Bibliometric Models for Management of an Information Store. II. Use as a Function of Age of Material

Ralph H. Parker
School of Library and Informational Science, University of Missouri-Columbia, Columbia. MO 65211

The distribution of use of items in a large information store is examined. The findings suggest that a simple exponential distribution is inadequate and that a multifactor exponential model describes the process of obsolescence more precisely.

Introduction

It has long been held that the utility of items in an information store tends to decline with age. The validity of this belief has been tested by researchers in different millieux with varying assessments of the aging process. This study, which evolved from a project to determine the possibility of limiting the size of a large university library [1], is concerned with creation of a close-fitting bibliometric model of aging in a large heterogeneous information store.

If the size of an information store is to be limited, two conditions must exist: (1) Some items are more useful than others, and the differential utility between the most productive and least productive must be sufficiently great to permit significant selection and retention decisions. This condition has been discussed in the preceding article [2]. (2) Use of items in an information store declines with age in such a manner and to such a degree that the decline can offset the forces of growth. The temporal distribution of this decline is examined in some detail in this study.

Previous Studies

Numerous studies over the past four decades have resulted in the consensus that the decline in use of materials follows a negative exponential curve. Most of these studies have been based on analysis of citations in scholarly publications or of the contents of other selective bibliographies. Only a few have dealt with the use of large heterogeneous information stores such as libraries.

The negative exponential distribution of utility by age was first suggested in 1944 by Gosnell [3] in his study of the titles in Shaw's *List of Books for the College Library* [4] and of Mohrhardt's *List of Books in the Junior College Library* [5].

In 1960, Burton and Kebler [6] discussed the half-life of scientific literature based on citation analysis. Other significant studies include Cole's [7] study of use of journals versus age and Brookes' [8] analysis of obsolescence of journals in a special library. Brookes calls attention to the sampling errors which can distort obsolescence rates; he also notes that his use of citations was the result of accessibility of data, but that library user demands would have been preferable.

Sandison [9] in 1971 observed that perhaps the negative exponential distribution was an inadequate description of the rate of decline over the entire time span and also pointed out that the growth of literature affects the apparent obsolescence rates. Line and Sandison [10], in a review article in 1974, pointed out weaknesses in studies of obsolescence, calling attention to both synchronous and diachronous studies. Griffith and associates [11] have pointed out that the aging process operates most rapidly among journals which support a research front, but the aging depends not only on the materials but also on the user.

There have been fewer studies of obsolescence based on observed use of materials in libraries. Gosnell suggested, as a sort of footnote, that conclusions might also apply to college libraries, citing Stieg's study of circulation records at Hamilton College. The Stieg article [12] had not characterized the age distribution, and unfortunately it contained no raw data. From the percentages he gave, reconstruction of the rate of decline in use yields an annual rate of 0.056 or 5.6%.

In 1961, Kilgour [13] analyzed the circulation of the Yale Medical Library. His study does not give the figures

Received May 1980; revised April 1981; accepted November 13, 1981

for circulation by date of publication, but by reconstruction from a graph the rate of decline can be calculated as approximately 0.13 or 13%. Because of the small numbers for ages above 15 years, it is difficult to differentiate random deviations from trends.

Both the Stieg and Kilgour studies may be described as synchronous studies, since they examine use, during a single period of time, of materials published through longer periods of time. A weakness of this type of study, as pointed out by Line and Sandison [10] and others, is that over the periods of time studied, i.e., the 20th century, there has been a continuing growth of the literature. Library acquisitions have followed that trend, so that the earlier periods (older materials) are underrepresented as compared with the later (newer) periods.

A second type of study attempts to follow the use of books of a certain date through a period of time. This diachronous approach has certain advantages, including elimination of the underrepresentation of access to older materials, but it introduces other more serious disadvantages; it is expensive and time consuming and it involves uncontrollable changes in users and conditions. Fussler and Simon [14] attempted to combine aspects of both types of study by using two consecutive five-year periods of observation.

Morse [15], in his study of MIT libraries, takes a somewhat different approach to the effects of aging on use of materials. He suggests that use is a Markov process, that the use in the next year can best be determined on the basis of use during the immediately preceding year. He suggests that the use R in year $t + 1$ can be estimated by adding a constant α to the product of a second constant β and the use R_t in year t, thus $R_{(t+1)} = \alpha + \beta R_{(t)}$. Both α and β may vary from subject to subject.

Hindle [16] has applied the Morse technique to both a university library and a public library in England, with constant values somewhat different from Morse's.

