

Analysis of Collection Development at the National Library of Medicine

BY DAZA FOX CRAIG, *Librarian*
Bibliographic Services Division

PAULA MEISE STRAIN, *Collection Development Librarian**
Technical Services Division

National Library of Medicine
Bethesda, Maryland

ABSTRACT

This paper reports the major findings of a study of collection development activities at the National Library of Medicine (NLM) from 1965 to 1977. The CATLINE file was the source of the data; analyses were performed on classification number, date of entry, and language. An overview analysis of the data base is presented for major subject and form classes. An in-depth subject analysis of the monograph collection was performed using the NLM call number. An analysis by date of entry revealed that the subject content of CATLINE has varied only slightly over the years; the most notable change was a recent decline in the related and peripheral subject areas. The language analysis indicated that 83% of the data base consisted of works published in English, German, Russian, and French. Throughput processing time was measured for English language monographs for selected years.

EVALUATION of a library collection has always been a difficult and seldom undertaken task, especially for large libraries. Collection of the necessary data for analysis and evaluation of a large collection is often prohibitively expensive in terms of both manpower and time. As records of library holdings become available in computerized form, compilation of collection data for evaluation purposes becomes more feasible. Libraries can now more readily obtain information about their collections, so that decisions relating to collection development priorities and funding can be supported by facts, rather than by intuition.

A variety of methods of evaluating data on library collections has been utilized. Quantitative evaluations have been most widely used over the years, particularly in larger libraries, but qualitative evaluation methods have also been documented. A survey of the literature on collection

evaluation shows some definite trends: (1) smaller libraries tend to compare their holdings to larger libraries; (2) larger libraries rely heavily on quantitative rather than qualitative information [1]; (3) qualitative evaluations are usually accomplished by comparing comprehensive subject bibliographies, standard lists, and catalogs of special collections to the card catalog [2-3], and (4) the adequacy of library collections is usually measured in terms of user needs [4-5].

The National Library of Medicine (NLM), with an estimated 1.5 million unique volumes, has many of the characteristics and problems associated with large research libraries. Collection evaluations of any kind have been infrequent and, until now, performed by non-NLM staff. In 1944 a team of outside consultants conducted an in-depth study of the serial collection as part of an overall survey of the library [6]. Recently librarians at other medical libraries have used NLM's computerized catalog, CATLINE (Catalog-online), to evaluate their own collections. The CATLINE file contains all serial and monographic titles cataloged at NLM since 1965. In 1976 Byrd used CATLINE to compare the subject coverage at two biomedical libraries (the University of South Dakota Health Science Library and the University of Nebraska Medical Center Library) with NLM [7]. A study by Kronick and Bowden reported their findings in comparing the collections of two health sciences libraries in Texas to CATLINE [8]. In the past, NLM staff members have used CATLINE to provide management with data about the collection on an ad hoc basis, but this paper reports the first use of CATLINE by NLM to comprehensively evaluate its own collection development activity during the past thirteen years.

NLM has the mission of providing comprehensive coverage of the world's biomedical literature

*Mrs. Strain is now a full-time homemaker.

on a current and archival basis. The library's collection development policy and guidelines are provided in the *NLM Scope and Coverage Manual* [9]; subjects that are in scope for the library are categorized as being either core, related, or peripheral. NLM's collection development goals are the establishment of a comprehensive collection in the core biomedical subjects, a research collection in the subjects closely related to biomedicine, and a reference collection in subjects peripheral to medicine and the health and life sciences. To achieve these goals, a group of selectors with subject and language expertise reviews and selects material appropriate for NLM's collection. Catalogs of special collections in medical subjects, medical works listed in national bibliographies, and specialized subject bibliographies are searched routinely to provide quality control of the library's numerous blanket-order arrangements in the U.S. and abroad.

The purpose of this paper is to report the results of a quantitative study of the CATLINE data base as an indication of post-1965 collection development activity at NLM in regard to subject, date of entry, and language. The study, performed in the summer of 1978, sought to examine collection activity in each subject area to establish whether there has been a shift in subject emphasis over the years. A similar objective was established for analysis by language. The authors set out to determine the distribution of languages represented in the collection to establish whether there has been a discernible change in the ratio of English and foreign language materials. Limited analysis on selected years was also performed to determine the length of time (throughput time) between date of publication and date of entry into CATLINE.

The authors formulated two major hypotheses: (1) the 1977 revision of the NLM scope and coverage guidelines has effected a change in the collection in specific subject areas; and (2) processing changes and manpower limitations in the library's Technical Services Division during the past few years have affected the growth patterns of the collection in terms of language and currency.

