

Marcos A Rodriguez. Knowledge Discovery in a Review of Monograph Acquisitions at an Academic Health Sciences Library. A Master's Paper for the M.S. in I.S. degree. March, 2008. 45 pages. Advisor: Diane Kelly

This study evaluates monograph acquisition decisions at an academic health sciences library using circulation and acquisitions data. The goal was to provide insight regarding how to allocate library funds to support research and education in disciplines of interest to the library user base. Data analysis revealed that allocations in 13 subject areas should be reviewed as the cost of circulation was greater than the average cost of circulation of the sample and the average cost of monographs was higher in these subject areas than the average cost of monographs in the sample. In contrast, 13 subjects returned cost of circulation rates lower than the average cost of circulation of the sample. These subjects merit stable budget allocation or increased allocation depending upon collection needs. Overall, this study found that this library is allocating a majority of resources to subjects with above average rates of use.

Headings:

College and university libraries – Acquisitions

Medical libraries and collections – Collection development

Acquisitions/Evaluation

Knowledge Management

Decision support systems – Case studies

Information systems -- Statistics

KNOWLEDGE DISCOVERY IN A REVIEW OF MONOGRAPH ACQUISITIONS AT
AN ACADEMIC HEALTH SCIENCES LIBRARY.

by
Marcos A Rodriguez

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information Science.

Chapel Hill, North Carolina

April 2008

Approved by

Diane Kelly

INTRODUCTION

Taking advantage of technological advances in content management systems, a large number of academic libraries have adopted integrated library systems within the last 10 years. These academic institutions have implemented these systems with the intent to streamline and automate the acquisition, cataloging, and management of traditional and electronic collections that had previously been performed in separate systems or manually. Over the course of this transition, “the average ARL library would have needed to spend nearly 45 percent more in 2003 to cover the monographic market than would have been necessary in 1994” (Stoller, 2006, p. 49). This inflation in the prices of monographs has been met with an average of 39.5 percent increase in monograph expenditures over that same period, “suggesting that ARL libraries are falling behind” (Stoller, 2006, p. 49). Studies by Webster (1993), Crotts (1999), Wise & Perushek (2000), Agee (2005) and Knievel et. al. (2006), all discuss the issue of increased costs in light of cyclical, static, or even reduced budgets for materials acquisitions in discussions on identifying new ways to assess collection development practices.

In response to this challenge, libraries and collection development research have started to rely more on statistics based models and goal programming based approaches to collection development (Kao, 2003, p. 134). Previous research using computerized library system data for collection development has explored the use of aggregated circulation information or a combination of circulation and budget expenditure information divided by subject area to inform collection management decisions. Facing

limited resources and increased costs, the impetus has been on academic libraries to efficiently acquire resources to support education and research. At the Duke University Medical Center Library, specific methods that have been employed have included: collection reviews involving input from library users, reviews of authorized lists of core titles in specific disciplines such as Doody's List of Core Titles and Brandon Hill lists, statistics of online content use, and journal impact factors to evaluate collection development activities. At present, this library is exploring the use of acquisitions and circulation data gathered from the integrated library system to feed into an evaluation of the monograph collection development process.

The field of knowledge management is concerned with utilizing technology and human ability to create, distribute, renew, and apply knowledge through knowledge discovery to allow an organization to adapt to changes in the environment in which it operates. (Malhotra 1998) Knowledge discovery in the context of this study is considered, "the extraction of knowledge from data warehouses by building information from a series of patterns produced by a knowledge-based system" (Baskerville, 2006, p. 97). Research that has used knowledge management methodology in the context of library decision making has focused on optimizing budget allocations in light of considerations that, "the budget is increasingly limited" (Wu, 2003, p. 401), and "utilization of materials . . . should be able to reflect the final allocation acquisition budget," in terms of relative expenditures (Kao, 2003, p. 134).