Rouse and Rouse [17] have studied the use of 10,485 monographs requested by interlibrary loan. They made corrections to actual demand based on the number of titles available from American publishers by year. They assumed two factors, similar to Morse, one constant the other affected by time.

Method

None of the studies of the library use, upon which the hypothesis of exponential distribution of decline in use is based, appears to have undertaken a *close* fitting of the theoretical frequencies to observed data. The researchers or others analyzing their data have shown that the curves are generally exponential. The small samples yield such low frequencies for older materials that distinguishing between a trend and sampling error is difficult. Hence, if there were a different rate of decay for older material, it might not be detected; and if there is underrepresentation of older material, the phenomenon could mask or exaggerate any reduction in rate of decay which actually exists.

This author undertook, as part of a larger study [1], to analyze the distribution of use by the length of time the material had been in the library, using a sample of almost 400,000 uses. Since the accession number of each book was included in the computer circulation record, it was possible to approximate the number of volumes in the library acquired in each time period observed without having to determine the publication date of each item.

To correct for generally increasing accessions, an index of use was computed for each time period. The current year and each of the preceding three years were calculated separately, years four through eight and all further years were grouped in five-year intervals. The index was computed as follows:

$$I_i = C_i/V_iV/C, \tag{1}$$

where I_i is the index for any interval, C_i is the circulation of items acquired during that interval, V_i is the number of volumes acquired during that interval, C is the total circulation in sample, and V is the total volumes in the library. The last term of the equation (V/C) is used to standardize the results by eliminating the effects of the size of the circulation sample.

The resulting index of use is given in Table 1. It ranges from 3.69 for material acquired in the current year (age 0) down to 0.16 for material acquired 71 (69–73) years earlier. The results of sampling error are apparent in some of the older time periods.

These indices are used to prepare adjusted cumulative

TABLE 1. Volumes in library and circulation by age.

Age[a]	Volumes in Library	Circulation	Index
0	47,080	39,355	3.69
1	29,456	21,010	3.15
2	68,768	38,984	2.50
3	73,665	29,351	1.77
4-8	351,802	130,281	1.63
9-13	265,367	50,762	0.84
14-18	181,412	19,448	0.47
19-23	105,895	12,573	0.52
24-28	94,951	12,029	0.56
29-33	86,938	8,555	0.43
34-38	88,889	7,199	0.36
39-43	46,013	3,010	0.29
44-48	56,104	3,801	0.30
49-54	45,893	2,020	0.19
55-58	37,320	1,690	0.20
59-63	37,103	1,951	0.23
64-68	32,109	1,316	0.18
69-73	19,465	709	0.16
74+	33,999	1,758	0.16
Total	1,702,220	285,989	

[a]Years in library, volumes acquired in fiscal year 1972–1973 are shown as 0.

TABLE 2. Cumulative percentage distribution of use by age of material.

Years in Library	Cumulative Percent	
	Observed	Adjusted[a]
0	100.00	100.00
1	89.81	91.74
2	84.37	84.69
3	74.27	79.07
4	66.62	75.13
9	32.88	56.88
14	19.73	47.37
19	14.69	42.11
24	11.39	36.29
29	8.28	30.02
34	6.07	25.21
39	4.21	21.08
44	3.43	17.93
49	2.45	14.57
54	1.93	12.44
59	1.49	10.2
64	0.98	7.63
69	0.64	5.62
74	0.46	3.83

[a]Corrected for underrepresentation of older materials.

distribution of use by period. Table 2 shows both the observed and the adjusted distributions. The effects of underrepresentation are obvious when graphically displayed in Figure 1.

If the rate of decay of use is exponential when the index values are plotted on a semilogarithmic scale, the curve will be linear. Figure 2 shows clearly that the experimental (adjusted) data do not conform to the exponential model.