The 1977 revision of the NLM scope and coverage guidelines was aimed at more clearly defining the library's collection responsibilities, particularly in the sciences and other subjects that are peripherally related to medicine. The effect of this revision had not yet been measured. The major processing change in the past decade had involved the establishment of a priority system to direct the flow of materials through acquisitions and catalog-

ing. Recently published English language materials in the core medical subjects are given top priority, while works published in foreign languages and titles older than three years are assigned lower priorities. Reductions in cataloging manpower and the resulting decrease in foreign language capability among staff were expected to affect the amount and kind of foreign language materials being cataloged.

The scope of the study was the entire universe of CATLINE, with the exception of 2,800 records that represent non-NLM holdings. This small percentage of records (0.02%) represents items cataloged as part of the cooperative cataloging program with the Harvard, SUNY (Syracuse), and UCLA biomedical libraries. The population size, therefore, was 175,479 records representing unique titles in the NLM collection.

The study was primarily aimed at analyzing the characteristics of the cataloged monograph collection, although some summary data on other types of materials have been presented. The NLM practice of classifying some items by form rather than subject enabled us to eliminate serials, government series, theses, and pamphlets from certain portions of the study.

This study may not be representative of the entire NLM collection because it does not include materials cataloged prior to 1965 (pre-CATLINE). The study also excludes the 22,000 titles currently in NLM's brief-listed collection. This group consists of older materials, items published in more esoteric foreign languages, and publications in many of the noncore subjects. In-house access to these items is available through an on-line in-process file, INPROC. In spite of these limitations, this study does provide a useful quantitative analysis of collection development at NLM based on that major portion of the current collection which has been cataloged.

PROCEDURES

Selected fields of the CATLINE record were chosen for quantitative assessment. These included the call number (to allow subject breakdown by class number), language, year of publication, and date of entry into CATLINE. A special file was created by copying these and other selected fields from each CATLINE unit record representing an NLM title into a software data management system named INQUIRE (a registered trademark of Infodata Systems Inc., Falls Church, Virginia). This system provides the major support for techni-

cal services processing at NLM and allows for data analysis and flexibility of output.

This new file was named INQCAT. The authors were able to query the file on-line, but because of the size of the file, most report generation was performed at night in batch mode. The data in these reports were analyzed and evaluated; the findings will be summarized in this paper.

OVERVIEW ANALYSIS OF CATLINE

An analysis was made of the distribution of items in CATLINE by major subject and form classes (see Fig. 1). The total study population of 175,479 records was sorted on the first three characters of the call number to obtain these data. Subject-classed material comprised 65.5% of CATLINE (114,900 items); these titles were the focus for subsequent in-depth analysis. The subject-classed portion of CATLINE represents the monograph collection except for a small percentage (3.5%) of subject-classed serials that are primarily in the Bibliography (Z) class. These serial items were not eliminated from the study population, since this small margin of error did not justify expenditure of the required amount of additional computer time. CIP (Cataloging in Publication) items were also included, even though these records represented titles not yet received by NLM.

Titles classified in the form numbers W1-W9

accounted for 30.8% of CATLINE (54,103 items). A small group of miscellaneous items (COSATI-classed technical report literature, films, and other audiovisuals) comprised the remaining 3.7% of CATLINE (6,476 items).

Subject analysis of these form-classed and miscellaneous items was not undertaken in this study, since subject access to these items is provided only through assigned MeSH headings; this portion of CATLINE could be analyzed at a later date. Subject classification is provided for some form-classed materials through the OXNLM call number, but this service did not begin until 1975. Subject analysis of records with an OXNLM call number was not included in this study because only a small percentage of records for monographic serials contain this additional call number.

SUBJECT ANALYSIS

The subject-classed materials were subdivided into groups corresponding to the major divisions in the NLM collection: the core monograph collection containing works in medicine and preclinical sciences, and the noncore monograph collection containing works in the related and peripheral subjects. Fig. 1 indicates the distribution of titles within these two groups: the core monograph collection, representing 49.8% of the CATLINE titles; and the noncore collection, accounting for 15.5% of the titles. The detailed results of the subject analysis follow.

Core Monograph Collection

NLM's core monograph collection consists of works classed in the twenty-seven medical classes (W-WZ) and in the eight classes in the Library of Congress Q schedule designated for the preclinical sciences (QS-QZ). The NLM scope and coverage policy provides for comprehensive collection of the world's biomedical literature in these subject areas. Works classed in the medical classes represent 40.2% of the CATLINE records (70,532 titles). The relative volume of materials cataloged in each of the twenty-seven medical classes (W-WZ) is shown in Fig. 2. The median number of items within each of the twenty-seven classes was 2,045 items, or 2.9% of the total for medicine. As indicated by the chart, the subject areas that contained the greatest number of titles were WM - Psychiatry (7,884 titles), WA - Public Health (6,790 titles), WB - Practice of Medicine (5,109 titles), and WL - Nervous System (3,770 titles). The subject area containing the least number of titles was WR - Dermatology, which included only

