Notes on Operations

A Simple Method for Producing Core Scientific and Technical Journal Title Lists

Stanley J. Wilder

The objective of this paper is to present a simple method for constructing core journal title lists in scientific and technical (ST) disciplines. This method is the invention of Bensman and is based on the theoretical foundation outlined in Bensman and Wilder (1998). The method requires the creation of a new measure of value called the Estimated Annual Citation Rate (EACR), which is derived from the Journal Citation Reports' total citation variable. EACR allows researchers to compare the relative value of ST journals, and because it is an annual estimate of citations, it can be compared directly to subscription price to produce a measure of cost-effectiveness. The method is described along with an illustrative exercise using journals in physics and chemistry, and the value and cost results are presented.

The work of Bensman and Wilder (1998) on optimizing scientific and technical serials holdings in an inefficient market is first and foremost theoretical in nature and intended to provide a model for understanding the social and economic dynamics of scientific and technical (ST) journal literatures. The essence of their work lies in their assertion that each ST discipline operates within its own social stratification system. These systems are marked by a high degree of consensus on what is important research, which individuals and institutions produce it, and what journals publish it. Bensman and Wilder established that this systemwide consensus exists, that it is measurable, and highly stable over time.

Given that the consensus within ST disciplines includes journal literatures, we may speak of the "value" of individual ST journal titles as an objective, quantifiable attribute. Further, the stability of journal value over time suggests that today's important journals will tend to remain important in the future. This in turn suggests the possibility of developing meaningful "core collection" lists for ST journals. Such lists would be discipline specific, not institution specific, hence they would apply equally well to colleges and universities without regard to size or academic rank.

Simple value rankings for ST journals would have obvious benefits for scientists and university administrators interested in focusing their promotion and tenure efforts and boosting their departmental rankings. For librarians, however, the double-digit inflation over the last decade in ST journal prices creates another, more immediate use for such lists, provided that the ranking method also allows for comparison of value to price. Bensman and Wilder (1998) found, as others have previously (e.g., Barschall 1988), that ST titles high in value are many times more cost-effective than lower value titles. This finding suggests that cost effectiveness data have the potential to give library administrators an important new tool in their struggle to control costs while safeguarding the measurable scientific value of their collections.

In this paper I will present a simple method, invented by Bensman, for developing lists that accomplish both

Stanley J. Wilder (swilder@rcl.lib.rochester.edu) is Assistant Dean for Information Management Services, River Campus Libraries, University of Rochester, Rochester, N.Y.

Manuscript received June 10, 1999; accepted for publication September 10, 1999.

objectives: the creation of value rankings for discipline specific journal titles and the production of cost-effectiveness data for these titles. I will also illustrate the use of this method using physics and chemistry journals from the 1997 Journal Citation Reports (JCR).

Subject Sets, Value Measures, and a New Measure of Journal Value

The method for producing ST core lists relies on a set of basic assumptions in regard to the subject groupings for journal titles and the value measures used to rank them. Both issues have an important bearing on how the eventual core list results should be interpreted. Therefore, before discussing the production of ST core lists, I first address the basic building blocks of subject sets, value measures, and the development of a new measure of journal value.

Subject Sets

The definition of subject sets is of critical importance to the development of core ST journal lists. Subject groupings that are either too broad or too narrow allow outliers to negate significant relationships and produce unusable results. The solution presented in Bensman and Wilder (1998) made use of Library of Congress subject schedules and the Louisiana State University course catalog to create 33 subject groupings for the range of ST disciplines.

But any system for producing subject sets that offers similar refinement could presumably return similar results. The JCR contains such a subject system, one that has the advantage of being ready-made and available in electronic format for easy manipulation. The JCR's system is the result of a complex mapping analysis conducted each year, drawing the journals that cite each other most often into subject clusters (Small and Garfield 1985). This approach gives the JCR subject sets a dynamic quality that allows them to keep up with the pace of change among ST disciplines.

There is no escaping, however, the "fuzzy" nature of sets in library and information science. These sets are "fuzzy" in the sense that overlapping disciplines do not allow for clear boundaries. The fuzziness of subject sets creates an important element of subjectivity in the creation of core title lists, and this subjectivity should be acknowledged and used to establish a good fit with institutional goals. The basic quality and cost relationships of ST journals are stable and predictable but differences in set definition inevitably produce differences in results.

