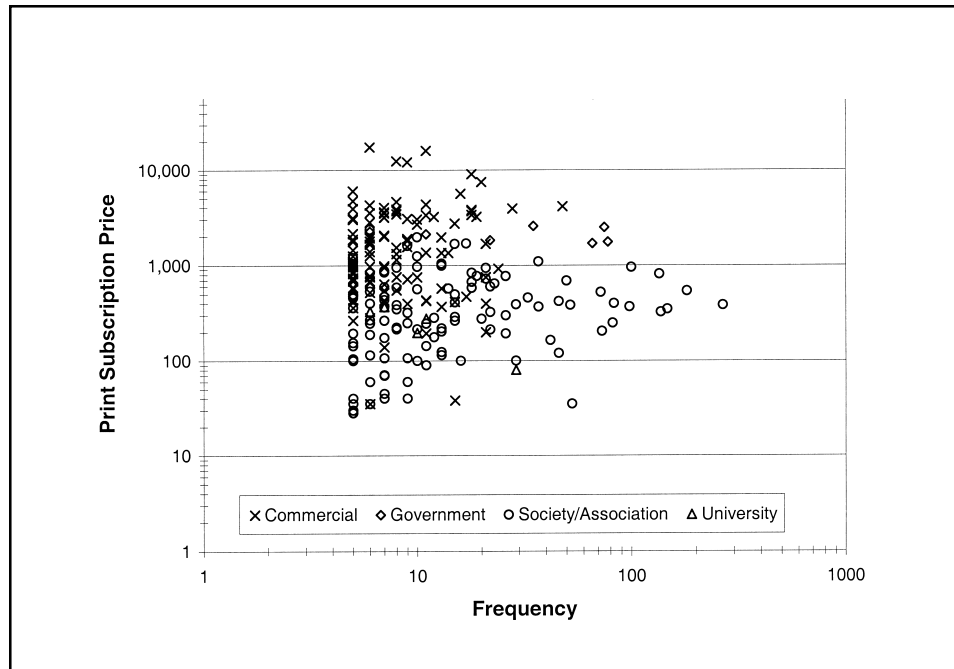


Figure 3. Pricing vs. Frequency of Journals that Account for 80% of Citations (N=240) in Biosis Previews

Comparison of Results to Impact Factor Measures

To verify whether journal impact factors provide comparable results to the local citation data used in this study, the top ten journals identified in this study were checked in ISI's Journal Citation Reports (table 2). JCR uses subject categories, since various disciplines have very different publishing and citation patterns. For example, the highest impact for a journal in the Multidisciplinary Sciences category in 2000 was 25.8, compared to 2.4 for Horticulture. As such, impact factors should not be compared across subject categories. Six of the top ten journals listed in table 2 were somewhat comparable; the others are not easily predictable from one measure to the other.

Our top-ranking journal in this study was *Phytopathology* from the American Phytopathological Society and while it is not the highest-ranked journal in the Plant Sciences category of JCR, the high frequency of citations reflects the research focus and strength of two departments at Cornell University. The generic metrics of the JCR simply cannot provide the campus-level data crucial to making informed decisions about the local importance of individual titles.

Discussion

Where to spend our e-Journal Money

Few libraries have money to maintain their current print subscriptions, let alone afford electronic access to all titles within their collections. This means that librarians need to

Table 2

Top 10 Journals from Bibliometric Analysis Compared to Impact Factors

Rank	Frequency of First-Authored Citations	Title	JCR Category	JCR Category Rank by Impact Factor	Impact Factor (in 2000)
1	266	Phytopathology	Plant Sciences	22 of 137	2.145
2	181	FASEB	Biology	1 of 51	9.249
3	147	J. Dairy Science	Agriculture, Dairy & Animal Sci	1 of 44	1.823
4	137	HortScience	Horticulture	11 of 20	0.470
5	135	BioPhysical Journal	BioPhysics	11 of 66	4.462
6	100	PNAS	Multidisciplinary Sciences	3 of 49	10.789
7	98	IOVS	Ophthalmology	2 of 41	4.373
8	83	Biology of Reproduction	Reproductive Biology	2 of 23	3.605
9	82	J. Animal Science	Agriculture, Dairy & Animal Sci	2 of 44	1.715
10	78	J. Biological Chemistry	Biochemistry & Molecular Biology	27 of 310	7.368

be discriminating selectors, focusing our e-journal money on titles that are considered core to the collection.