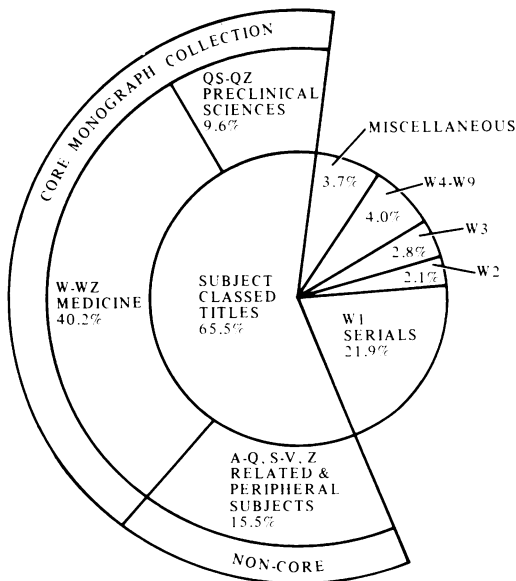


FIG. 1—Distribution of Subject and Form Classes in CATLINE.

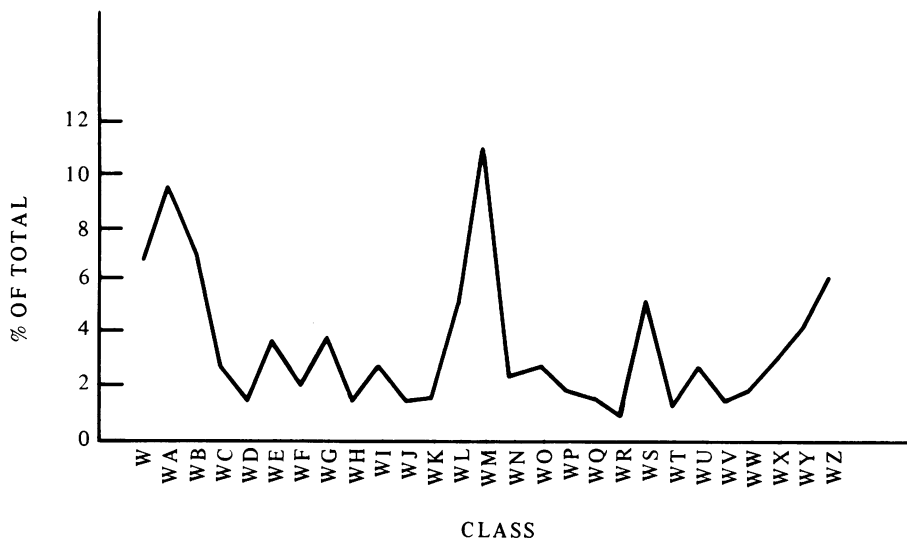


FIG. 2.—Distribution of Materials in the Medical Classes.

743 records. Geriatrics (WT), Urogenital System (WJ), Hemic and Lymphatic System (WH), and Otorhinolaryngology (WV) were also low, each subject area representing slightly more than 1,000 titles.

Titles classed in the preclinical sciences represent 9.6% of the CATLINE records (16,874 items). The distribution of materials cataloged in these eight classes (QS–QZ) is shown in Fig. 3. The subject area that represented the greatest

number of titles by far was QV - Pharmacology (4,682 titles). Parasitology (QX) contained the fewest number of titles (559). The median number of titles for each of the preclinical subject classes was 2,000.

The indicated distribution of titles in the medical and preclinical subjects did not differ significantly from the results reported by Kronick and Bowden in their study of CATLINE in early 1977. Unlike our present study, their subject analysis of CATLINE through 1976 included form-classed titles with OXNLM call numbers. Despite differences in population size and types of material included in the respective studies, the data seem to indicate that in the one and one-half years between the two studies, the distribution of subjects in CATLINE has remained fairly constant. This constancy may be indicative of rather stable levels of publishing activity in each of the medical and preclinical subject areas.

Noncore Monograph Collection

The related and peripheral subject classes represented 15.5% of the CATLINE titles (27,494 items); Fig. 4 shows the distribution of these titles by class. The subject area that contained the greatest number of titles was the Q schedule, Science (excluding QS–QZ). A more detailed analysis showed that the large majority of these Q schedule titles were QL - Zoology (2,658 titles), QD - Chemistry (1,147 titles), and QH - Natural History (973 titles). Bibliography (Z) and the Social Sciences (H) were the next most highly

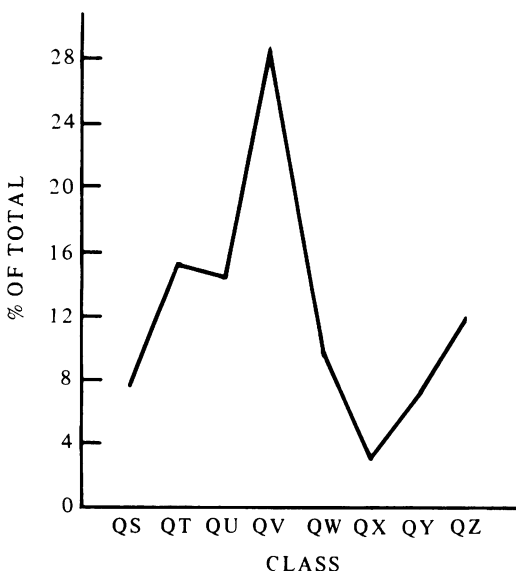


FIG. 3.—Distribution of Materials in the Preclinical Sciences.