Value Measures

Journal value is fundamentally a human construct. Bensman and Wilder (1998) suggest that faculty perception of the scientific value of journals is an amalgam of the following five elements:

- Subjective judgment as to value or utility
- The social status of the scientists publishing in the journals
- Size in both the physical and temporal aspect
- Subjective comprehensiveness dictating its importance to a wider range of faculty
- Personal advantage of whether the rater could publish his work there

When quantified, faculty perceptions produce journal rankings that are startlingly similar to those derived through analysis of library use (Bensman and Wilder 1998), and to those derived using the JCR's total citation variable (Bensman 1996; Bensman and Wilder 1998). The strength of the correlation between the three measures is fundamental to the assertion that value exists, and that these variables measure it.

Of the three measures—faculty perceptions, library use, and total citation—the last mentioned, total citation, has several advantages. First, it reflects the judgement of the most actively publishing scientists, from virtually the entire universe of scientific endeavor. Total citation is also easily collected and manipulated using the Web version of the JCR. It is by far the least expensive data to collect, and it produces fewer adverse political side effects than faculty surveys. Finally, total citation is attractive because the JCR also includes invaluable data such as journal half-life, ISSN, and detailed subject categories. For these reasons, the method for preparing core title lists presented here is based on total citation. The JCR's impact factor variable was not considered as a potential value measure because it corrects for size, an important component of both journal cost and value.

There are two important problems, however, with using raw total citation figures to derive core title lists. First, the JCR does not track title changes in the manner prescribed by the Anglo-American Cataloguing Rules, hence working with total citation requires the researcher to adjust total citation to account for each title's entire backfile, through title changes, mergers, and divisions. Second, raw total citation data cannot be compared to price data, as total citation covers the entire length of a title's existence, while each journal's price reflects the cost of a single year's publication.

A New Measure of Journal Value

Although Bensman and Wilder (1998) analyzed JCR and total citation, the development of a simple method called for the development of a new and easier to apply measure of journal value. The Estimated Annual Citation Rate (EACR) was created to correct for the deficiencies in raw total citation data and be a measure that would reduce the need to adjust the citation count over the life of the title and enable direct comparison with annual price data. The formula for calculating EACR is as follows:

$$\text{EACR} = \frac{\text{total citation/2}}{\text{half-life}}$$

The vast majority of titles have an established half-life value that is found in the JCR data, but adjustments are required for two sorts of exceptions. First, the JCR does not calculate half-lives greater than 10, hence a further modification is needed to estimate these half-lives:

Citations needed to reach half-life total (CHL)
= (total citation/2) - total citation for period 1988–97

Average Citation Rate for the last 3 years (ACR3)
= total citation for period 1995–97/3

Years needed to half life (YN)
= CHL/ACR3

Estimated half life (EHL)
= YN + 10

Estimated annual citation rate
= (total citation/2) / EHL

This adjustment is much easier to perform using the Web version of the JCR.

Second, the JCR does not calculate half-lives for journals that receive fewer than 100 citations per year, hence these titles are listed in the JCR with no half-life data. These titles should first be checked to determine whether their low citation count was the result of a recent title change. If not, these low-value titles, which are determined by a low number of citations, can be eliminated from the data sets rather than attempting to estimate the half-life of a very small number of citations.

The EACR calculation is analogous to the JCR's impact factor variable in that it produces a citation rating and shows how the journal is used, but it has important differences. It is derived using the JCR's half-life variable, which measures the number of years necessary to reach one half of a journal's total citation figure. Incorporating half-life thus allows EACR to correct for the differing temporal frameworks of ST disciplines (math journals typically have much longer half-lives than physics journals) and for different types of journals (review and applied journals are not cited in the same way as basic research journals). Further, in almost every instance, 50% of citations constitutes a bigger and thus more stable sample than the two years used to derive impact factor. This is particularly important given that a single key article can have an enormous but short-lived impact on a journal's citation patterns (Garfield 1997).

Having created a new variable intended to measure journal quality, it is necessary to validate it statistically. The Spearman rank correlation, a simple, nonparametric statistical test, was selected to establish the degree of correlation between total citation and EACR. To conduct the Spearman tests, the data were ranked in descending order by total citation and EACR. The Spearman rank formula is as follows:

$$R_S = \frac{1 - 6\Sigma d^2}{N(N^2-1)}$$

In the above formula, R_S is the Spearman rank, d is the difference in the ranks of total citation and EACR, and N is the number of pairs of ranks. The results of the Spearman test fall between -1 and 1, with -1 indicating a perfect negative relationship between the two variables, and 1 indicating a perfect positive one. The results of the total citation/EACR Spearman tests for the physics and chemistry data sets are shown in table 1.