This study identifies journals in which Cornell University researchers frequently publish and by extension represent the scientific communities to which our researchers belong.

This study identifies journals in which Cornell University researchers frequently publish and by extension represent the scientific communities to which our researchers belong. Considering the similarity to publishing patterns of University of Wisconsin-Madison researchers, the results of this study may be generalizable outside of Cornell.

This study does not identify which journals our researchers cite, although if such a study were done, it would probably look similar, with the exception of large multidisciplinary science titles moving higher up the ranks. Lascar and Mendelsohn recently illustrated a



high degree of uniformity between where researchers publish and what they cite from a sample of articles in structural biology.²⁴

Where should we spend our e-journal money? It is obvious that we should spend it where it is most effective. Figures 1 and 2 provide us with a guideline for where to start—purchase access to the most important journals and move out as budget allows. Purchasing the top 240 e-journals would cover 80 percent of the articles published by our researchers; 409 titles would cover 90 percent.

The results of the Cornell analysis illustrate that society or association publications dominate the top ranking journals and are significantly less expensive than their commercial counterparts. Many small nonprofit societies and associations, some of which publish only one title each, lack the financial backing and technological infrastructure to move their journals online. This difference gives commercial publishers a marked advantage over nonprofit entities. Those journals that are easily accessed online will be read, cited, and ultimately supported. The society journal that isn't online may be at serious risk of losing readership and ultimately vanishing.

Some societies who do move online have been criticized for the high percentage increases they charge for electronic access to their journals. For a large university, the cost of some of these journals may be in the ballpark of 100 percent over print. Commercial publishers can offer electronic access free with print, or for a nominal 5–20 percent increase over print. The difference of paying 100 percent over print for a prestigious \$400 society journal, however, is still much cheaper than paying an additional 10 percent for a \$2500 commercial journal (\$800 total versus \$2750). The arithmetic of purchasing an entire publisher's package at a single-digit increase over print looks even less favorable, especially in light of the fact that many of the journals included are titles we've never selected in print.

Supporting nonprofit publishers moving to the online environment is ultimately beneficial for libraries since societies and associations provide a low-cost, high-impact, alternative to commercial journals. Without the competition in the online marketplace, commercial journals will gradually win out and libraries

will undoubtedly pay the price with higher subscription fees. Libraries will be left with fewer publisher options, and little (if any) ability to pick and choose individual titles.

Supporting nonprofit publishers moving to the online environment is ultimately beneficial for libraries since societies and associations provide a low-cost, high-impact, alternative to commercial journals

Significance of the Librarians' Dilemma

The identification of a core journal collection alone does not solve the Librarians' Dilemma. Many publishers only offer an all-or-nothing deal for their e-journals, making individual selection a moot point. Other publishers have made individual title pricing more expensive and more restrictive than purchasing the entire package.

In a study of 203 core journals in the fields of political science and economics, Suzanne Gyeszly illustrated that many of the high-priced journals offered through

Elsevier's ScienceDirect received no usage at all during the academic year, despite having to pay for electronic access.²⁵ With licensing clauses that forbid or severely restrict the canceling of current subscriptions, libraries are forced to pay for dual access (paper and electronic) for journals of limited use.

The University of Wisconsin Libraries among others has refused to sign Big Deal packages.¹ The California State University system has also balked at the Big Deal approach arguing that it forces libraries to pay for access to all titles in order to get the ones they desire. Using the influence of the twenty-two-campus university, California State system has put out an RFP, looking for a distributor to put together a custom package of 1,279 titles called the Journal Access Core Collection.²⁶

Which journals and publishers libraries decide to support is an individual decision for each library and we should base our decisions on what is best for our own user community. By applying Bradford's Law, we can make wise use of our serial money; without it, we can easily use it up purchasing large packages of mostly irrelevant titles.