This analysis will serve as a case study to introduce a knowledge management framework into a collection development review process at the Duke University Medical Center Library. Utilizing technology and human ability to create, distribute, and apply

knowledge, the expectation was to assist the library organization adapt to increased monograph costs. Therefore, this study involved going through the process of data preparation, data selection, data cleaning, incorporating appropriate prior knowledge and proper interpretation of the results through finding useful patterns in the data. This process has been defined as knowledge discovery in databases (Fayyad, 1996, p. 28). such, this analysis was intended to allow the library to build information from a series of statistical patterns retrieved from the integrated library system. Therefore, an argument can be made that this study utilized a knowledge management framework using statistical analysis as a form of data mining in a review of collection development activities.

Given the increased costs of developing and maintaining academic library collections, an analysis of collection management and usage information from integrated library system data records may provide insight regarding how to efficiently allocate limited funding to support research and education in disciplines of interest to the library user base. Following, the research question guiding this effort was: Is the library allocating its financial resources in a manner that provides levels of use that support continuing with collection building that mirrors past decisions? The future holds continued development of integrated library systems, budget challenges and organizational change for libraries. Therefore, continued exploration of how library computer systems may be utilized by libraries to assist with management decisions for collection development is a worthwhile endeavor.

RELATED WORK

Morse (1968), Simmons (1970), Jenks (1976), and Lancaster (1982) conducted some of the earliest studies that examined data sets gathered from electronic library systems to evaluate collections management activities. They also provided early lessons in utilizing statistics in for this purpose. Morse developed one of the first statistical models of circulation activity in relating Markov processes to book circulation histories at the M.I.T. science library. In his analysis of 9 years of circulation data he found that, “the expected circulation next year of a book . . . appears to be roughly .4 plus about a half of its last-year’s circulation, independent of the age of the book (at least out to an age of 5 years)” (Morse, 1968, pp. 93). Likewise, Simmons conducted a study that looked at circulation of materials over a semester to analyze what additional copies should be purchased. His findings lead him to suggest that the, “most effective role of comparative analysis (of material circulation) may be to illustrate patterns of use rather than circulation history of individual volumes” (Simmons, 1970, pp. 62). From these studies, an interest in assessing circulation of materials by subject areas would become a common research method and was adopted for this research effort to provide a logical breakdown of materials for specific medical disciplines.

Jenks (1976) introduced the use of Library of Congress classifications of books in a study that compared relative use of books across academic departments at Bucknell University. His analysis provided information relating the subject matter of monographs and their circulation yet he limited his recommendations to performing follow-up evaluations of the collections for academic departments found to have high and low usage. Expanding upon the framework introduced by Jenks, Lancaster (1982) included

evaluation of holdings in particular subject areas in a framework for evaluating collection building by usage. One method he proposed was to analyze the percentage of overall holdings in each subject area versus the proportion of total circulation to calculate underuse and overuse data for each subject. In comparing actual relative use of materials versus an expected rate of usage, he proposed a metric for evaluating collection development using circulation data broken down by subject area. For this study, a metric for computing expected budget allocation using the mean cost of monographs purchased was used in a similar manner to evaluate collection development in terms of actual versus expected cost of use by subject.

Among the earliest literature exploring the potential for using computerized library systems in library decision making, Edwin Cortez (1983) proposed organizational management decision making that utilizes information gathered from such systems. In his discussion, Cortez posits that evaluation of automated library systems should be conducted in the context of both how, “effectively they handle day-to-day operations,” and “their ability to manipulate and generate information for management” (Cortez, 1983, p. 22-24). Reed-Scott also argued for the benefits of using computer systems for macro management decision making in that collection management information systems would be essential for, “collection managers to exploit machine-generated data for improved decision-making and effective use of collection resources” (Reed-Scott, 1989, p. 48).

Analyses by Hawks (1988) and Knutter (1987) also discussed the potential for using computerized systems in management decision making. However, their frameworks provided detail at the level of library functional areas, including collection

development. In discussion on collection development, Hawks described the potential for using information for circulations and patron material requests to support purchase decisions in that, “usage may warrant consideration for future allocations to subject areas in high demand” (Hawks, 1988, p. 133). With respect to acquisition expenditures, Knutter discussed the potential for gathering data on collection growth over time, detailed financial information, and data related to who made purchasing decisions (Knutter, 1987, p. 137).