COLLECTION DEVELOPMENT AT THE NATIONAL LIBRARY OF MEDICINE

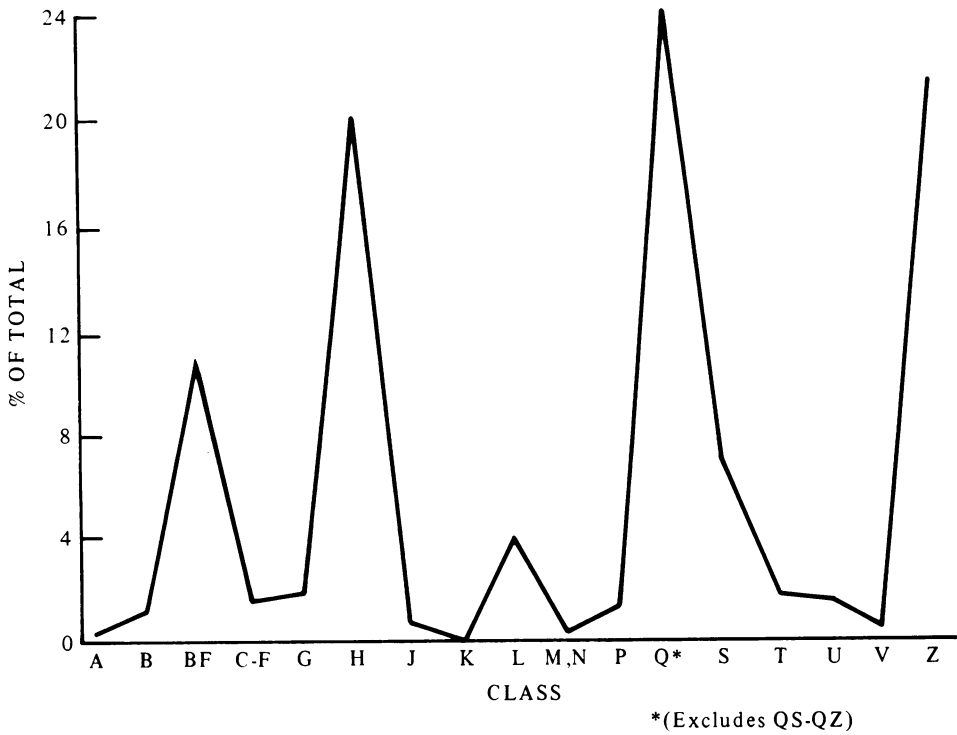


FIG. 4.—Distribution of Materials in the Related and Peripheral Classes (LC Schedule).

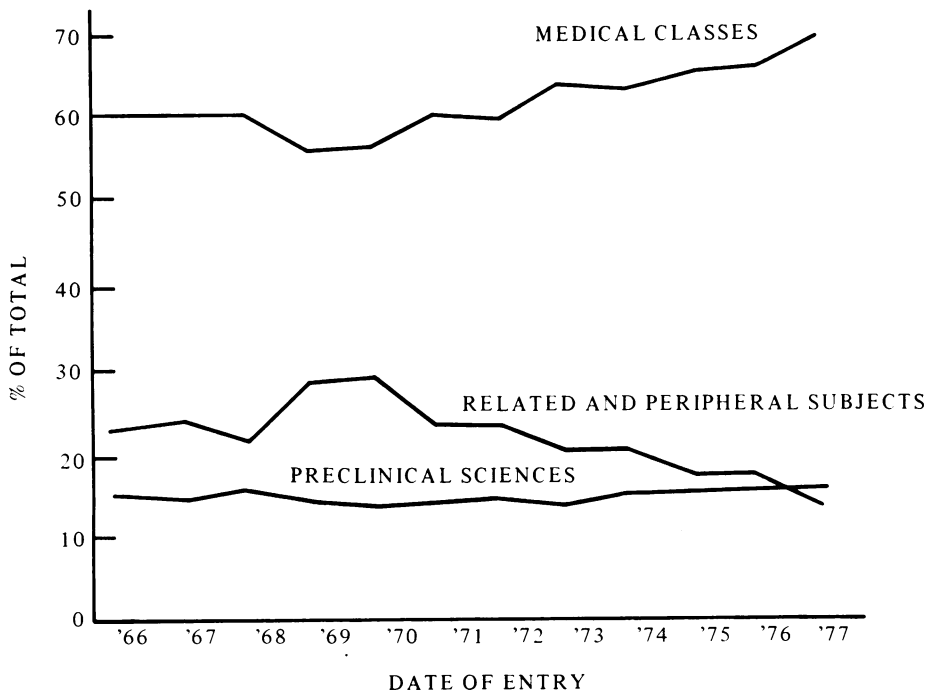


FIG. 5.—Subject Analysis of Monographs by Date of Entry into CATLINE.