Appendix 1. Data

The data and methodology used in the journal frequency and price analysis can be found at: <http://people.cornell.edu/pages/pmd8/>. The methodology provided in this paper can be modified to work on any bibliographic database, given that the dataset can be downloaded and manipulated. Kushkowsky et al. used a similar methodology, employing several indexes in a single study of journals in industrial relations.²⁷

Appendix 2. Further Reading on Bradford's Law

For a thorough examination of bibliographic analysis, readers are referred to: "The Bradford Distribution: A Review of the Literature 1934-1987," "Bradford's Law and the Bibliography of Science," "Numerical Methods of Bibliographic Analysis," and "Bradford Ranking Conventions and their Application to a Growing Literature."²⁸⁻³¹ Metadata analysis on numerous studies indicates that the specifics of the law (slope and intercept of the line) is not dependent on details such as the subject of the literature, time period, or search technique, but of some probabilistic mechanism underlying the law.³²

Bradford's Law can be used to define the lower limits of a dynamic library collection. The upper limit is a factor of the library's budget and collection priorities.³³ B.C. Brookes established a quantitative method for finding the optimum percentage of journals for a given field. Most journals follow an exponential law of obsolescence, such that the appearance of new journals is counterbalanced by the disappearance of the old.³⁴

Acknowledgements

No paper should be written in a vacuum without the intellectual feedback of colleagues. Special thanks goes to Cornell colleagues Greg Lawrence, Suzanne Cohen, Leah Solla, and Barbara DiSalvo, and to Bill Walters (St. Lawrence University), Ken Frazier (U. Wisconsin), Claudia Lascar and Loren Mendelsohn (City College of New York), and Heather Joseph (BioOne).



The author is a Life Sciences Bibliographer at the Albert R. Mann Library at Cornell University, Ithaca NY; he may be contacted via e-mail at: pmd8@cornell.edu.

Notes

1. Kenneth Frazier, "The Librarians' Dilemma: Contemplating the Costs of the "Big Deal"," *D-Lib Magazine* 7, no. 3 (2001).
2. Suzan D. McGinnis, "Selling Our Collecting Souls: How License Agreements Are Controlling Collection Management," *Research Collections and Digital Information* 31, no. 2 (2000): 63–76.
3. S. C. Bradford, "The Documentary Chaos," in *Documentation* (London: Lockwood, 1948), 144–159.
4. Encyclopædia Britannica, *Diminishing returns, law of* ([cited 27 July 2001]); available from <<http://search.eb.com/bol/topic?eu=30967&sctn=1>>.
5. Richard L. Trueswell, "Some Behavioral Patterns of Library Users: The 80/20 Rule," *Wilson Library Bulletin* 43 (1969): 458–61.
6. Quentin L. Burrell, "The 80/20 Rule: Library Lore or Statistical Law?," *Journal of Documentation* 41, no. 1 (1985): 24–39.
7. Tina E. Chrzastowski and Brian M. Olesko, "Chemistry Journal Use and Cost: Results of a Longitudinal Study," *Library Resources and Technical Services* 41, no. 2 (1997): 101–111.
8. Katherine W. McCain and James E. Bobick, "Patterns of Journal Use in a Departmental Library: A Citation Analysis," *Journal of the American Society for Information Science* 32, no. 4 (1981): 257–267.
9. Donald A. Windsor, "Rational Selection of Primary Journals for a Biomedical Research Library: The Use of Secondary Journal Citations," *Special Libraries* 64, no. 10 (1973): 446–451.
10. F.W. Lancaster et al., "The Relationship between Literature Scatter and Journal Accessibility in an Academic Special Library," *Collection Building* 11, no. 1 (1990): 19–22.
11. Eugene Garfield, "How ISI Selects Journals for Coverage: Quantitative and Qualitative Considerations," *Current Contents* May 28, no. 22 (1990): 5–13.
12. Eugene Garfield, "The Significant Scientific Literature Appears in a Small Core of Journals," *The Scientist* 10, no. 17 (1996): 13.
13. George Soete and Athena Salaba, "Measuring the Cost-Effectiveness of Journals: Ten Years After Barschall," (University of Wisconsin Madison Libraries, 1999), 17.
14. Henry H. Barschall, "The Cost of Physics Journals," *Physics Today* 39, no. 12 (1986): 34–36.
15. College of Agricultural and Life Sciences Faculty Taskforce, Division of Biological Sciences, Albert R. Mann Library, "Journal Price Study: Core Agricultural and Biological Journals," (Ithaca, NY: Cornell University, 1998).
16. ARL, "Measuring the Cost Effectiveness of Journals: The Wisconsin Experience," *ARL Bimonthly Report* 205 (1999).
17. Eugene Garfield, "The Impact Factor," *Current Contents* June 20, no. 25 (1994): 3–8.
18. Eugene Garfield, "Using the Impact Factor," *Current Contents* July 18, no. 29 (1994): 3–5.
19. Tina E. Chrzastowski, "Journal Collection Cost-Effectiveness in an Academic Chemistry Library: results of a Cost/Use Survey at the University of Illinois at Urbana-Champaign," *Collection Management* 14, no. 1–2 (1991): 85–98.
20. Elizabeth Pan, "Journal citation as a predictor of journal usage in libraries," *Collection Management* 2, no. 1 (1978): 29–38.
21. Amy Dykeman, "Faculty Citations: an approach to assessing the impact of diminishing resources on scientific research," *Library Acquisitions Practice & Theory* 18, no. 2 (1994): 137–146.
22. Janet Hughes, "Use of Faculty Publication Lists and ISI Citation Data to Identify a Core List of Journals with Local Importance," *Library Acquisitions: Practice & Theory* 19, no. 4 (1995): 403–413.

