
The landmark citation method: analysis of a citation pattern as a collection assessment method*

By Catherine B. Soehner, M.L.S.
Technical Information Specialist

MEDLARS Management Section
National Library of Medicine
8600 Rockville Pike
Bethesda, Maryland 20894

S. Tanner Wray, M.L.S.
Head of Circulation Section and Coordinator of Regional
Document Delivery System†

National Agricultural Library
10301 Baltimore Boulevard
Beltsville, Maryland 20705

Daniel T. Richards, M.S.
Director of Biomedical Libraries‡

Dartmouth College
Hanover, New Hampshire 03755

The landmark citation method is a new collection assessment method based on the citation record of a single landmark article. This citation record is developed by identifying sources which cite the landmark article. A bibliography, extracted from the citation record, is then used to complete an assessment of the collection. This method was developed and used to assess the biotechnology collection of the National Library of Medicine. The information gained from this study, in addition to demonstrating the technique, also provided insight into the evolution of the biotechnology literature.

Collection assessment can be defined as measuring or determining the degree to which a library has acquired the books, journals, and other materials it intended to acquire, especially in relation to a written policy statement. The practice is common; indeed, researchers have noted that "collections have been evaluated since libraries began building them, and the literature related to the many methods available is large" [1]. Mosher traces programmatic collection evaluation back to the mid-nineteenth century, to Jewett's 1849 comparative analysis of citations in sev-

eral lists of notable books in international law, chemistry, and anthropology [2]. The lists were compared to the holdings of major libraries to support the budget request for the Smithsonian Library, which Jewett directed.

It was not until the 1930s and 1940s that true collection assessment emerged, probably as a reflection of the overall documentation process undertaken by many libraries. Over time, the codification of collection development policy and selection criteria, and the drafting of procedures for implementing those policies and criteria, provide the framework for developing a collection. This also serves as the structure within which a collection assessment is done.

To achieve the goals of most collection assessments, some measure is taken of the collection in comparison to both an external yardstick (a *universe*) and the collection development policy. A universe may be a bibliography or bibliographies in publications, or it may be established by compilation from a variety of

* Expanded version of a paper presented May 21, 1990, at the Ninetieth Annual Meeting of the Medical Library Association, Detroit, Michigan.

† Formerly, technical information specialist, Office on Smoking and Health, Centers for Disease Control, Rockville, Maryland.

‡ Formerly, collection development officer, National Library of Medicine, Bethesda, Maryland.

sources. The measurement consists of checking items in the universe in scope against a device containing the bibliographic records for the collection, typically a library catalog.

Collection assessments cannot be based on the premise that it is possible to examine every item or title. In any case, such an approach is not necessary as long as the assessment proceeds under a well-structured plan that includes a variety of methods to open windows into different aspects of the literature—by type, geographic origin, use, and so on. From these analyses of small segments of the literature, inferences can be drawn and applied to the overall subject collection. Such an assessment also provides substantial statistical data and produces a solid base for interpretation and planning.

This paper describes and evaluates an original collection assessment method, the landmark citation method. This technique was developed and used to assess the biotechnology collection of the National Library of Medicine (NLM), to describe the relationships within the biotechnology literature, and to make recommendations regarding the collecting process. NLM is mandated by Congress to collect, maintain, and disseminate biomedical information; therefore, the primary focus of this assessment was on the collection itself rather than on the ability of the collection to meet the needs of a specific user group.

BIOTECHNOLOGY AT NLM

Biotechnology is multidisciplinary, and its literature is derived from a wide range of subjects. Most aspects of biotechnology, especially those related to medicine and health, are collected comprehensively at NLM. In the 1987 biotechnology supplement to the *Collection Development Manual of the National Library of Medicine*, biotechnology is defined as that "body of knowledge which relates to the use of organisms, cells, or cell derived constituents for the purpose of developing products which are technically, scientifically, and/or clinically useful" [3]. The same document formalized NLM's commitment to collecting biotechnology literature at a comprehensive level. Comprehensive collections contain all significant works of recorded knowledge in all applicable languages from all time periods and geographic origins.