All of the correlations are significant at the 0.01 level. Clearly, EACR is very highly correlated with total citation in every subject grouping.

EACR is a useful measure because the data are easy to collect, the result is easy to calculate and there is a direct comparison to annual price. Also, the results are not confounded by an adjustment for size and there is no elimination of titles with fewer than 100 citations annually.

The Method

To construct core journal lists using the EACR variable, one begins with the most recent electronic edition of the JCR to collect the data that go into the

Table 1. Total Citation/EACR Spearman Tests

Physics	0.90	Chemistry	0.93
Physics Applied	0.92	Chemistry Analytical	0.96
Physics Atomic Molecular	0.90	Chemistry Inorganic and nuclear	0.88
Physics Condensed Matter	0.90	Chemistry Medicinal	0.95
Physics Fluids Plasmas	0.92	Chemistry Organic	0.92
Physics Mathematical	0.93	Chemistry Physical	0.85
Physics Nuclear	0.94		
Physics Particles Fields	0.92		

EACR variable. While the CD-ROM version will suffice, the Web version is more flexible and speeds the process of data collection enormously. One then performs a careful analysis of the JCR's subject groupings to arrive at complete coverage of the discipline under consideration. Once a group of subject categories has been identified, one searches the JCR, and then downloads the titles and the variables that go with them into a spreadsheet.

The most time-consuming aspect of this analysis is the collection of price information for the year covered by the JCR data. Ulrich's and vendor catalogs such as those produced by Blackwell and Harrasowitz cover most titles, but some require more diligence. When a price cannot be found despite concerted effort, experience shows the title is likely to have very low total citation counts, and that many of them are noncommercial titles published in Eastern Europe or Asia. These titles can be eliminated from the data set. Once price data have been added to the spreadsheet, one adds formulas to estimate half life where necessary, and then EACR and cost/EACR.

Once the titles in the subject grouping are sorted by EACR in descending order, one confronts the problem of deciding where the "core" list of high value titles ends and the "noncore" begins. Choosing a cutoff point inevitably involves an element of judgement, and users of core lists should understand that there is generally very little to distinguish titles that are just over the cutoff point and those just under it. As a preliminary test, two cutoff points were considered: 75% of the total subject group EACR, and average subject group EACR. The first cutoff borrows from Bensman and Wilder (1998), who developed cancellation/new subscription lists that satisfied 75% of value as measured by faculty score. In part, 75% was chosen for the support lent it by Trueswell's 80-20 rule, and in part because it was not deemed politically realistic to satisfy less than the top 75% of ST value. As a practical matter, however, consensus on the worth of ST journal titles largely disappeared after the top 50% of journals in Bensman and Wilder subject groupings.

The second cutoff considered is the average EACR for each subject group. The advantage of this approach is its conceptual simplicity; it says that core journals are those that rank in the top half of each JCR subject grouping. The effects of these cutoffs will be discussed in relation to the exercise presented below.

An Exercise Using Physics and Chemistry Journals

To illustrate the method, journals in the 1997 JCR's physics and chemistry subject groupings were chosen. These subject areas were chosen because they are the most expensive journals in the U.S. Periodical Price Index (Dingley and Alexander 1998). Table 2 shows the JCR listings of narrower disciplines for physics and chemistry.

It is important to note that these subject sets do not reflect the full range of journal literatures of interest to physicists and chemists. For example, the JCR maintains a biochemistry subject grouping that falls outside the general chemistry group. Further, it is not uncommon for a single title to appear in more than one subject set, and a title that is not core in one subject area may well be core in another.

Having completed the construction of the data sets and sorted them by EACR, we return to the question of where to divide core titles from noncore. Throughout the chemistry and physics sets, the surprising result is that there is not much to choose between the two cutoff points. The average EACR core lists are generally more inclusive, but only by a few titles. Tables 1 and 2 present the core title lists for physics and chemistry. Table 3 presents a summary description of the core/noncore data using the more inclusive measure for each subject group. In every subject group, the resulting rankings echo the Bensman and Wilder finding that value is highly concentrated in a relative few titles in each subject group.