To test for this linearity, the expected values for any particular year may be completed by using the equation of the exponential distribution:

$$I_t = I_0 a e^{-at}, \qquad (2)$$

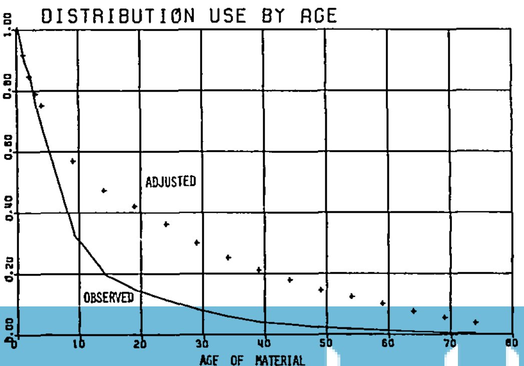

FIGURE 1. Cumulative distribution of use by age

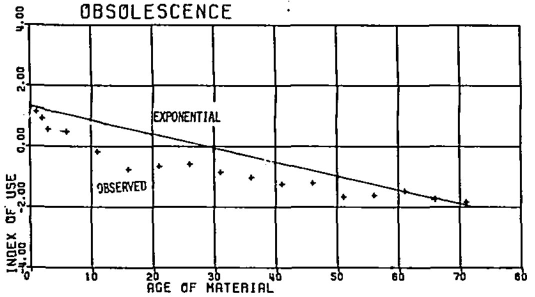

FIGURE 2. Index of use by age of material.

where I_t is the index of use at time t, I_0 is the index of use at time of acquisition, and a is the annual rate of decay.

The best fit of the exponential line to the empirical curve was by the least-squares difference method. This line (Fig. 2) has a slope of -0.045 and cuts the empirical curve at year six (midpoint of the 4–8-year group) and near year 51.

Two-Factor Model

Following the findings of Morse [15], Sandison [10], and others, we shall hypothesize that more than one factor is involved in obsolescence of material in an information store.* This phenomenon may be analyzed without the necessity of assigning roles to each factor, but merely naming them for convenience. Let us consider the situation of two factors, each operating simultaneously with different exponential rates of decay. This differs from Morse [15] and Rouse and Rouse [17] in that they assumed a second factor unaffected by time, but Rouse and Rouse did suggest that a double exponential might better fit some subjects.

The equation for such a model would consist of the sum of the negative exponential curves of the two factors. We shall call the first the *ephemeral* factor and label it A, the second the *residual* factor labeled B. Thus, the equation becomes

$$I_t = I_0(Ae^{-at} + Be^{-bt}), \qquad (3)$$

where A and B are the weights of the factors A and B, and a and b are the annual rates of decay of A and B. Each factor may have any weight from 0 to 1 such that $A + B = 1$.

Using an iterative computer program which assigned

*Although we commonly attribute obsolescence to characteristics of the information source or of information itself, it is more appropriate to consider obsolescence as a social phenomenon. Thus, it occurs as a result of attitudes of those in the universe of users of information rather than as a result of characteristics inherent to the sources. An individual item may be in great demand in one environment and not in another, e.g., in a chemical library as compared with a general public library.

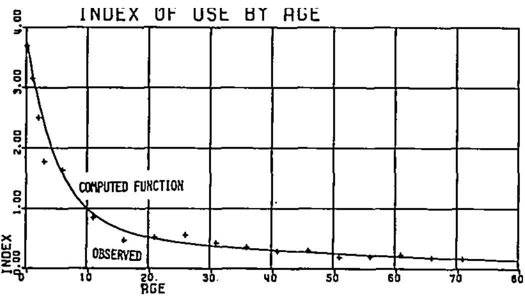

FIGURE 3. Index of use by age.

successive values to A, B, a, b, the values of I_t were computed and compared with the adjusted values in Table 1.

The sum of the squares of the deviations was computed for each set of values used. The computation continued with newly assigned values until the sum of the squared deviations could not be reduced. These derived values are $A = 0.813$, $B = 0.187$, $a = 0.18$, $b = 0.02$. The results of the computation of indices using this composite two-factor model are shown in Figure 3, which also shows the adjusted observations from Table 1. A semilogarithmic plot of the same computation shows that contribution of each factor at any point in time (Fig. 4).

The original study, by the author, from which the empirical data are derived, did not analyze use by subject. To check on subject differences, as well as to make an overall comparison with another set of data, computations based on an unpublished study of the UMC library in 1951 are shown for selected subjects in Table 3.

The ephemeral factor A appears to vary by subject; for biology, for example, this factor has a weight of 0.85 and a rate of decline of 0.18. For literature, the weight of factor A is 0.5 and the rate of decline is 0.1. For all subjects the decline in the B factor is 0.025.

