PROCEEDINGS OF THE CONTRIBUTED PAPERS SESSION
Biomedical and Life Sciences Division

Special Libraries Association

91st Annual Conference
Philadelphia, Pennsylvania
June 14, 2000

What journals, if any, should STILL be PUBLISHED?

David Goodman
Princeton University Biology Library

Abstract
This paper examines the actual use of individual articles from print journals. On the basis of the measured use and citation data for biology journals at Princeton and at the University of Wisconsin, it can be determined that representative middle-level biochemistry journals publish articles many or most of which will be read less than one hundred times, ever, by anyone. Publishing such articles as part of journals may have been necessary when there was no more affordable and flexible medium of distribution, but now that it is not necessary, its appropriateness is questionable. An article based scheme would seem a better response to the wide variation in use of published scientific papers.

I suggest that the publication of scientific articles be based not on journals, but on a suitable internet-based repository, with the authors and their organizations responsible for preparation. Journals, print and electronic, would be media for republishing significant articles, and would serve functions of quality control and aggregation. To bring this about, I suggest that universities and agencies consider for funding or personnel decisions only scholarly articles that have been first published by being made available without access charge from a suitably indexed and permanently maintained publicly accessible repository.

Introduction

Electronic or Print?

I am not discussing here the relative merits of electronic and paper journals. I will take it for granted that the electronic medium is preferable for the publication of most academic periodicals because of its advantages in distribution, storage, searching, linking, and incorporation of non-textual material. I recognize the greater ease of printed material for reading, and possibly for scanning and annotating--this is, however, an artifact of present technology; its amelioration merely requires the use of currently available printing facilities as needed, and the continued production of dual print editions of those few titles more conveniently printed centrally.

Journals or Articles?

The question here is to what extent publication of scientific papers in journals, even electronic journals, is necessary or cost-effective, as compared to an article-based mode of production and distribution. Are there any key virtues in the journal format that require its continuation?

The basic nature of the journal system is that the papers are divided among a large number of

different publications, each distributed at fixed intervals in portions containing several articles. With print, this clearly facilitates physical distribution, and greatly improves the storage of the individual parts, but it is obvious that these virtues are no longer relevant electronically. An advantage relevant for both electronic and print is that publication at intervals provides convenient units for current awareness, and provides a psychologically relevant unit for reading. With electronic publication, there are obviously other ways of providing such groupings that will be better than publication in arbitrary sections.

The division of print publications into individual journals facilitates references, conveniently distributes the publishing and editing responsibilities, and, at least potentially, groups related articles and articles of like quality together. With electronic publication, the provision of easily remembered citations is no longer relevant--digital identifiers do even better. The division of labor is equally well attained by any other device: by discipline, institution, or individual interest. The division by subjects and by quality is obviously irrelevant for the most common use of journals, the retrieval of a specific article referred to by another article or index. It can only be relevant for current awareness or browsing. For those purposes, there is a point in marking articles by quality (or by general interest) and by subject. Most people want to read everything in their exact specialty, and also major articles of importance in their general field, and of course the very most important articles in their broad area of interest. In principle, this can be done by a hierarchy of titles, but even when a subject has only one or two leading titles, a considerable number of equally important articles occurs in titles of a slightly lesser rank. Below that level, articles will be found in a broad scattering of journals. Even if it were possible to reorganize this rationally, the interdisciplinary nature of much research and the variety of individual interest would make a single hierarchy impractical. Thus, a much more efficient way of marking would be to mark the individual articles, allowing for their arrangement into any desirable grouping.

Since, then, it is not necessary to publish according to the present system, are there forces that should drive change to publication as individual articles? One force is the preference of authors to publish as soon as possible. Already, electronic versions of many titles now publish the articles as available, not waiting for the complete "issues." Even this delay is regarded by authors in an increasing number of fields as undesirable, and they release their material in unedited electronic form as individual items on servers. The desire of the readers to see the material as soon as possible has caused the use of these unedited articles in some pioneering subjects to be many times that of the same articles in journal form. The other force is economic. The dispersion of titles makes the provision of complete collections on any one subject financially impractical except for the very largest libraries. The very low use of many--or even most--articles makes the expense of their preparation in the current fashion inappropriate.