CONVENTIONAL ASSESSMENT

The two primary reasons for assessing any library collection are to determine the effectiveness of the acquisitions program and its procedures for implementing collection development policy and to enhance the service capabilities of the library by iden-

tifying and addressing weaknesses in the collection. Other rationales are to determine preservation priorities, to gather information for collection enhancement, and to judge the library's collection in comparison with those of other libraries.

An assessment plan is built upon choices—choices about sources or standards against which to measure a collection, choices among methods of examining specific aspects of a collection, choices in interpreting findings, and choices about the staff who will conduct the study. These decisions require professional judgment, awareness of methods, and, for collection of science materials, familiarity with scientific literature.

There is no single, universal method for conducting collection assessments. The American Library Association has published a useful pamphlet that provides a statement of basic principles and guidelines [4]. Hall compiled an especially helpful discussion of techniques for collection assessments [5]. These and other sources discuss collection-centered techniques (e.g., checking bibliographies, expert review of the collection) and use-centered techniques (e.g., circulation and in-house use studies, analysis of interlibrary loan statistics) and present the pros and cons of various approaches to collection analysis. Both Lockett [6] and Hall [7] recommend a combination of approaches for comprehensive collection assessment.

Due to the nature of the NLM collection and the commitment to comprehensive collection of biotechnology literature, collection-centered assessment methods were reviewed for possible use. Most of these methods require a well-structured bibliography of the related literature. These methods are problematic for emerging areas of study, such as biotechnology, which appeared as a distinct area of study in the 1970s and 1980s. In such cases, the literature itself is evolving and standard bibliographies generally are not available.

The complex nature of emerging scientific disciplines must be considered when planning a collection assessment. While scientific disciplines differ in mix and orientation of topics, they are in many ways interdependent. The primary literature of an emerging discipline often is found in the literature of older, more established fields. As a discipline matures, its literature develops unique characteristics. Thus, the literature of a new discipline often can be found in journals of related but more general subjects before the emergence of journals focusing specifically on the new field. This means that it can be extremely difficult to define precisely the universe of literature that supports a particular emerging scientific discipline. If an article of great importance dating back to the origin of a new discipline can be identified, an analysis of titles that cite that article can offer a glimpse of the early core of related literature.

LANDMARK CITATION METHOD

A new collection assessment technique, the landmark citation method, was developed in 1989 at NLM. This method is inherently collection-centered and is most useful for collections intended to be comprehensive in particular fields. The technique is citation-based but not in the traditional sense. Standard citation-based methods most frequently analyze citations from papers written by library users who relied on an existing library collection. Such an analysis provides an indicator of the degree to which the collection under review could have supported the research reported in the articles.

By contrast, the landmark citation method analyzes the citation record of a single landmark article in the field. The citation record is derived from *Science Citation Index*[®] (SCI[®]), by examining sources that cite the landmark article. A bibliography is developed by extracting from the citation record the titles of the journals and books in which the landmark article was cited. The bibliography then is compared to the collection development policy and the library collection to determine the degree to which the collection contains the literature of the field and meets the stated collecting goals.

The present study derived a bibliography from the literature citing a specific DNA sequencing technique. Because DNA sequencing is critical to biotechnology research and products, this bibliography can be extrapolated to the broader discipline. Consequently, the assessment results can be used to describe the NLM collection for the field.

It should be noted that, because SCI does not include all scientific literature, any subset derived from SCI shares its inherent limitations. SCI provides extensive coverage of mainstream scientific journals but includes few monographs, conference proceedings, nontraditional materials, or journals outside the mainstream of science and technology. This was not considered a serious limitation in the present study, as mainstream journals generally are regarded as the primary literature. Perhaps a greater weakness of SCI as a basis for developing a citation record is its lack of comprehensive coverage of foreign-language journals.

Identification of landmark article

A research scientist at the National Center for Biotechnology Information identified six articles that were important in the development of biotechnology and helped develop the methodology for this study. The articles were published between 1970 and 1977. The citation frequency for each article was determined using SCISEARCH[®]. A single article, Sanger et al., was selected initially due to its extraordinarily

high citation rate (more than 11,000 citations between 1978 and 1988) [8]. Moreover, both general and specialized works cited the article. As the assessment would reflect only a portion of the biotechnology literature, it was advantageous to select an article cited by as broad a range of works as possible.