TABLE 1
CLASSES SHOWING GREATEST DEGREE OF VARIABILITY 1966-1977

Class	Earlier Collecting Effort		Current Collecting Effort (1977) % of Year's Total	Change %
	Entry Date	% of Year's Total		
H - Social Sciences	1970	7.2	3.3	-3.9
Q - Science, General Aspects	1970	7.4	3.6	-3.8
W - Medical Profession	1972	3.1	6.4	+3.3
WM - Psychiatry	1968	5.9	8.7	+2.8
WY - Nursing	1970	2.0	4.7	+2.7
WZ - History of Medicine	1968	5.4	2.7	-2.7
S - Agriculture, Veterinary Medicine	1967	3.5	0.8	-2.7

represented classes, containing 5,905 and 5,522 titles, respectively. These three classes, together with BF - Psychology (3,033 titles) and S - Agriculture and Veterinary Medicine (1,969 titles), represented 85.5% of the noncore collection. The classes with the least number of items were M - Music (9 titles) and K - Law (12 titles).

SUBJECT ANALYSIS BY DATE OF ENTRY

This analysis examined the degree of change in the subject content of CATLINE over the years. The date-of-entry field, representing the date on which each cataloging record was input into CATLINE, was correlated with the broad subject groupings for monographs. Fig. 5 indicates that the percentage of monographs cataloged each year in the medical classes (W-WZ) has significantly increased over the years, while the percentage of titles in the related and peripheral subjects has markedly declined, especially in 1977. These trends correspond to increased emphasis on collection of publications dealing with health care delivery, hospital administration, and education of health care personnel, and decreased emphasis on collection of materials in noncore subjects, which followed the revision of the NLM *Scope and Coverage Manual*. The percentage of preclinical science monographs has remained stable over the years.

Each individual subject class was then analyzed in detail in terms of date of entry. The percentage of each year's total cataloging input was calculated for every subject class; then the years representing the lowest and highest percentages were determined. Overall there was very little variation from year to year in percentage of titles cataloged within each subject class. Table 1 indicates the seven subject classes that exhibited the greatest degree of variability over the years. The decrease in the

number of titles in the Social Sciences (H), the nonclinical sciences (Q), and Agriculture and Veterinary Medicine (S) may be explained by changes in collection policy. The increase in collection levels in Medical Profession (W), Psychiatry (WM), and Nursing (WY) may be attributed to a possible increase in the publication of literature in these areas, and/or improved acquisition and cataloging of these materials by NLM. The decrease in the collection in the History of Medicine (WZ) class may be an indication of decreased collection or cataloging of medical miscellany, which is included in this class. Further investigation of these trends is warranted.

LANGUAGE

All of the records in CATLINE representing NLM holdings were analyzed according to language. There were 177,045 values in the

TABLE 2
LANGUAGE ANALYSIS OF CATLINE

Language	Total No. of Records	% of CATLINE Total
English	102,113	57.7
German	17,208	9.7
Russian	13,475	7.6
French	12,805	7.2
Japanese	6,051	3.4
Italian	5,219	2.9
Spanish	5,200	2.9
Polish	2,432	1.4
Dutch	1,792	1.0
Portuguese	1,391	0.8
Chinese	1,376	0.8
Czech	1,303	0.7
All Others	6,680	3.8
(52 Languages)		
Total	177,045	100

COLLECTION DEVELOPMENT AT THE NATIONAL LIBRARY OF MEDICINE

language field indicating that 1% of the CAT-LINE records list more than one language. Table 2 indicates that more than half the titles cataloged since 1965 were in English (57.7%). German, Russian, and French language materials comprised an additional 24.5% of the data base. Foreign language items were then compared to English language items in broad subject and form

categories. In the preclinical sciences there were only slightly more English language items than foreign. For core medical monographs and serials (W1), the margin of English over foreign language titles was represented by a few thousand records. In one class (W6 - Pamphlets), there were almost twice as many foreign language titles as English titles. These comparisons are presented in Fig. 6.