Use of Individual Articles

I examine in this paper the actual use of individual articles from journals, based on the available data from print journals. In the near future, it will be possible to make much more precise and direct measurements using the extensive data that will be available from electronic journals. For now, it is necessary to make use of measurements using printed journals, and realize that the many uncertain factors make it possible to construct only reasonable inferences. These approximations will, however, serve as a baseline for the future observations.

The strategy here is to:

1. measure the overall use of print journals
2. correlate the use of the print journals with citation frequencies for the journals
3. measure the citation frequencies of individual articles
4. infer the use of the individual articles.

Procedure

Library Environment

This paper reports the use of titles in the Princeton Biology Library, a specialized library in a relatively small but very active research and teaching university. This library has the mission of providing an extensive research and teaching collection specific to areas of biology currently of interest in the University. In the absence of schools of medicine, agriculture, and environmental science, no attempt is made to provide complete coverage of the entire field of life sciences; provision of material in peripheral subjects has always relied upon the collections of more comprehensive institutions.

This library serves primarily the medium-sized Department of Molecular Biology, with 43 research groups, and the smaller Department of Ecology and Evolution, with 17 research groups. There are about 550 primary users, including the faculty, post-doctoral fellows, graduate students, and senior and junior undergraduates. The library has (for 1999) 590 active periodicals, with access to about one-fourth of them online (as of March 2000). Most neurobiology titles are shelved in the Psychology Library, and are not included in this analysis.

These measurements have been compared with the measurements of the use of journals at the University of Wisconsin at Madison, a very large research university with extensive programs covering the complete range of pure and applied biological sciences, including many professional schools. Their data are available as part of an extensive study carried out there over several years. Wisconsin has 34,560 students as compared to Princeton's 6400, and 1957 faculty as compared to Princeton's 760. At Wisconsin, those working in the biological sciences form a greater part of the population than at Princeton; the primary users of the group of biological sciences libraries at Wisconsin can be estimated to be 10 times that at Princeton.

Journal Reshelving

Reshelving of bound journals was measured at Princeton from Feb. 1999 through Jan. 2000, a period including all parts of the academic year. Reshelving of unbound issues was measured beginning Dec. 1999 and extrapolated to a full year. (The details will be reported elsewhere.) There are the obvious undercounting errors: Users may return items to the shelves directly (this is expected to be much more likely for unbound items and for those not photocopied); it is probably the principal source of error, and cannot be completely avoided in an open-shelf library. Users may read or copy more than one item per issue or bound volume; users may read or copy from an issue or volume that has not yet been reshelved. As we bind in very small units, neither is likely to be important for bound volumes except for the most recent few years of the very few most used titles. For unbound issues, it should also only affect only the very few most heavily browsed titles.

In analyzing the Wisconsin data, I made use of the 1998 tables as available on the web. Their data is from the entire group of libraries; they are used as found. The compilers guess the true values to be at least as great as the measured values, but less than twice as much.

Journal photocopying

Since journals must be photocopied within the library at Princeton, the counts on the library photocopiers measure the total photocopy use. To convert these counts to the total number of articles copied, it is necessary to know the percentage of total copying that is for journals, and the average length of an item copied. We estimate the percentage of copying that is due to journals as 90%. (The rest is primarily book chapters, with a small amount of class notes and exercises.) For interlibrary loan photocopying, where records are kept of both number of items and number of pages, the average number of pages per item is 10. The average length of articles copied by our own users is probably less, because of the small amount of interlibrary loan for many of the most

heavily used titles, most of which are characterized by short articles; we therefore assume an average length of 8.

Citations

To measure the citations to each of the different journals, we used the 1990-plus segment of the Dialog version of Science Citation Index. We examined the number of citations made from all authors whose address was Princeton University (or University of Wisconsin at Madison) to all articles in all the years of each journal, and divided by 10 to get the citations per year average.

To examine the citations to particular articles we used the ISI Web of Science version of Science Citation Index using the "Cited Reference Search." For a limited number of journals, all in the field of biochemistry, we searched for the citations to the individual 1990 items published in each of these separate journals. We then counted the number of citations to each of 50 articles from each journal using an appropriate sample. This method does not detect those items with zero citations. We controlled for this in a few cases: using the "General Search," we searched for the articles published in a particular journal in 1990, and examined the number of citations for each. Technical limitations restrict this to titles publishing fewer than 300 items in that year.