It appears reasonable to suggest that the decline in residual factor B is relatively constant for all subjects, at about 2-2.5% per year. The annual decline rate for ephemeral factor A varies from about 10% for literature and fine arts, to near 20% for science. The difference in

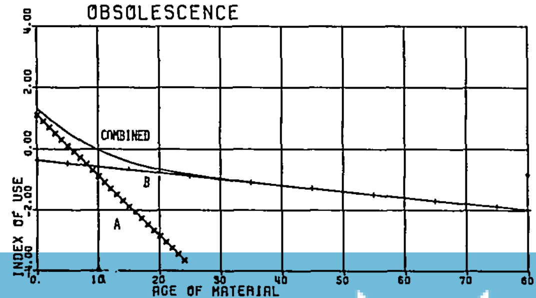

FIGURE 4. Index of use by age showing contributions of two factors

TABLE 3. Obsolescence factors, selected subjects.

Group	*A* Factor		*B* Factor	
	Weight	Rate	Weight	Rate
1973 — Total	0.813	0.18	0.187	0.02
1951 — Total	0.833	0.12	0.167	0.025
1951 — Biology	0.85	0.18	0.15	0.025
1951 — Literature	0.5	0.1	0.5	0.025

rate between A and B factors results in the disappearance of the A factor in influencing the index of use after about 30-35 years. Since the B factor appears to be uniform, the use, after about 35 years, of all subjects follows the same pattern of decay.

Conclusion

We may conclude that a multifactor decline in usage better describes the phenomenon than a simple exponential equation. The significance of this multifactor decline is that the decision on when to discard material becomes more difficult. If the retention policy is to discard when, let us say, the index of use declines to 20% of the initial index value, discarding will take place much sooner than if the decline were at a simple exponential rate; but if policy calls for retention until the index reaches 5% of the initial index, discarding will be postponed; assuming a simple exponential decay of about 4.5% per year, the decline to 20% will occur in 36 years; using the combined factors it will be achieved in less than 13 years. The decline to 5% on the single-factor model will occur after 65 years; with the multiple-factor model, approximately 71 years will be required. The comparisons will differ for specific subjects with different mixes of factors. Thus, development of an effective retention policy is more complex than previously thought.

References

1. Parker, Ralph H. *A Stochastic Analysis of Books Circulated from Elmer Ellis Library, 1972-1973.* Columbia, MO: University of Missouri; 1974.
2. Parker, Ralph H. "Bibliometric Models for Management of an Information Store: I. Differential Utility Among Items." *Journal of the American Society for Information Science.* 33:124-128; 1982.
3. Gosnell, Charles. "Obsolescence of Books in College Libraries." *College and Research Libraries.* 5(1):115-125; 1944.
4. Shaw, Charles B. *A List of Books for College Libraries.* Chicago: American Library Association; 1931.
5. Mohrhardt, Foster E. *A List of Books for Junior College Libraries.* Chicago: American Library Association; 1937.
6. Burton, R.E.; Kebler, R.W. "The Half-Life of Some Scientific and Technical Literature." *American Documentation.* 11(1): 18-22; 1960.
7. Cole, P.F. "Journal Usage Versus Age of Journal." *Journal of Documentation.* 19(1):1-11; 1963.
8. Brookes, B.C. "Obsolescence of Special Library Periodicals:

Sampling Errors and Utility Contours." *Journal of the American Society for Information Science*. 21(5):320–329; 1970.
9. Sandison, A. "The Use of Older Literature and its Obsolescence." *Journal of Documentation*. 27(3):184–189; 1971.
10. Line, Maurice B.; Sandison, A. "Obsolescence and Changes in the Use of Literature with Time." *Journal of Documentation*. 30(3):283–350; 1974.
11. Griffith, Belver; Servi, Patricia N.; Anker, Anita L.; Drott, M. Carl. "The Aging of Scientific Literature: A Citation Analysis." *Journal of Documentation*. 35(3):179–197; 1979.
12. Steig, Lewis. "A Technique for Evaluating the College Library Book Collection." *Library Quarterly*. 13(1):34–44; 1943.
13. Kilgour, Frederick G. "Recorded Use of Books in Yale Medical Library." *American Documentation*. 12(4):266–269; 1961.
14. Fussler, Herman; Simon, Julian. *Patterns in the Use of Books in Large Research Libraries*. Chicago: Univ. Chicago P.; 1969.
15. Morse, Philip M. *Library Effectiveness*. Cambridge, MA: MIT Press; 1968.
16. Hindle, A. "Markoff Models of Book Obsolescence." *Information Processing & Management*. 15(1):17–18; 1979.
17. Rouse, S.H.; Rouse, W.B. "Analysis of Monograph Obsolescence at Two Levels of an Interlibrary Network." *Information Processing & Management*. 15(5):219–225; 1979.