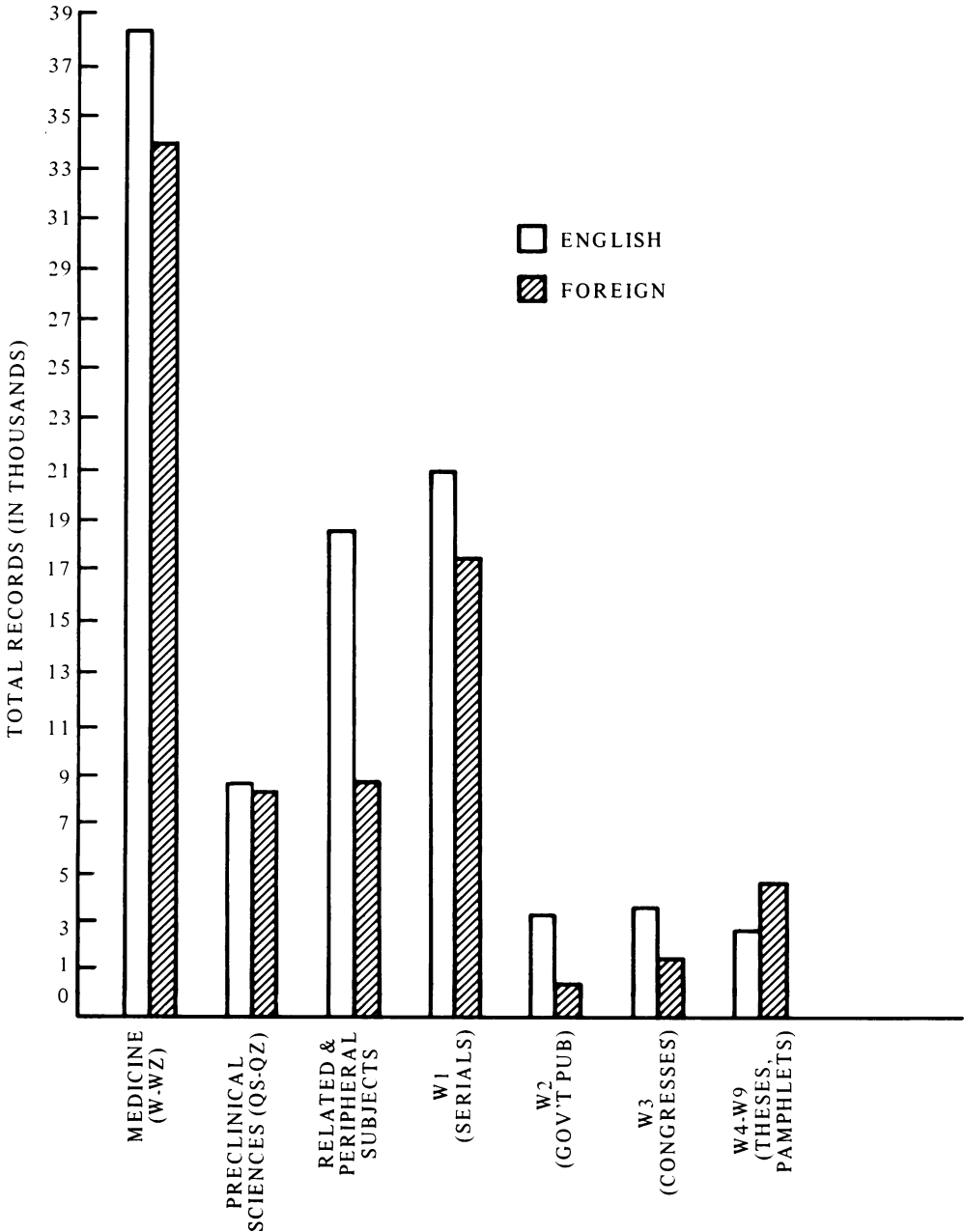


FIG. 6.—Comparison of English and Foreign Language Items by Broad Subject and Form Classes.

