

VERN M. PINGS

Development of Quantitative Assessment of Medical Libraries

Assessments of library services can be used for a variety of purposes, including improvement of services, management justification, and budget preparation. The Institute for the Advancement of Medical Communication has developed methods for testing an academic research medical library's ability to provide specific services. These include the ability to deliver books and journals to primary clientele, to supply answers of simple fact, to deliver documents through inter-library lending, and to verify and correct citations. Other methods were tested for evaluation and comparison of library service policies and utilization of services by a library's primary clientele. A random alarm device was developed. The use of this device has been tested and has demonstrated that many library "statistics" can be collected randomly and provide the same information that can be acquired by maintaining "total" counts of such services as circulation and library use.

ONCE MAN LEARNED to record his experiences, the invention of libraries was inevitable. Libraries, in an abstract sense, are man-made institutions, and one reason for their existence is to support the social organizations man defines. The object of study of natural scientists assumes that there is a regularity in nature; this regularity is "given" and

Dr. Pings is Medical Librarian in Wayne State University. Preparation of this paper was supported in part by USPHS Contract PH 43-66-540 from the National Library of Medicine. The data and many of the concepts in the paper were developed by project members supported by this contract. The principal investigator was Richard H. Orr, M.D., Director, Institute for the Advancement of Medical Communication. The principal collaborators were Edwin Olson, Ph.D., Senior Research Associate, Institute for the Advancement of Medical Communication; Irwin H. Pizer, Director of the Library and Associate Professor of Medical History, State University of New York, Upstate Medical Center; and the author.

although scientists search for ways to control regularity, any investigation into the purpose of nature soon moves into the disputatious realm of philosophy and theology. Any investigation of a man-made institution, on the other hand, must include in its methodology and approach the purpose of the institution if it is to produce any results that can be used to alter, improve, or justify the institutional function.

Libraries, as social entities, have only relatively recently come under the scrutiny of systematic study. Certainly these studies have contributed to the improvement of library service, but the methodology of library surveys established by the turn of the century has changed little. Lyle, while criticizing this lack of innovation, rationalizes that investigation of academic library functions has rested upon busy, practicing librarians who, for the sake of convenience perhaps, have usually taken "a greater interest in the measurable and organizational problems of finance, physical plant and equip-

ment, and operational problems than they do in some of the more intangible problems of policy, relationships, and library use."¹ Most of us can easily agree with Clapp's remark that the problems of the research library arise from the gap that exists between what its users require and what it can supply.² We need to find out how to evaluate, or at least make decisions about, the function of the library in the research and related programs which society supports.

This paper is a summary report of a project undertaken between July 1966 and June 1968, supported by a contract from the National Library of Medicine to the Institute for Advancement of Medical Communication "to develop methods for collecting objective data suitable for planning and guiding local, regional, and national programs to improve biomedical libraries and the biomedical information complex." This succinct statement requires amplification to make the nature of the project clear. Unlike most studies in which the purpose is to collect data and test hypotheses, where methods are only the means to these ends, the emphasis is completely reversed in this project—it is *methods*, rather than data, that are of primary interest.³

GENERAL APPROACH

Requirements

In general, three considerations guided the project throughout, which served not only as a perspective within which to work but also as a check on progress and purposes. First, any survey method developed had to be applicable to academic biomedical libraries. As expansive as one might like to have been, the

work could not be made applicable to *all* biomedical libraries, at least without some modification. It had to be recognized from the outset that attitudes of librarians and their supervisors to standardization of all kinds have too often been ambivalent. They have argued in their writings for uniform standards and practices; they have set up committees to design standard codes; and at the same time nearly every library is non-standard to a greater or lesser degree in many of its operations and records.⁴ The efforts of the project were therefore directed primarily to developing instruments that could be applied to the medical school library. Any instrument that was useful for this diverse, although circumscribed, group of libraries should be adaptable for use in other libraries. The second general requirement was that any method of data collecting should be one that could be applied in an operating library by the staff. Finally, as a corollary to the last requirement, the methods used and the resultant analysis of data must have meaning to librarians as well as to nonlibrarians, users, and administrators.

Perspective

The concept and the desirability for interlibrary dependence since its promulgation at the turn of the century, has all but become part of the ethic of librarianship. Yet nearly all the investigative work on libraries involves but one library or, at most, the relationship of a main library to a group of subunits with the effort to determine whether the administrative unit with its external constraints of space and finances is functioning efficiently. Rather than trying to refine old and search for new approaches to studying individual libraries, it was decided to try to define methods that would be applicable in evaluating

¹ Guy R. Lyle, "An Exploration into the Origins and Evaluation of the Library Survey," in Maurice F. Tauber and Irene R. Stephens, *Library Surveys* (New York: Columbia University Press, 1967), p. 3-22.