Despite this optimism, research on this topic also reflected technological and organizational limitations that prevented the utilization of library computer systems in the manner described above. Knutter discussed the risk of information overload as an organizations’ ability to collect, organize, and manipulate data far outstripped their ability to interpret and to apply them (Knutter, 1987, p. 143). “The practical problem of digesting the massive amount of data generated by these systems has not been dealt with effectively,” as well (Reed-Scott, 1989, 48-49). In a follow-up analysis, Hawk reflected on limitations of computer systems to capture all manner of circulation activity and the need for manual statistics generation to, “yield the information needed as standard reports may be unsuitable for the purpose at hand,” due to system inflexibility and lack of functionality” (Hawks, 1992, p. 15).

In her analysis, Knutter also considered factors influencing a library’s ability to use circulation data for collection development decision making. These factors included the comprehensiveness of the data, the collection of in-house use statistics, and the inclusiveness of collections in the computer systems, and the availability of programs to compile, manipulate, and analyze the use and user data (Knutter, 1987, p. 133). In the

course of this research project, the challenges and limitations mentioned by the research related to the quantity and quality of data as well as suitable software applications to retrieve and organize data had implications for the resulting analysis.

The management oriented literature mentioned above was supplemented by research that focused specifically on using electronic circulation information to inform collection development practices. Day & Revill (1995) conducted an analysis using circulation data to analyze the average use of materials purchased and compared the proportion of purchases in particular subject areas that circulated. In their study, they were able to “provide data on the performance of individual items and help to better match library acquisitions to demand,” that enabled them to, “more strongly justify our share of the University’s budget” (Day, 1995, pp. 156). Similar to Jenks’ (1976) work, Crotts (1999) conducted a study that explored interrelationships between circulation, expenditures and student enrollment by subject area to develop a model for allocating subject funding for monographs. Using a cost/usage variable for each subject compared against an average demand value calculated using data over a five year period, Crotts recommended budget allocations that present a, “realistic level of expenditure for materials in relation to usage” (Crotts, 1999, pp. 270). This evaluation metric was adopted for this study to compare cost per use of materials in each subject area with an average cost per use statistic for all monographs purchased by the library.

Within a medical library context, Kraemer (2001) conducted a study that analyzed circulation data in relation to average cost of monographs purchased in particular subject areas. Of interest is that Kraemer introduced consideration for the types of books within subject areas to potentially allocate more funding based upon analysis of relative usage of

monographs both within and across subjects. Utilizing more formal statistical methods, Chen (1997) incorporated circulation data in a data analysis framework for library management to score library resource use efficiency and Wise & Perushek (2000) utilized a goal programming framework that utilized counts of monographs purchased in subject areas and percentage of overall circulation by subject area to inform collection management planning. Studies conducted by Aguilar (1986), Knievel et. al. (2002), and Ochola (2006) also incorporated counts of item circulation in subject areas but compared those with the ratio of interlibrary loans versus holdings in subject areas as measures of use in collection development analysis. Each of these studies reflected an increased interest in directly link circulation statistics and budget allocation, which was the motivation for this research effort.

In light of this body of literature exploring the use of circulation data, there is continued resistance to using automated system generated data in evaluating collection development practices. Carrigan (1996) conducted a study of collection development officers at 79 ARL member libraries that revealed of the 45 responding libraries did not use data produced by automated circulation systems due to factors ranging from limitations of the system to not being convinced of the value of the data gathered (Carrigan, 1996, p. 434). Casserly & Ciliberti's (1997) survey of 49 collection development librarians at academic libraries using automated library systems revealed that system derived data was found to be less useful than available and computer systems were, at the time, not able to provide the same quality of data gathered manually regarding complex aspects of system use (Casserly, 1997, p. 79).

Despite this resistance, Peters (1996) and Atkins (1996) continued the tradition of supporting the use of library computer systems to support management and collection development begun in the previous decade. Peters conjectured that the movement to utilize systems in this manner was at that point a grassroots movement rather than a management tool and expounded upon the potential for improving the automated systems and, in the context of collection development, enabling expression of need, through circulation, to drive some collection development activities (Peters, 1996, pp. 21-23). Atkins mirrored this sentiment in arguing that only in libraries, “where freedom to experiment and hire programmers has existed has the full potential of automated systems to provide library management statistical data been realized” (Atkins, 1996, pp. 16). Subsequent arguments for the use of statistics ranged from issues related to, “the cost of books increasing . . . and with no end in sight, it becomes most obvious that subject allocations cannot continue to be based on precepts unsupported by the actual demand for materials” (Crotts, 1999, p. 271) to “usage data are even more important in light of remote storage facilities and the attendant storage decisions that have been adopted by many U.S. libraries” (Kniewel, 2006, p. 49). Of note in Atkins’ analysis, his discussion covered the potential of data mining of automated systems for collection management and planning. In this regard, his research bridged previous applied research and recent research that has incorporate knowledge management methodologies to inform library collection development decision making.