Data and Analysis

Overall Use of Print Journals

The total use for a print journal is the sum of the library use and the use out of the library; this additional use will be due to personal copies, to re-copying of articles, and to use of reprints. For the present purposes, "major" journals can be defined as those for which personal copies are available, and "minor" one as those for which they are not. This will obviously vary locally. At the Princeton Biology Library during one entire year, the photocopy count was 370,523 pages. Making the adjustments described in "Procedure," this corresponds to the copying of 41,684 articles. Our measured shelf use in this period was 26,955 bound reshelvings and 5659 unbound, giving 32,624 observed uses.

To find the true total in-library use requires estimating four constants: the relative amount of undercounting of bound and unbound, for photocopy and for print. Obviously there are many possible values. An analysis (to be reported in detail elsewhere) indicates that at our library the plausible values for the total use is between 60,000 and 70,000. The median is 65,000, which is 1.97 times the unadjusted reshelving value. The best factors to use in adjusting the figures for a particular journal are 1.61 times the bound reshelvings plus 3.76 times the unbound reshelvings.

We have also used the figures from the University of Wisconsin study. The uncorrected use shown by their figures is generally in the range of 10 times our use; this was at first disquieting, but is what should be expected in a university with its size and programs.

Relative Use of Print and Electronic

For journals available in electronic form, added use is expected and observed. It is yet unknown to what extent the electronic use replaces library use, and to what extent it adds to it. Such data as we have is for titles available in both formats, and for major titles only, where personal print copies have been available. It generally shows additional electronic use of about 10 times the uncorrected library use. Correcting for measurement error as above, the use is about 5 times as great. This increase represents both replacement of the out-of-library use of print copies and the increased use due to the greater convenience. If we assume an equal contribution, each value represents twice the in-library use. Thus we have a very rough estimate for the "major" titles that the increased use of electronic journals over their print counterparts is about a factor of 2, not the 10 that the raw data suggests.

Use and Citation
The data for Princeton shows, as expected, that the use of journals roughly parallels their citation frequency. The values of uses/year and citations/year is shown in Figure 1, for those titles for which uses/year is greater than 2, and citations/year is greater than 0.2. Titles that have been cancelled for more than one year and titles that have begun publication within the last 5 years are not included in this table.

The use/citation ratio for a journal at Princeton is the ratio of the adjusted uses per year at Princeton, to the average number of citations per year of that journal by all Princeton authors. The average value is 21; this corresponds to the total of all citations and all uses, and is influenced greatly by the most heavily used journals. The best fit linear estimate is 13.46, and is the more appropriate factor for an average journal. The individual ratios show interesting variations. Some of the variations reflect peculiarities of the local academic program, but most probably simply reflect the small sample size, especially for unbound reshelvings.