23. Biosis, *Biosis Previews* [Web] ([cited July 26 2001]); available from <<http://www.biosis.org>>.
24. Claudia Lascar and Loren D. Mendelsohn, "An Analysis of Journal Use by Structural Biologists with Applications for Journal Collection Development Decisions," *C&RL* 62, no. 5 (2001): 422–433.
25. Suzanne D. Gyeszly, "Electronic or paper journals? Budgetary, collection development, and user satisfaction questions," *Collection Building* 20, no. 1 (2001): 5–10.
26. Michael Rogers, "Cal State Proposes New E-Journal Buying Model," *Library Journal* 124, no. 3 (1999): 107. Update available from: <<http://www.calstate.edu/SEIR/JACC/JACC.shtml>>
27. Jeffrey D. Kushkowsky, Kristin H. Gerhard, and Cynthia Dobson, "A Method of Building Core Journal Lists in Interdisciplinary Subject Areas," *Journal of Documentation* 54, no. 4 (1998): 477–488.
28. Mary W. Lockett, "The Bradford Distribution: A Review of the Literature, 1934–1987," *Library & Information Science Research* 11, no. Jan (1989): 21–36.
29. B.C. Brookes, "Bradford's Law and the Bibliography of Science," *Nature* 224, no. 5223 (1969): 953–956.
30. B.C. Brookes, "Numerical Methods of Bibliographic Analysis," *Library Trends* 22, no. 1 (1973): 18–43.
31. Michael H. Heine, "Bradford ranking conventions and their application to a growing literature," *Journal of Documentation* 54, no. 3 (1998): 303–31.
32. M. Carl Drott and Belver C. Griffith, "An Empirical Examination of Bradford's Law and the Scattering of Scientific Literature," *Journal of the American Society for Information Science* 29, no. 5 (1978): 238–243.
33. William Goffman and Thomas G. Morris, "Bradford's Law and Library Acquisitions," *Nature* 226, no. 5249 (1970): 922–923.
34. B.C. Brookes, "Optimum P% Library of Scientific Periodicals," *Nature* 232, no. 5311 (1971): 458–461.