The knowledge management research field has roots in information economics and organizational strategy research in the mid 1990s and has moved from “buzzword” status to a position of practical intellectual strength for management (Baskerville, 2006,

pp. 86, 84). The field is generally focused on exploring the “synergy of data and information processing capacity of information technologies, and the creative and innovative capacity of individuals.” (Malhotra, 1998) A sub-discipline within knowledge management is data mining, which is concerned with using large stores and flows of data that are available for decision making. Further, “these stores and flows can be used for knowledge ‘discovery’ through the means of complex tools to aid in the logical and practical digesting of data into information,” (Baskerville, 2006, p. 96). From this perspective, statistical analysis of integrated library system data may be considered a form of data mining in that the purpose is to gather, process, analyze, and generate information to inform collection development decisions. However, research that has applied data mining in the context of libraries has involved the development of automated agents or algorithms to facilitate data analysis of large quantities of data. Banerjee (1998) presented one of the first discussions for use of data mining in library management as he reflected on requisites for successfully utilizing data mining. He also raised issues related to lack of standards and technological hurdles to implementation (Banerjee, 1998, p. 30-31). Guenther (2000) discussed the use of data mining in a health sciences library and evaluated the requisite technologies and strategies necessary to apply data mining within a library setting. Noteworthy was her discussion on making data application neutral to facilitate importing data into a single database for analysis (Guenther, 2000, p. 62). In this analysis, use of an integrated library system provided a common framework that facilitated the collation of acquisitions, cataloging, circulation and other data collection systems into one dataset.

Literature involving application of data mining and knowledge discovery into studies analyzing library collection development practices has emerged in the last five years. Nicholson (2003), Nicholson & Stanton (2004), and Nicholson (2006) developed and expanded a framework termed bibliomining, which is data mining specifically to examine library data records (Nicholson & Stanton, 2004, p. 248). At the core of this framework is the concept of a central data warehouse on a computer system to organize the collection, organization, and analysis of data gathered from all of a library's computer systems. Citing resistance by integrated library system vendors to provide sophisticated analytical tools that would promote useful access to raw data, Nicholson's main contention is the importance for libraries to create data warehouses that permit queries and matches across multiple heterogeneous data sources. Nicholson argued that "only by combining and linking different data sources can managers uncover the hidden patterns that can help the understanding of library operations and users" (Nicholson, 2004, p. 251-252). With respect to collection development, bibliomining,

may provide insight as to how those items got into the library. By looking for correlations between low-use items and subject headings, publisher, vendor, approval plan, date, format, acquisitions librarian, collection development librarian, library location and other items, managers might discover problem areas in the collection or organization (Nicholson, 2004, p. 255).

Kao et. al. (2003), Wu (2003), and Wu et. al. (2004) also developed a knowledge management framework that utilizes data mining of circulation data to assess use of materials by particular academic departments in their subject areas. Kao et. al. introduced this information into a budget allocation model that derived relative expenditures in different subject areas based upon the analyses of the circulation data. In a follow-up study, Wu (2003) incorporated additional pre-processing of data and

weighted calculations of subject usage by departments versus the concentration of purchases in subject areas to calculate budget allocations. Wu et. al. (2004) completed a follow-up study that explored material acquisitions in the context of specific departments as opposed to relative comparisons across departments. By analyzing the relative use of subject materials, the goal was to predict user needs that could be used by librarians to reflect actual needs when acquiring materials. (Wu, 2004, p. 723) At this time, the results are inconclusive and further research is necessary to realize the goals set forth by these researchers.