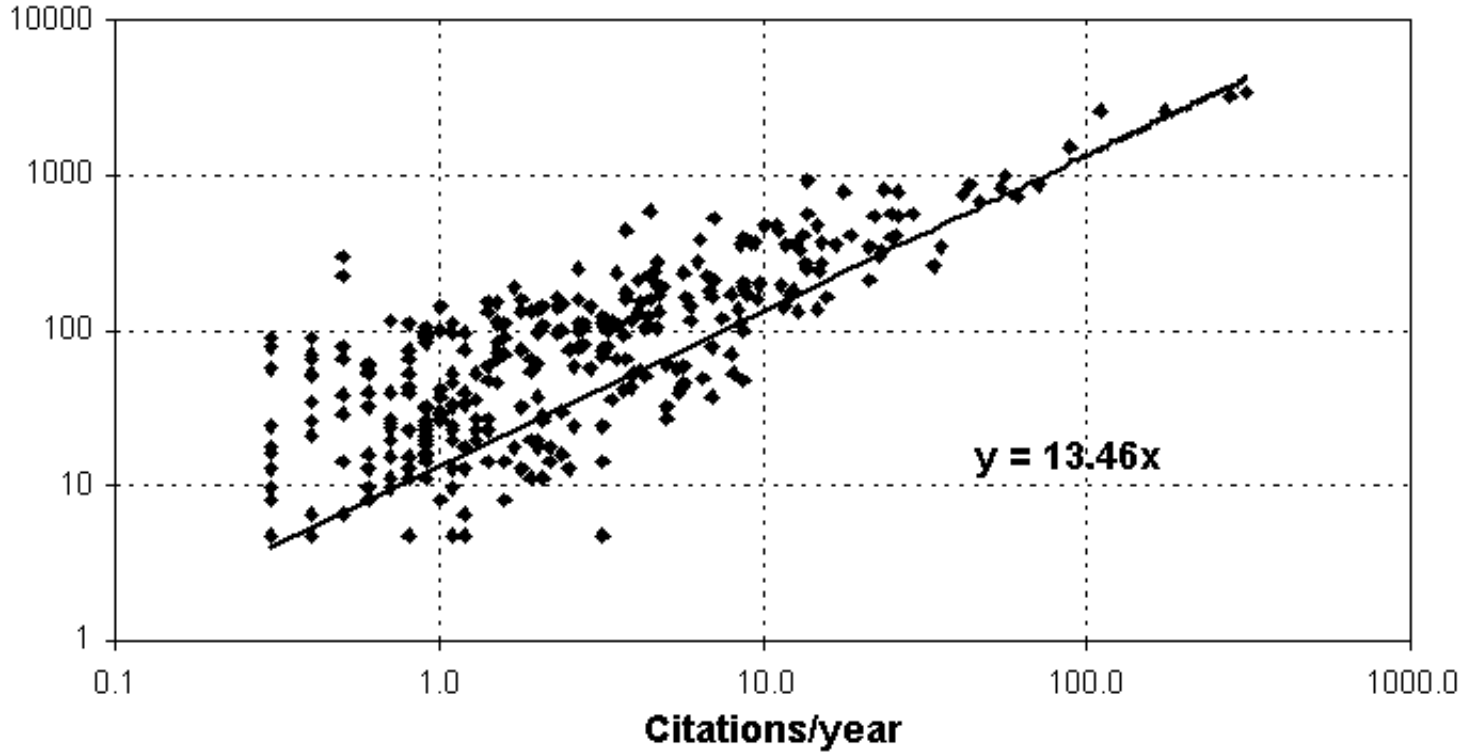

Figure 1. Use and Citation of Journals

We can examine the use to citation ratios for specific journals; this paper considers six specific biochemistry titles. The use and citation data and this ratio for these journals at Princeton and Wisconsin are given in Table 1 below.

	BBRC	BBA	Arch	BC-HS	BCCB	IJBC
Princeton citations since 1990	213	337	85	10	16	8
Princeton uses per yr.	214	263	111	53	69	18
Uses per year / citations per yr. at	10.0	7.8	13.1	53	43.3	22.5

Princeton						
Unbound uses per year at Princeton	4	11	0	0	2	
Wisconsin citations since 1990	2280	2637	1376	70	75	106
Uses per year at Wisconsin (uncorrected)	1968	2677	989	99	52	79
Uses per year / citations per yr. at Wisconsin	11.6	9.8	13.9	7.1	14.4	13.4

Table 1. Citations and Use of Biochemistry Journals

The total use to citation ratios range from 7.1 to 53 for the individual titles. The ratio for the six titles combined is 10.9 at Princeton, and 11.2 at Wisconsin. Considering the differences between the universities, and the numerous corrections, this is remarkably close, and implies that the numerical relationship may be approximately true for the academic world as a whole. Therefore if a journal has 1000 total citations to its entire run from anywhere in the world, it would be expected to have 1346 uses per year in the whole world. If it is valid to consider this group of titles separately, then for these six biochemistry journals, 1000 total citations corresponds to 1125 uses; this value will be used in the further discussion.

Use and Citations of Articles

A journal is the sum of its articles. The use of the journal is the sum of the uses of each of its separate articles, and the citations to the journal are the sum of the citations to each of the individual articles. We can therefore approximate the use of the individual articles (which we cannot yet directly measure) from their citation frequencies, which can be determined from Science Citation Index. The application of these numbers to the total world-wide use and citations of the biochemistry journals discussed is shown in Table 2 below.