² Verner W. Clapp, *The Future of the Research Library* (Urbana: University of Illinois Press, 1964).

³ A series of articles is being prepared describing what methods were tested and how they were selected and developed. The first two articles in this series have been accepted for publication in the July 1968 issue of the *Bulletin of the Medical Library Association*.

⁴ Cf. J. W. Jolliffe, M. B. Line, and F. Robinson, "Why Libraries Differ—and Need They?" in Nigel S. M. Cox and M. W. Grosse, *Organization and Handling of Bibliographic Records by Computer* (Hamden, Conn.: Anchor Books, 1967), p. 62-68.

chosen for discussion here are those methods which are ready to be applied directly to academic and resource medical libraries and with modification to all libraries, biomedical and others.

Inventory of Library Services to Individual Users

The goal here was to devise a standardized procedure for gathering information needed to describe in detail the services a library provides to each of its user groups and to provide a framework to relate other methods of data collecting to be tested. The viewpoint adopted, as already pointed out, was that of the user; in other words, only those library activities were considered which a user could see as a direct service or as a policy which would affect his time, effort, or convenience in using the library. After much discussion and testing, fifty-three separate services were delineated. In an effort to standardize this checklist two methods were used. First, alternative policies for each service were categorized in an hierarchical array from the most restrictive to the least. For example, a library may not circulate any of its serials; it may restrict circulation of unbound issues only; it may restrict just the current issues; or it may allow all serials to circulate to faculty but not to students (see Fig. 1). Any hierarchical arrangement has to be arbitrary, but this checklist provides a convenient format to demonstrate variations in service between different user groups. Further, it provides the possibility of assigning a score to each service. A perfect score for the paragon of libraries would be 1,000. We say "possibility" because the one way to establish numerical values for such a qualitative matter as service is by agreement among a group of experts. The second method used to standardize this checklist was the development of an annotated "question-tree" for each of the services to accompany the checklist (see Fig. 2). If a survey of many libraries

were undertaken, this forced-choice method of collecting information through interviews would give some assurance that the same meanings and the same purposes were applied. Although originally devised as an interview guide, it can be used by institutions for self-evaluation and for teaching students and new staff of library service policies.

Inventory of Policies on Services to Other Libraries

The checklist described above included services relating only to individuals. When one begins to include interinstitutional services in a policy checklist, the behind-the-scenes library functions cannot be ignored. Another inventory was developed listing sixty-nine possible services and covering 230 alternative policies involving interlibrary relationships. No "interview guide" was prepared for this checklist because it was assumed that anyone using it would first familiarize himself with the individual-user checklist format.

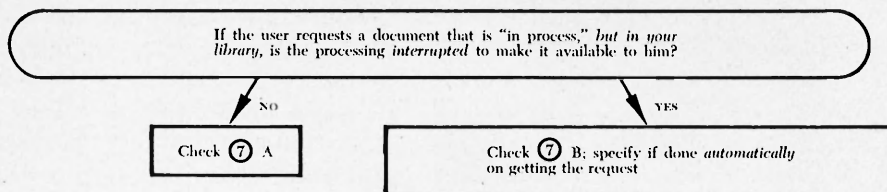
Document Delivery Test

Two separate tests have been developed.

Medical research libraries. In this method, a Document Delivery Test (DDT) is administered at a library by searching its collection for each document in a standard test sample consisting of about three hundred citations drawn at random from a large pool representing documents cited by U.S. biomedical scientists. The availability of each test document in the given library is translated into numbers that reflect how long it would take a user to obtain the different test documents from the library. A single figure-of-merit, which may be called the Capability Index (CI), is then calculated from the mean "delivery time" for all documents in the test sample. The CI for a library ranges from 100, if its users could have obtained all of the test documents in ten