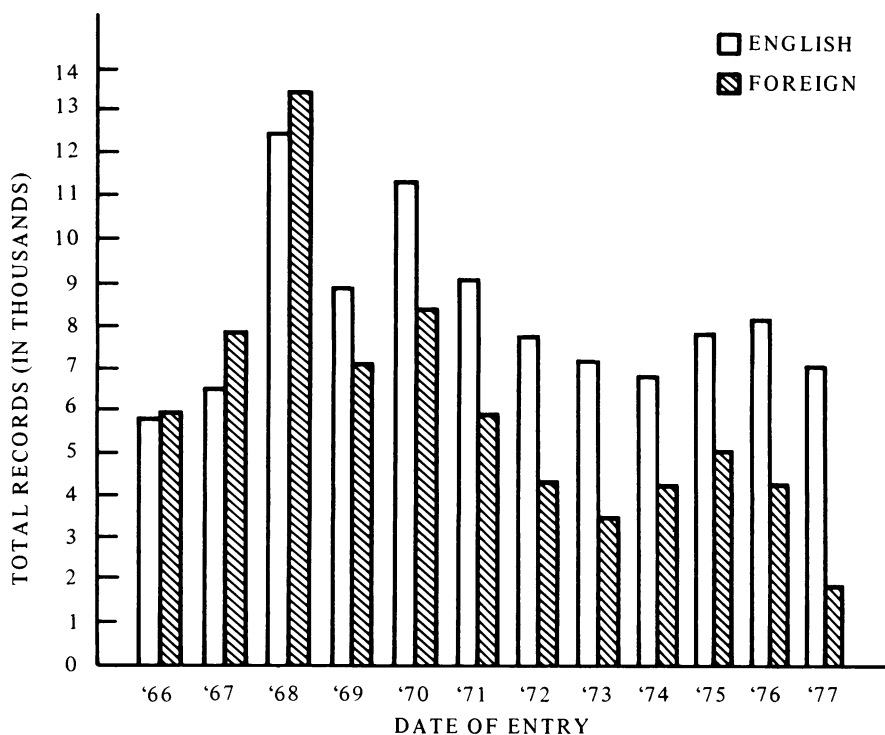


FIG. 7.—Comparison of English and Foreign Language Items by Date of Entry.

An analysis of English versus foreign language items by date of entry (Fig. 7) showed that the number of foreign language items being cataloged has steadily declined. In 1966–68 there were more foreign items cataloged than English, but beginning in 1969 English language materials received higher priority in cataloging, with the intent of supporting the need for cataloging copy in the biomedical community.

THROUGHPUT TIME

Analysis of throughput time for processing English language monographs was performed for selected years (1977, 1972, and 1968). These dates were chosen to represent the collection at different stages of development: current collecting efforts (1977), five years previous (1972), and early collecting efforts (1968). The year 1968 was specifically chosen for study since it was the calendar year preceding the development and implementation of the priority system for book processing.

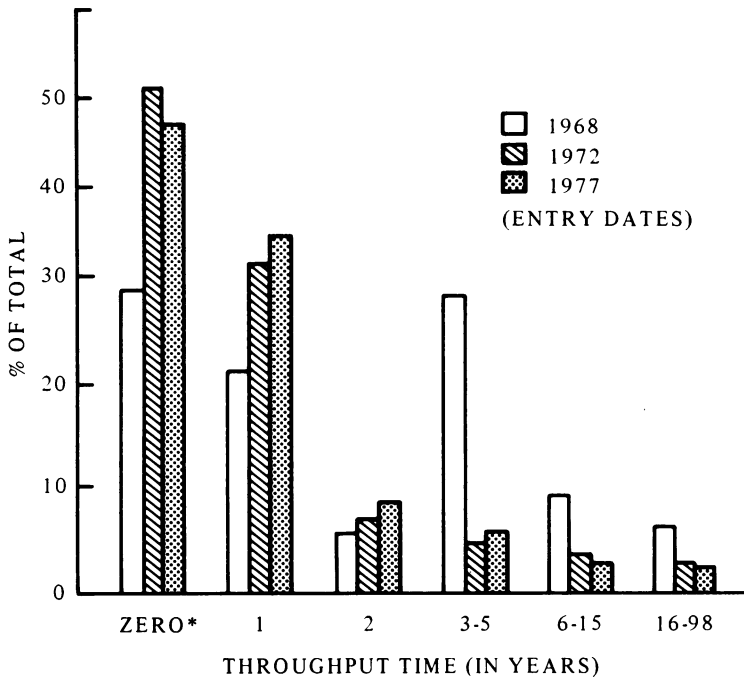
Throughput time was computed by subtracting the year of publication from the year of entry into CATLINE. Since possible delays in acquisition are not accounted for in this present study, the data represented in this analysis can be considered to be a composite of both acquisition and cataloging

throughput times. An analysis of acquisition throughput time and its related influence on cataloging throughput time is a topic for possible future study using other automated files that support technical services at NLM. Fig. 8 indicates that in 1977 and 1972 approximately 47–50% of the monographs were entered into CATLINE in the same year in which they were published, and that 31–34% were entered within the calendar year following publication. In 1968 a high percentage of retrospective cataloging was completed. Approximately 50% of the titles cataloged were two years or older, and only 29% were cataloged in the same year in which they were published. The priority system was developed to reverse this processing trend and to give higher priority to the current literature.

SUMMARY AND CONCLUSIONS

A recent article on collection evaluation by Mosher listed some of the important benefits to be gained from a well-planned program of collection evaluation. Since our present study of the NLM collection was a quantitative rather than qualitative analysis, many of the benefits relating to collection adequacy that he described were not realized. However, we now have a better under-

COLLECTION DEVELOPMENT AT THE NATIONAL LIBRARY OF MEDICINE



*Entry is the same as publication year

FIG. 8.—Throughput Time for English Language Monographs.

standing of the overall characteristics of the monograph collection and can begin “to focus human and monetary resources on collection areas most needing attention” [10].