	BBRC	BBA	Arch	BC-HS	BCCB	IJBC
Articles / yr. (1990)	1601	2023	513	208	217	208
Impact factor (1992)	3.583	2.610	2.435	1.839	1.473	1.155
Rank (1992)	32/157	47/157	48/157	68/157	86/157	107/157
Total citations worldwide (1998)	58,706	68,402	22,030	2819	1795	2397
Total citations to one year, ever	53,369	62,184	20,027	2563	1632	2179
Total citations / average article	33.33	30.73	39.04	12.32	7.52	10.47
Total uses / average article	375	3464	439	139	85	118

Table 2. World-wide Use of Biochemistry Journals

From each of the biochemical titles being discussed, we examined 50 randomly chosen articles published in 1990, as described in "Procedure." The data are show in Figure 2 below. It will be observed that for each of the journals, most of the citations are represented by only a few articles, while many of the articles receive only a very few citations, consonant with the usual expectations. For each title, one-fifth of the articles account for one-half or more of the total citations. The figure above includes only those articles that had one or more citations; we also determined for two titles the percentages of the articles that had no citations at all. We can calculate the percentage of the articles in each journal destined to have less than 50 and 100 uses, worldwide ever; the results are shown in Table 3 below.

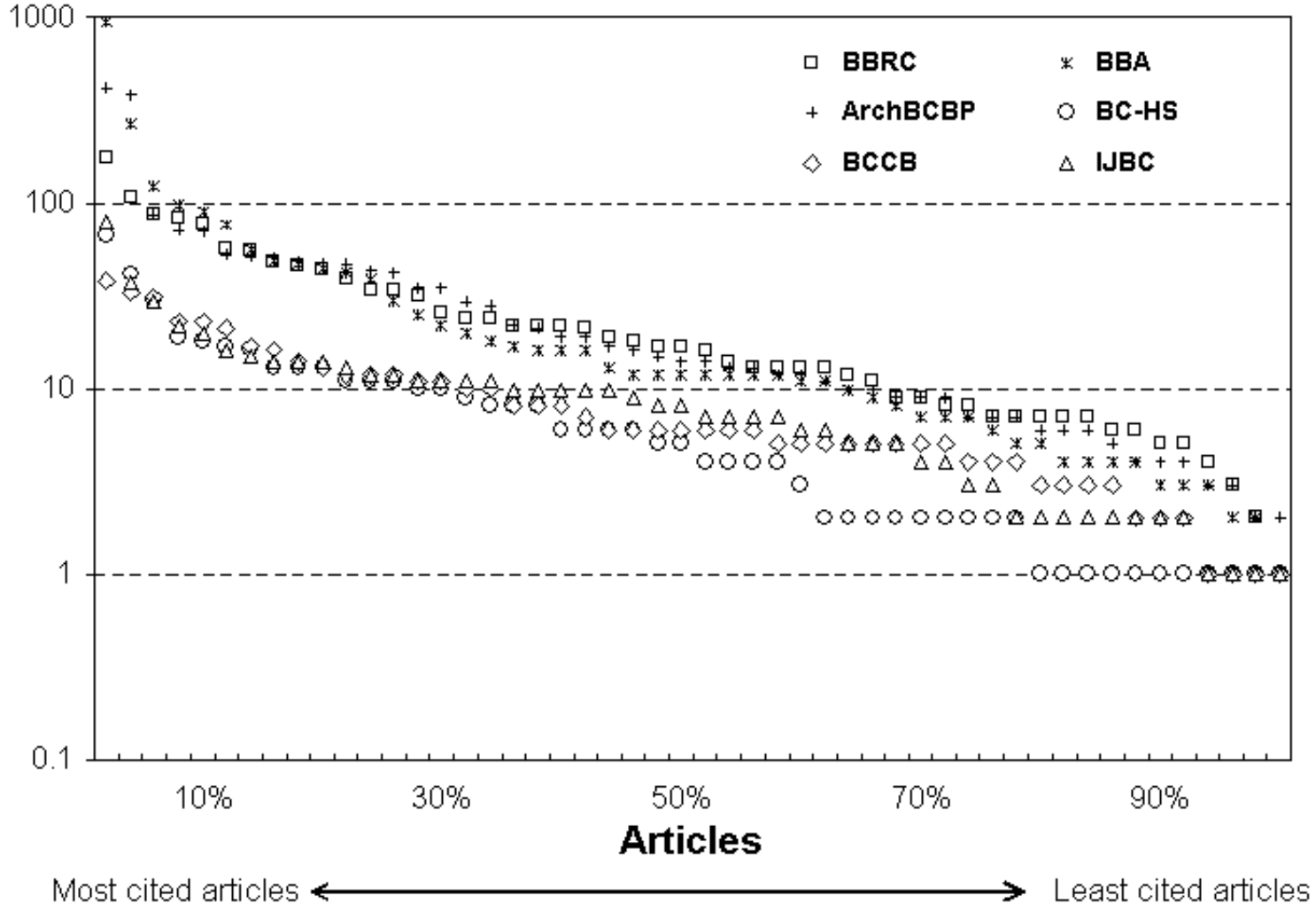

Figure 2. Citation and Use of Individual Articles

	BBRC.	BBA	Arch	BC-HS	BCCB	IJBC
Articles / yr. (1990)	1601	2023	513	208	217	208
Percent with no citations					9%	11%
Percent with 4 or fewer citations = fewer than 50 uses in 10 yrs.	8%	20%	14%	50%	35%	39%

Percent with 9 or fewer citations = fewer than 100 uses in 10 yrs.	34%	36%	34%	70%	92%	92%

Table 3. Use per individual article

The interpretation of these numbers depends upon economic factors. On the one hand, surely use of even 40 times is real use, even though it will be spread out over many years and many universities. Even a certain proportion of articles that are never cited can be seen as an acceptable consequence of the unpredictability of scientific progress. On the other hand, clearly the distribution of the hundreds of thousands of relartively low use articles produced each year should be done in the least expensive suitable fashion.

Discussion

Journal quality

The journals being discussed are respectable titles. They represent average quality journals, as can be seen in Table 2 from their citation position in the middle of the impact factor range for the subject. The articles in them are used, and are clearly worth publishing for the scientific community. The question of how the articles should be published depends upon the availability and costs of alternatives. In the print era, there was no other practical way to publish these articles except in journals if the data in them were to be distributed. The increasing expenses of doing this were met by increasingly limiting the distribution of all but the top tier of journals to the decreasing number of institutions that could afford them. Now the alternatives include various possible ways of electronic article-based distribution.