The importance of the Sanger article to biotechnology was validated by consulting several sources describing the history of biotechnology. These sources described Sanger as a pioneer in biotechnology. He received the Nobel Prize in chemistry in 1958, and in 1975, he and two colleagues developed and implemented the first enzymatic sequencing technique. Two years later, this technique resulted in the sequencing of the lambda bacteriophage, consisting of 48,502 bases. This sequencing technique has become standard in biotechnology and led to Sanger's second Nobel Prize in chemistry [9]. The selected article describes this sequencing technique. This evidence clearly established the article as a landmark in biotechnology.

Development of citation record

A complete analysis of the more than 11,000 articles and books citing the landmark article was beyond the resources available for the study. A decision was made to draw a 10% sample, stratified by year of publication. The sample was identified in the volumes of SCI for 1975–1988. The print version of SCI was used rather than the online database, principally because of the high cost and the difficulty of printing only a sample using the online database SCISEARCH. The sample consisted of every tenth citation in cumulated SCI volumes prior to 1985 and annual volumes for 1985–1988.

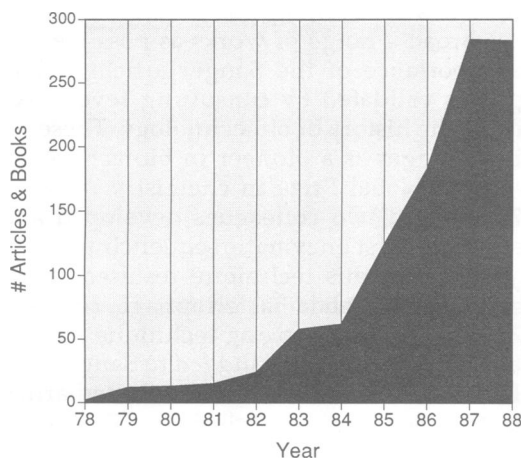
A rapid increase in the number of works citing the landmark article is apparent in the distribution of the sample by year of publication (Figure 1). The data underreport 1988 citations, as many of the citations published in 1988 did not appear in SCI until the 1989 volume. The 1989 volume of SCI had not yet been printed at the time of the study, hence the artificially level numbers between 1987 and 1988.

Use of cumulated volumes of SCI resulted in a slight over- or underrepresentation of some years. In 8 of the 11 years studied, the goal of 10% of citing items was met, with an actual range of 10% ± 2%. The variation for the other three years was as high as 3.8%. This sampling variation did not diminish the validity of the assessment results. The final sample of 1,060 items consisted of 1,058 articles and 2 books.

Development of bibliography

For each citation, the title of the journal or book, the publication year, and the language were recorded. To handle the large volume of information collected, a

Figure 1
Number of articles and books in sample citing landmark article, per year



database was developed using dBase III+® and Lotus®. These software packages were selected primarily due to their sorting and searching capabilities.

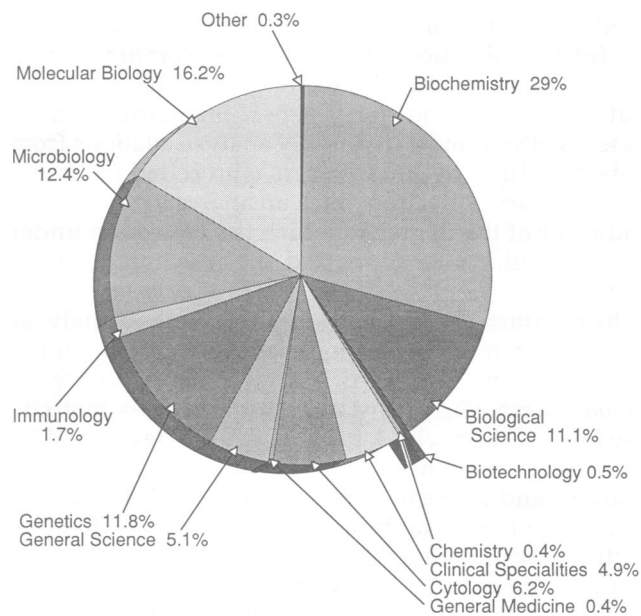
Scope Determination

The 1,058 articles citing the landmark article appeared in 135 unique journal titles. The two books and 135 journals constituted the initial assessment bibliography. The landmark article was cited in an increasing number of journals each year. This pattern of increase was similar to that in Figure 1. These journals and books were categorized using more than 100 subjects. The subjects were based loosely on indexing and cataloging subject headings in *Medical Subject Headings—Annotated Alphabetic List*. Each title was assigned up to three subject headings initially. Using these subject categories, the NLM collection development policy was reviewed to determine if the titles were in scope for the collection [10]. Eight journal titles were judged to be out of scope for the NLM collection and were eliminated from the sample, leaving a total of 127 journals and two books for the final assessment bibliography.