At this time, research focused on using data mining to inform collection development decision making is still in early stages of theory and methodology development. In contrast, research that utilizes statistical analysis to inform collection development decision making has a longer tradition of demonstrating the use of complex tools to aid in the logical and practical digesting of data into information in the context of libraries and should not be abandoned in light of the potential for data mining via algorithms or automated agents. In his discussion Wu (2004) reflected on an important consideration for using automated data mining.

With regard to the application of knowledge discovery in databases, data preparation is an important process in order for the discovering mechanism to perform. In spite of many knowledge discovery tools available . . . this process is a highly domain-specific task that may require domain knowledge and a large amount of time to accomplish (Wu, 2004, p. 723).

In contrast to automated data mining techniques, statistical analysis is more readily applicable in a variety of contexts for evaluation. Given the state of the research literature in moving beyond statistical analysis to produce automated metrics to inform collection

development decisions, the statistical analysis in this study seeks to bridge the ideologies of statistical based research and automated data mining research.

METHODS

This study makes use of acquisitions, cataloging, and circulation statistics data gathered from an integrated library software system. For the purposes of this study, acquisitions data was defined as information related to the order and purchase of materials including order date, order type, and purchase price. Cataloging information was defined as information related to the bibliographic information assigned to materials such as call number, collection, and enumeration information such as volume and copy number. Circulation statistics was defined as events logged in the circulation system as the check-out of materials to library users. Data for three fiscal years spanning from July 1, 2004 to June 30, 2007 were selected for this analysis.

Following retrieval from the system, cataloging and circulation data were combined with the acquisitions information to create a properly formatted dataset with expenditure information, catalog classification information, and circulation statistics. The integration of this data was chosen because acquisitions and cataloging information were not sufficient to properly identify materials and link circulation information to materials in the sample. Additionally, the acquisitions data did not completely reflect all library acquisitions during the period of interest due to changes in staffing and workflow patterns. Use of the cataloging information allowed for remediation of a majority of issues related to data cleanup. Following data cleanup, the focus of the analysis was on monograph expenditures for items in the general circulating collection; therefore, several filters were utilized to restrict the dataset to appropriate records for analysis.

The first filter removed all items donated as gifts to the library collection as well as materials acquired from budget funds separate from the fund for monographs. These materials included serials and standing orders and history of medicine materials. The second filter removed materials with non-standard circulation policies, including electronic books, materials purchased for reserves and reference collections, and materials purchased for library staff use. The third filter removed materials collected that were not of interest in the context of this analysis. These materials included graduate and doctoral theses for supported academic departments and materials collected for the leisure reading collection that are not cataloged using Library of Congress or National Library of Medicine classifications. The resulting dataset for analysis contained 1365 items in 10 Library of Congress classes and 35 National Library of Medicine classes. To facilitate data analysis, the 18 items classified using Library of Congress subject headings were combined into one data group.

EVALUATION METRICS

This research proposal utilized statistical analysis of circulation and acquisitions information as a means for introducing a knowledge management framework in the assessment of budget allocations and expenditures for monographs in one academic health sciences library. For this analysis, one of Crott's (1999) measures for computing "costs" of circulation was used to compute an average cost of circulation for each subject area in the sample. In Crotts' analysis, he calculated the ratio of expenditure to circulation of materials in each subject as well as the number of books circulated per dollar expended (Crotts, 1999, p. 267). The ratio of expenditure to circulation was adopted for this study as an actual cost of use measure (ACU). See (1) on next page.

$$\text{ACU} = \frac{\text{Budget Expended on Subject}}{\text{Number of Circulations within Subject}} \quad (1)$$

In Crott's analysis, the lower the average cost per use of materials in specific subject areas relative to the average cost per use of the entire sample indicated a positive rate of return for the funds allocated by the library (Crotts, 1999, p. 267). In contrast, higher average cost per use indicated a high level of expense in purchasing materials in that subject area in relation to the user demand. Similarly, this study will compare at actual cost per use measure (ACU) to the average cost per use of the sample to determine which subjects are, "less costly or more costly to circulate" (Crotts, 1999, p. 267). A significant limitation in Crotts' analysis was related to his not addressing issues related to differences in costs of monographs across subjects.