Since these are general journals, no one reads the journal systematically to follow a specific subject, because there will be only a very few articles of relevance. Since these are not among the very highest quality journals, no one reads them as wholes to follow the most important developments in biochemistry . Neither would anyone even scan them for articles of personal interest, because their proportion of articles of high interest and quality is so low. (This is confirmed by our data for the use of unbound issues, shown in Table 1, which shows extremely low use for any of these titles.) Articles in these journals are accessed through the indexes and through citations. The articles in them of value to an individual's purposes are identified by indexing services and citation networks. They would be found and used as readily if they were available by themselves.

Article quality

For the journals discussed a small fraction of the articles receives high use. (In each case, one-fifth of the articles account for one-half or more of the total citations.) The six titles discussed seem to fall into clearly different groups in terms of use; the more used three are used much more than the others. It should be emphasized that these titles are shown as illustrative only. The data are given for one time point only, and do not reflect changes in their quality and use of the titles. Only a few of the journals in only one field of study are analyzed.

Biochemistry is a field with a very high number of citations per article and a great many interested workers. The figures in most fields of study would be much lower in absolute terms. In fields with a very small number of workers, the value of the work may nonetheless be very great, and it is not a correct employment of figures such as these to judge the social or scientific worth of the subject areas. It would seem correct, however, to say that the lower the number of interested users, the

more attractive is non-conventional publication.

Currently, the desirability of the good articles in a journal pays for the distribution of the others. If articles were published individually, it would still be possible to devise schemes for distributing the cost of publication as individual articles, if they will all in general be used enough. Publishing the bulk of the articles that appear in the journals in a suitable database, with possible republication in overlay journals customized to personal interests would seem the most appropriate and consistent way of publication. In particular, it will be a much more direct way of making them available outside a small number of major institutions, without the cumbersome methods of conventional document delivery.

There are journals with much lower use. It would seem obvious that if the middle-level journals analyzed here are not necessarily worth publishing as such, then certainly the lower use ones are not. Some of those journals are simply low quality general journals; some are extremely specialized titles. It is sometimes argued that even very low use specialized titles within large subject fields are worthwhile, because they concentrate the articles for the benefit of the possibly small number of workers on that specific topic. But such specialized titles within a large active field will contain only a small proportion of the articles in their particular subject niches, because most authors will choose to publish articles of high quality in the more general titles to reach a wider readership. Individual publication would seem appropriate here also--overlay journals or personal awareness selections are a much more direct route to access to specialized subjects.