Comparison to the collection

The journal and book titles in the bibliography were compared to the SERLINE® and CATLINE® databases to determine whether they were in the NLM collection. Of the 127 journal titles, 126 or 99.2% were present. This result indicates that NLM successfully achieves its policy objective to collect biotechnology journal literature at the comprehensive level. Both book titles were in the NLM collection; however, the

Figure 2
Subject grouping of citing journal and book titles



data set was too small to support a conclusion in regard to books.

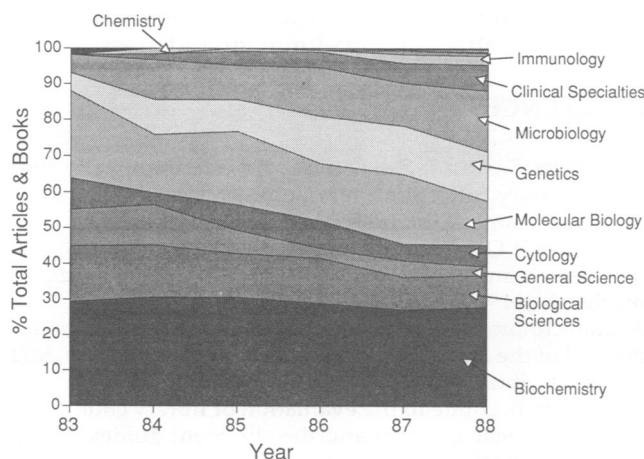
Literature characterization

Data assembled using this method, in addition to validating a collecting practice, provide an opportunity to characterize the literature related to the discipline in question. For the present study, this characterization was accomplished by consolidating the numerous specific subject areas assigned to each journal and book into thirteen broader categories. Each journal and book was assigned only one broad subject heading. Both in-scope and out-of-scope titles were analyzed.

Figure 2 depicts a preliminary subject representation of journals and books citing Sanger's article. Each article in the sample counted as one point for the subject assigned to the journal. Likewise, each book counted as one point for its assigned subject. It is evident that the early biotechnology literature derives from many different scientific disciplines, including the more established fields of biochemistry, molecular biology, and microbiology.

Additional analysis was performed to examine the changes in the relative ratios of subjects over the final six years in the study. These years contained sufficient numbers of citations to reflect meaningful trends. These variations over time are presented in Figures 3 and 4. These figures depict the increasing impor-

Figure 3
Variation in subject classification of titles over time

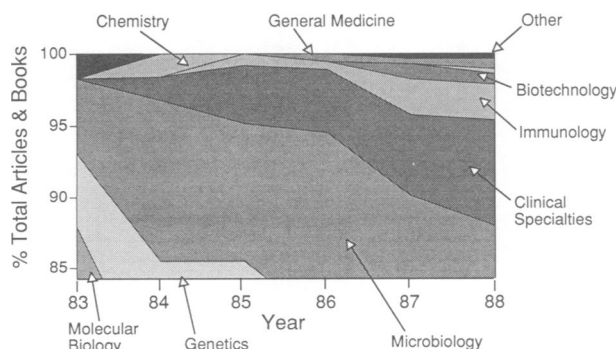


tance that Sanger's sequencing technique assumed in a widening range of scientific fields. Figure 4 reveals the emergence of a distinct biotechnology literature. Such data can provide an approximate timing for the emergence of a distinct core literature. However, definitive statements about the emergence of the core literature of biotechnology would require verification using other, independent measures.