Certain subjects in the core monograph collection have been identified as areas of apparent low collecting activity. Day, Bowden, and Kronick have studied some of these subject areas for selected years [11]. Further study is needed, particularly in terms of the publishing activity in each of these subjects. Shifts in collecting levels in some subject areas have been noted in recent years, but the areas exhibiting change were the subject areas for which the library has deliberately changed its collection and cataloging practices. The revision of the NLM scope and coverage guidelines has indeed effected a change in the library’s monograph collection in medical subjects, so our hypothesis was shown to be correct. The language distribution of the collection has changed discernibly since 1969 when the cataloging of English language publications was assigned top priority. Reduction in the number of foreign language titles cataloged was related to a decrease in the number of catalogers with expertise in the less widely known foreign languages. The analysis of throughput time indicated the positive effect of the priority system

established in 1969 for cataloging recently published materials on a higher priority basis than older materials. The hypothesis that processing changes and manpower limitations would affect the growth patterns of the collection in terms of both language and currency was supported by our data.

Since the findings from this analysis of CATLINE have already proved useful to NLM management staff, it is reasonable to project that subsequent studies, if performed at regular intervals, will prove worthwhile. Changes in procedure at the outset of future studies (that is, elimination of subject-classed serials and CIP records) could provide for better segregation of data than we were able to achieve. Deletion from CATLINE of duplicate records for CIP titles was completed before our study, so we did not encounter the problems of data validity cited by Day, Bowden, and Kronick. Analyses of the NLM serial collection and the uncataloged numbered arrearage are necessary in order to provide a more complete description of the library’s collection development activities in recent years.

The increase in the number of libraries with machine-readable holdings records may lead to an increase in the number of collection evaluations

performed in the future. The use of computers to summarize and correlate data and to provide access to the cataloged holdings of large research libraries allows for both quantitative analysis of an individual library's collection and qualitative comparison of that collection with the collections of other libraries. The data gained from such analysis can provide librarians with definitive information about the one thing that is the very heart of all libraries, the collection.

ACKNOWLEDGEMENTS

The authors thank Dan Tonkery, former chief of the Technical Services Division, NLM (now associate university librarian for technical services, UCLA), for his valuable suggestions on the preparation of this article, and Ted Catchings, applications programmer, for technical support in creation of the INQCAT file.

Other NLM staff deserve thanks for their support at various stages during this study: Lillian Kozuma, Sally Sinn, and Carolyn Davis of the Cataloging section, and Dan Carangi, graphics director.

REFERENCES

1. LANCASTER, FREDERICK WILFRED. *The Measurement and Evaluation of Library Services*. Washington, D.C., Information Resources Press, 1977. p. 165-206.
2. BONN, G. S. Evaluation of the collection. *Libr. Trends* 22: 265-304, Jan. 1974.
3. CASSATA, M. B., and DEWEY, G. L. The evaluation of a university library collection; some guidelines. *Libr. Resour. Tech. Serv.* 13: 450-457, Fall 1969.
4. WEBB, W. H. Project CoED: a university library collection evaluation and development program. *Libr. Resour. Tech. Serv.* 13: 457-462, Fall 1969.
5. CLAPP, V. W., and JORDAN, P. J. Quantitative criteria for adequacy of academic library collections. *Coll. Res. Libr.* 26: 371-380, Sept. 1965.
6. METCALF, KEYES D., ET AL. *The National Medical Library: Report of a Survey of the Army Medical Library*. Chicago, American Library Association, 1944. p. 19-21.
7. BYRD, G. D. The shelflist as an administrative and acquisitions tool. In: *Abstract Proceedings of Medical Library Association Seventy-fifth Annual Meeting, June 12-17, 1976, Minneapolis, Minnesota*.
8. KRONICK, D. A., and BOWDEN, V. M. Management data for collection analysis and development. *Bull. Med. Libr. Assoc.* 66: 407-413, Oct. 1978.
9. NATIONAL LIBRARY OF MEDICINE, TECHNICAL SERVICES DIVISION. *Scope and Coverage Manual of the National Library of Medicine*. Bethesda, Maryland, May 1977.
10. MOSHER, P. H. Collection evaluation in research libraries: the search for quality, consistency and system in collection development. *Libr. Resour. Tech. Serv.* 23: 16-32, Winter 1979.
11. DAY, R. A., BOWDEN, V. M., and KRONICK, D. A. Comparison of holdings of NLM (CATLINE) with those of resource libraries. *Bull. Med. Libr. Assoc.* 67: 25-30, Jan. 1979.

Received July 1, 1979; accepted September 25, 1979.