There are also journals with much higher use. Some of them are so large that people do not usually read or even scan each article, but typically read several articles in each issue. Some are small enough, and important enough, that many users will read or scan most of the articles. Each field has one or two such titles; it may well be viable to continue publishing them, and probably even in print format as well as electronic.

Practical article-based schemes are necessarily electronic; similar ideas were conceived and even attempted in the paper era, but faced the obvious logistic difficulties. Article-based electronic publication does not necessarily exclude journals. Journals could remain as media for republishing significant articles, electronically or in print. These journals could be organized by quality, by subject, by local interest, or any other factor. They could, if desired, incorporate peer review or alternative mechanisms for evaluating and maintaining quality. Many good journals might choose to replicate their current practices and policy. Even now, there are journals for which people are very willing to pay for paper personal subscriptions although the entire scientific content is freely available on the web. There would be a market for appropriate selections and cumulations; the most successful new journals have been review journals. Would the economic role of journal publishers be less under this scheme? It would depend on whether they published material that enough people would want to buy.

Objections to article-based schemes

Quality control is currently carried out by peer review based upon conventional journals. I will not discuss here whether peer review is necessary, effective, or desirable; what is clear is that there are now many ways to implement it. The cost of conventional publication is not justified by the peer review because almost none of the cost goes towards this; the reviewers and the editors are not usually compensated in money.

The operation of the tenure system is a more intractable problem. As long as elite journals exist, there will remain administrative pressure, even at teaching institutions, to publish in them. I am cynical enough to ascribe this to the desire and possibly the need of academic administrators to

tenure as few faculty as possible. Change will be up to the leading institutions, which, in principle, judge work by its actual quality, not merely where it is published.

Other objections, such as reliability, long term access, archival preservation, and network problems, are common to all electronic formats, and need not be discussed here.

Commercial, Non-commercial, or Informal?
Article-based methods are an improvement over journals for most scientific material, but they do not by themselves necessarily solve the basic problem of expense. The costs of publishing on an article-by-article basis are not necessarily lower than publishing journals. An electronic-only journal has no significant cost of distribution; neither do its articles. The cost of preparation in each case is determined by the technical quality desired for the material. It may of course be preferable to maintain a high quality for all formal publications, but a good standard of presentation can be obtained by a variety of desktop methods. The cost of overhead depends in each case on the organizational structure. This might be every bit as expensive as at present; however, article-based publication does permit some drastically less expensive systems.

The cost of publication by societies or other non-commercial organizations is typically on the order of half that of commercial enterprises.. Some of the difference is due to the necessary profit in a commercial enterprise, but many organizationally sponsored titles also try to return a profit to the parent organization. Some of it is due to higher overhead costs. Some of it is related to the typically higher quality and consequent larger readership of the non-commercial titles. Much lower costs seem only achievable by informal publication; the cost of running an e-print server which relies on the authors for formatting can be very low. From the available examples, attaining very low costs requires such a method.

Proposal
I suggest that the publication of scientific articles be organized as suitable internet-based repositories, with the authors and their organizations responsible for preparation and quality. Journals, print or electronic, would continue as media for republishing significant articles. These publications could be undertaken by commercial publishers, but also by individuals, academic departments, institutions, professional organizations, or professional editors. The cost of producing them as virtual journals would be so low (at the most basic level, such a publication need be no more than a list of links) that there would be potential for great customization and experimentation, and even for commercial entrepreneurship.

To bring this about, I suggest reversing the current rule, and establishing a standard that no reputable journal consider an article unless it has previously been made effectively available without access charge on a appropriate suitably indexed and permanently maintained publicly accessible server. An immediately practical change to accomplish this, without impairing in any way the ability to publish in any manner an individual might choose, is for universities and funding bodies to consider authors' publications in appointment, promotion, funding, or tenure decisions only if they have been first or simultaneously published in that manner.

RETURN TO CONTRIBUTED PAPERS CONTENTS

RETURN TO BIOMEDICAL AND LIFE SCIENCES DIVISION WWW PAGE