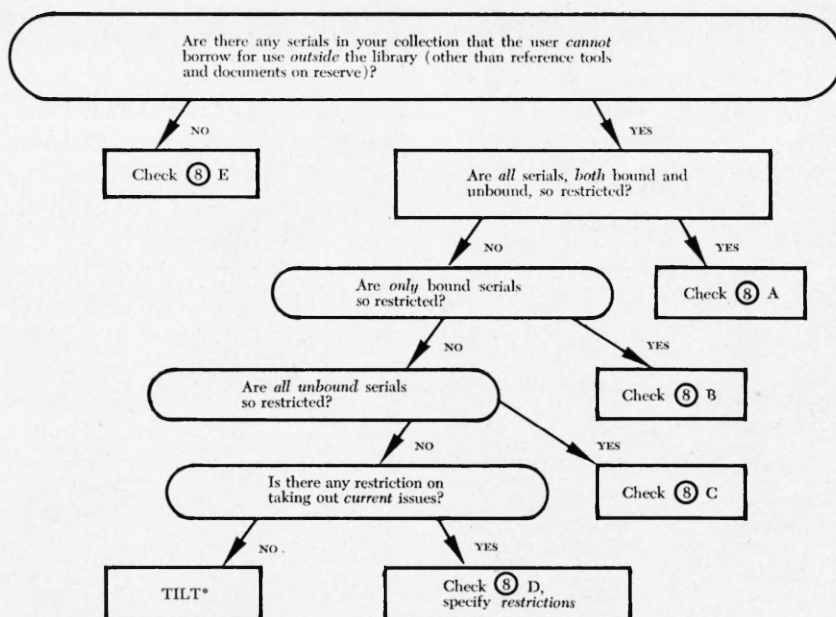
SECTION ⑦

AVAILABILITY OF "IN-PROCESS" DOCUMENTS



SECTION ⑧

CIRCULATION OF SERIALS



° Here, as elsewhere, when a question tree leads to TILT, it means that some dichotomy has either broken down or has been misunderstood. Often this can be remedied by starting down the tree again and clearing up yes-no choices by more explanation or by specifying the exceptions that caused the difficulty. For example, in this tree, TILT might be reached if the only serials not circulated were specific titles or specific volumes. In this case the difficulty could be handled by checking 8 E, and specifying the exceptions.

Fig. 2—Illustration of the annotated "question-tree" used with the library service policy checklist.

minutes or less, to 0, if none of the test documents could have been obtained in less than a week. Different options for calculating the CI allow one to correct, or control, this index for (a) the number of test documents that happen to be in use within the library at the time of the test; (b) the number of test documents that happen to be in circulation at the

time of the test; and (c) how long it typically takes the library to obtain documents by interlibrary loan. The CI's on repeated tests of a given library, with *different* samples from the same citation pool, can be expected to vary within ± 5 points in 95 out of 100 such tests. At the same level of confidence, one can conclude that the document delivery

capabilities of two libraries tested with the same sample are significantly different if their CIs differ by as much as six points.

Reservoir libraries. A special DDT was devised to determine a library's ability to relieve the National Library of Medicine of its load in providing documents. A total of 305 interlibrary loan requests received at the National Library of Medicine during fiscal 1967 were randomly selected to serve as the sample for this DDT. The procedure for performing this test is identical with the DDT at a research library, and the calculation of the CI is accomplished in the same way. A different time base must be used in scoring, however; supplying documents through interlibrary loan in a matter of minutes is obviously unrealistic.

Evaluation of the DDT. The dependability of the DDT rests on the care with which a citation pool is developed from which to select sample citations. The selection and verification of citations is expensive, and test samples can only be practicably developed if they are to be applied to a large group of libraries.⁵ Selecting documents from a citation pool provides an objectivity that is not possible if recommended lists or the circulation records of individual libraries are used; the source of the citation pool determines the validity of the CI. The CI from the sample designed for resource libraries should not be interpreted as reflecting a library's ability to serve *all* user needs: for example, materials to support educational programs and current awareness. The CI resulting from applying the test sample for reservoir libraries measures a different function from the sample developed from citations of biomedical research papers.

⁵ At the time of writing the University City Science Center is applying the tests so far described to over a hundred medical school libraries and to a selected group of reservoir libraries. This survey is undertaken by a contract with the National Library of Medicine.

Interlibrary lending. The perspective used was that of studying a library complex rather than individual libraries. One function of such a complex is interlibrary lending. Accurate knowledge of the processing time involved in lending a document to another library is useful in assessing the ability of a complex to supply documents to individual users. While many of the activities of interlibrary loans, such as mail deliveries, are beyond the control of libraries, a large share of the routines of an interlibrary loan transaction are under the direct control of a library. A method was devised to evaluate the routines involved with the lending procedures. A form was designed to accompany each request which required the recording of the time in performing actions in filling the request. Since each record was an actual time, this information can distinguish a library's "total processing time," which can be further divided into two segments: "request-processing time," ending when the library's staff has the requested document physically in hand; and "loan-processing time," starting where the request process finished and ending with the original or facsimile copy ready for delivery. Processing time, either total or its two segments, can be calculated then at any specified percentage level: for example, 80 per cent of requests are processed within x hours. This method has been carried out at six libraries, who have not found it cumbersome to collect data. This procedure can serve as a monitoring device not only to assess a library's ability to fill a request within certain time limitations, but also to provide information on the internal processing operation for exploring alternative routines which can lead to improved document delivery performance.

Citation and Information Services

One of the services all libraries provide is assistance to users by identifying

documents that may contain needed information or by supplying the information directly. Two tests were developed.