Summary of Highlights from the Exercise

Skew in value: 20% of titles account for 82% of total EACR in the combined physics list, and 22% of titles account for 78% of total EACR in the combined chemistry lists. The skew in value at the top end is equally evident among the subgroupings, where the average EACR of core titles is many times higher than in noncore groups. The skew in value is nonetheless understated in these data for two reasons: first, some low-value titles in the JCR data were excluded as described above. Second, many titles that would fall in these subject groupings are not sufficiently valuable to be included in

Table 2. JCR Disciplines for Physics and Chemistry

Chemistry	Physics
Chemistry Analytical	Physics Applied
Chemistry Inorganic and Nuclear	Physics Atomic Molecular
Chemistry Medicinal	Physics Condensed Matter
Chemistry Organic	Physics Fluids Plasmas
Chemistry Physical	Physics Mathematical
	Physics Nuclear
	Physics Particles Fields

Table 3. Summary Results

Subject Category	No. of Titles Core	Noncore	Avg $/EACR Core	Noncore	Total Cost Core	Noncore	Avg EACR Core	Noncore
Physics	7	48	1.95	12.61	19,650	41,959	3,188	123
Physics Applied	14	44	1.98	16.95	26,255	47,321	1,412	92
Physics Atomic Molecular	6	23	3.85	16.48	26,847	31,001	1,985	143
Physics Condensed Matter	10	35	5.11	23.53	36,432	61,380	2,129	123
Physics Fluids Plasmas	3	15	1.58	22.64	4,262	14,537	1,327	96
Physics Mathematical	7	16	5.89	17.48	17,667	19,464	722	93
Physics Nuclear	4	16	4.70	32.46	25,601	24,247	1,496	132
Physics Particles Fields	4	11	4.35	43.45	25,359	18,394	2,114	121
Total	**55**	**208**			**182,073**	**258,303**		
Chemistry	15	70	1.29	9.93	13,266	43,079	1,553	98
Chemistry Analytical	12	41	4.87	20.85	40,136	50,972	996	110
Chemistry Inorganic and nuclear	8	23	4.19	35.34	24,044	46,174	1,123	92
Chemistry Medicinal	7	19	1.77	10.34	5,911	14,770	603	88
Chemistry Organic	9	23	2.40	9.21	30,591	27,142	1,777	181
Chemistry Physical	20	59	4.12	17.41	51,397	81,994	1,021	120
Total	**71**	**235**			**165,345**	**264,131**		

the JCR to begin with. Were it possible to include the universe of available titles in each subject group, the skew in titles would be still more dramatic.

Subscription cost: In this exercise, core titles cost less than noncore titles. The chemistry core titles cost 60% of the noncore titles, and the physics core titles cost 76% of the noncore titles. Due to the disparity in the number of titles, however, the average cost of a core title is about two times that of a noncore title in chemistry, and about three times as high in physics. The advantage in average cost of the noncore titles is deceptive, however, because these lists contain many foreign association titles that are low in value and even lower in cost.

By the same token, there are some very expensive titles among the core lists. This is to be expected: the exercise is intended to identify the highest value titles in each subject area, and while high cost titles tend also to be low in value, there are exceptions.

Cost/EACR: A more revealing unit of comparison is cost per EACR. As table 3 demonstrates, core titles are many times more cost-effective than noncore titles throughout the subject groupings. The relative cost-effectiveness of the core lists constitute an additional incentive for using value-based core lists as the basis for collection decisions. It is also important because it echoes Bensman and Wilder's (1998) finding that high value titles are cheaper in terms of cost per use.

Conclusion

Core lists should never be used to make collection decisions in a mechanical way, absent judgement and local considerations. In addition, one must always be conscious of the inherent difficulty of set definition among fuzzy ST disciplines, and the difficulty in making fine distinctions between journals that fall just above or below the cutoff point. That said, one is left with the unavoidable fact that when ST journals are properly grouped according to subject, value is highly concentrated among a small number of titles. Further, the high value titles are many times more cost effective than the remaining titles. In this environment, core title lists seem almost to suggest themselves. This is fortunate for librarians and other academics faced with making real-world judgements about quality and resource allocation in ST disciplines.

Works Cited

Barschall, Henry H. 1988. The cost-effectiveness of physics journals. *Physics Today* 41: 56–59.

Bensman, Stephen. 1996. The structure of the library market for scientific journals: The case of chemistry. *Library Resources & Technical Services* 40: 145–70.

Bensman, Stephen J., and Stanley J. Wilder. 1998. Scientific and technical serials holdings optimization: A LSU serials redesign project exercise. *Library Resources & Technical Services* 42: 147–242.

Dingley, Brenda, and Adrian W. Alexander. 1998. U.S. periodical price index for 1998. *American Libraries* 29 no. 5: 82–90.

Garfield, Eugene. 1997. Dispelling a few common myths about journal citation impacts. *The Scientist*, 11, no. 3: 11.

Small, Henry, and Eugene Garfield. 1985. The geography of science: Disciplinary and natural mappings. *Journal of Information Science* 11: 147–59.