CONCLUSION

The landmark citation method is a new method that has been added to the array of collection assessment methodologies available to librarians. The method uses the citation record of a single landmark article in a field. The citation record is derived from *SCI* by examining sources that cite the landmark article. An assessment bibliography is developed by extracting

Figure 4
Enlargement of upper 16% of Figure 3



from the citation record the titles of the journals and books in which the landmark article was cited. This bibliography then is compared to the collection development policy and library collection to determine the degree to which the collection contains the literature of the field and meets the stated goals.

Prior to using any collection assessment method, librarians must consider both the advantages and limitations of the method. The landmark citation method has several advantages and limitations compared to other collection assessment techniques.

Advantages

Facilitates a cross-disciplinary assessment. Standard citation lists or bibliographies generally include materials from a single discipline. This method permits analysis of the literature from a wide range of disciplines.

Serves as an alternative to other methods. This method can be used when other methods are not feasible or are inappropriate. For example, when the discipline is emerging or when a discipline's literature is scattered in many fields, other collection assessment methods may provide insufficient data for meaningful assessment.

Eliminates bibliography bias. Many bibliographies are limited to the literature to which the compiler is exposed or with which the compiler is familiar. The landmark citation method is unbiased because it includes the significant and insignificant. For this reason, its bibliography presents a more global view of the literature of a field and is more representative of that literature.

Provides a capsule view of core literature. Because a landmark article is used as a base, citations should be found in a broad range of the literature related to the field. The citing works therefore may be considered the core literature of the new discipline.

Useful for describing the literature of an emerging discipline. The data derived from this method establish an outline of the development of a discipline's literature. Patterns of increasing or diminishing importance of the literatures of other disciplines can be observed. Such analysis promotes awareness of the connections among scientific disciplines.

Disadvantages

Restricts assessment to cited literature. This method reviews only the literature citing one landmark work and is based on a citation count. The method may overlook related literature that does not cite the land-

mark work. Furthermore there is no control over which authors or journals cite the landmark article.

Expense. This method requires a substantial commitment of staff resources and is more expensive than other assessment methods. Special online service offered by the Institute for Scientific Information was considered, but this proved to be even more expensive than the staff costs.

Limited coverage of nonjournal literature. *SCI* is limited in its inclusion of monographs, conference proceedings, nontraditional materials, foreign-language materials, and journals outside the mainstream of science and technology literature. Consequently, parallel studies for those formats should be undertaken to attain a more comprehensive view of the state of a collection.

No user consideration. As with most collection assessment methods, this method examines only a segment of a literature; inferences are made regarding the larger collection. The method is inherently collection-centered and, as such, is most useful for collections that are comprehensive for particular fields. It has limited applicability to user-oriented collections.

The landmark citation method is particularly well suited for emerging disciplines lacking a well-structured bibliography, for cross-disciplinary assessments, and as an adjunct to other collection assessment methods. The landmark citation method is a

new approach that facilitates broader understanding of the relationships among scientific disciplines. As scientific disciplines become increasingly interdependent, an awareness of these relationships is more critical for collection development and assessment.

REFERENCES

1. ROBBINS-CARTER J, ZWEIZIG DL. Are we there yet?: evaluating library collections, reference services, programs and personnel. *Amer Libr* 1985;16(11):724.
2. MOSHER P. Quality and library collections: new directions in research and practice in collection evaluation. *Adv Librarianship* 1984;13:211-38.
3. Biotechnology, [supplement to] collection development manual of the National Library of Medicine. Bethesda, MD: National Library of Medicine, 1989.
4. LOCKETT B. Guide to the evaluation of library collections. (Collection management and development guides, no. 2.) Chicago: American Library Association, 1989.
5. HALL BH. Collection assessment manual for college and university libraries. Phoenix: Oryx Press, 1985.
6. LOCKETT, op. cit., 1.
7. HALL, op. cit., 7.
8. SANGER F, NICKLEN S, COULSON AR. DNA sequencing with chain-terminating inhibitors. *Proc Nat Acad Sci* 1977;74:5463-7.
9. ANTEBI E, FISHLOCK D. Biotechnology: strategies for life. Cambridge, MA: MIT Press, 1986.
10. TECHNICAL SERVICES DIVISION, NATIONAL LIBRARY OF MEDICINE. Collection development manual of the National Library of Medicine. Bethesda, MD: The Library, 1985.

Received October 1991; accepted December 1991