Verifying and correcting citations. A library's capability for providing its users with the bibliographic tools they need to verify citations themselves ("self-use" service) and its capability for "staff-mediated" service, can be assessed by the same test. Samples of fifty incomplete or incorrect citations were selected from a large pool generated by biomedical researchers.⁶ Knowing the secondary sources that are needed to verify these citations can give an index of a library's capability for verifying citations. The same sample can be used for a staff-mediated test in which a reference librarian is given four hours to verify the citations. With a sample of this size, the reliability and sensitivity of both tests is high enough to be acceptable for many purposes. Both are practical to administer; given the test materials and written instructions, the head of any library can administer the test and score the results.

Simple-fact information service. Anyone faced with the task of teaching a course in specialized reference can sympathize with the efforts made to develop a method for evaluating a library's ability to supply answers to specific questions. The problem lies in selecting realistic questions which do not conflict with library service policies or are not merely a test of individual librarians' abilities. A group of biomedical scientists were given published research papers and requested to indicate the questions they would need to have answered if they had done the research reported. Two sets of test questions were thus developed. Just as with the citation verification test, upper and lower limits of reliability can be established with fifty questions.

⁶ Actually selected from the citation pools generated for the DDT described.

Library Service Records

Each library keeps certain "statistics" for management control. These data, or at least those that are published, are descriptive and in most instances are based on a total count of circulation and a total number of reference questions answered. Administrators would like to have more data of this kind rather than less, but the cost in staff time to collect the data is prohibitive. For management decisions, information from a proper sample of a universe is as satisfactory as knowledge of the total universe. The difficulty lies in obtaining an appropriate random sample on which to apply statistical analytic methods.

A battery-powered random alarm mechanism (RAM) was developed by a commercial organization according to specifications. The RAM measures about two by three inches and can be easily carried in a shirt pocket since it is about the size of a cigarette package.⁷ The alarm, a short beep, is amplified through a small speaker that can be clipped to the shirt pocket or coat lapel, or even carried as an ear plug. The interval between alarms varies from as short as ten minutes to as long as six hours.

The use of the RAM to sample a wide variety of in-library activities has been explored, including self-service photocopying, use of the card catalog and other bibliographic instruments, library visits, utilization of library space, and provision of staff-mediated reference services. These activities were selected because they represent library services for which current measurement techniques are either unreliable or impractical for routine applications. Not investigated were applications to the technical services of a library, but the time sampling technique is an obvious application. On the basis of the experience and

⁷ Women's fashions preclude such a convenient method of carrying the RAM, but it can be conveniently attached to a clip board which then can be carried and thus also serve as a means for recording the observations made.

data accumulated, it can be stated that the use of the RAM is not only practical and efficient of library staff time, but its use will make possible the collection of reliable data on some major services that, up to now, have gone largely unmeasured except in one-time studies.

OTHER AREAS INVESTIGATED

One area which has been investigated extensively and is beginning to reveal an approach for additional study might be characterized as the "utilization of library services." Although most scholars have what might be identified as their primary library, several library resources are available to them except in instances of geographic isolation. In an academic medical environment this may include department collections, branch libraries, hospital libraries, and the resources of other academic and public libraries. An understanding of the relative use of alternative resources by a particular library user population is necessary if realistic planning for local and regional library services is to be undertaken. In a biomedical environment the relationship of library service is especially complex because the concentration of research and specialized health care in academic medical centers has produced specific patterns of library functions and policies.

Utilizing published data from school catalogs, data from administrative reports and records such as payroll and personnel files, and information collected through a questionnaire to a random stratified sample of a library's primary users, statistical techniques have been tested to identify some of the variations that exist in different library environments. Thus, by combining the information acquired through library service records with the characterization of a user population, a measure of library utilization can be defined. This approach in studying library utilization not only produces a powerful instrument in understanding the function of library

service to specific populations, but it also provides valuable information for planning and management. How a library is used depends on many factors, some of which are beyond its control, including such things as availability of space and the improvement of other resources accessible to its users.

As with any investigative work, more problems are identified than solved. This is certainly true in this area. The aim of the study was only to test the applicability of certain data-gathering and analysis techniques on library utilization. Although the goal has been attained, and although the techniques may be useful if applied "as is," sampling plans that are applicable for academic medical libraries in general cannot yet be recommended. The wealth of data collected has demonstrated the potential and the need for further refinement.

FURTHER APPLICATION AND DEVELOPMENTS OF METHODS

This summary report has concentrated on the work accomplished toward the development of techniques for measuring library service which have been tested and are ready for general application and use for academic medical libraries. The use of these methods for other kinds of libraries should be practical; the assumptions underlying the methods, however, would have to be reviewed to determine whether the functions tested are appropriate. The theoretical considerations, limitations, and general characteristics of each of the methods reported here will be described in detail and published elsewhere. Their practicality for assessing the functioning of library and information complexes has been demonstrated, but of equal importance is the use of these methods for research in library administration. Data collected from diverse environments can be an important resource for investigators who want to test hypotheses and conduct exploratory analyses. ■■