To account for differences in cost of monographs across subjects, a measure using the mean cost of items across the sample instead of actual monograph prices was used as a baseline by which to compare actual cost per use across subject areas. To compute this measure for each subject area, the mean cost of the sample was first multiplied by the number of items purchased in a subject area to generate an expected budget expended on subject. See (2) below.

$$\text{Expected Budget Allocation on Subject} = (\text{Average cost of items in entire sample}) \times (\text{Number of items purchased in subject}) \quad (2)$$

The result was then divided by the total circulation of items in the subject to produce an expected cost of circulation statistic (ECC). See (3) below. As with the ACU measure, the higher the value of ECU, the higher the expected cost of circulation for a subject. To

compare relative costs across subjects the ACU was compared to the ECU for each subject.

$$\text{ECU} = \text{Expected Cost of Circulation} = \frac{\text{Expected Budget Expended on Subject}}{\text{Number of Circulations within Subject}} \quad (3)$$

In this analysis, the actual cost of use measure (ACU) for each subject was compared with an expected cost of use measure (ECU) for each subject. Subtracting ACU from ECU produced a measure that indicated whether the actual cost of circulation for a subject was higher or lower than that predicted by the expected cost of circulation measure. This resulting statistic served as a moving baseline by which to compare average costs of monographs across subjects.

The values for ACU yielded an indication of the relative strength of the dollar in terms of circulation demand for books within a subject similar to that calculated by Crotts in his analysis. Subjects with actual cost of use less than the mean actual cost demonstrate a strong user demand in relation to cost whereas subjects with actual cost of use more than the mean actual cost of use demonstrate weaker user demand in relation to cost. Further, the values for ECU yielded an expected value of the relative strength of the dollar in terms of circulation demand for books within a subject derived from the mean cost of monographs in the entire sample.

Further, for subjects in which $\text{ECU} - \text{ACU}$ is positive, the average cost of materials in the subject was shown to be lower than the average monograph cost calculated from the overall sample. Inversely, for subjects in which $\text{ECU} - \text{ACU}$ was negative, the average cost of materials in the subject was shown to be lower than the average monograph cost calculated from the overall sample. At the same time, the sign of

the difference between ECU and ACU indicated whether monographs in a particular subject were more (if positive) or less (if negative) than the sample mean cost. Therefore, in relating this data to collection development decisions, materials purchased in subjects demonstrating weaker user demand and higher average costs should be reviewed for applicability of those materials purchased for the library user base. Additionally, decisions on materials purchased in subjects demonstrating stronger user demand should be reviewed for possible increase in budget allocations to support user demand in light of higher or lower average material costs. See Table 1.

Table 1. Proposed Breakdown of Subjects Areas by Average Cost and Rates of Use

ECU – ACU value positive	Subjects with higher average costs and higher average rates of use <i>Consider for increased allocation.</i>	Subjects with higher average costs and lower average rates of use <i>Consider for decreased allocation.</i>
ECU – ACU value negative	Subjects with lower average costs and higher than average rates of use <i>Consider for increased allocation.</i>	Subjects with lower average costs and lower than average rates of use <i>Consider for decreased allocation.</i>
	ACU value lower than mean	ACU value higher than mean

RESULTS

The following section will detail the procedures for collecting and analyzing the circulation and acquisitions data in this study. As mentioned, this analysis was selective and included only circulating items in the main library collection with LC and LM classifications found in the integrated library system and were purchased between July 2004 and June 2007. The items that met these criteria numbered 1376 with a total count of 4544 circulations when the data was collected in February 2008. Descriptive information and statistics for these items, including breakdown by subject area, expenditures by subject area, and circulations by subject area are listed in Table 2.

As shown in Table 2, WG - Cardiovascular System, WE - Musculoskeletal Diseases and WL - Nervous System materials returned the highest number of circulations. WE, WG, and WL also accounted for the largest proportion of budget expenditure in the sample as well as the largest proportion of monographs purchased. Of interest is that QS - Human Anatomy, QV - Pharmacology, WX - Hospitals & Other Health Facilities, and LC items returned high numbers of circulations relative to the number of items and the mean item count, mean expenditure, and mean circulation across the sample were equal.

ANALYSIS

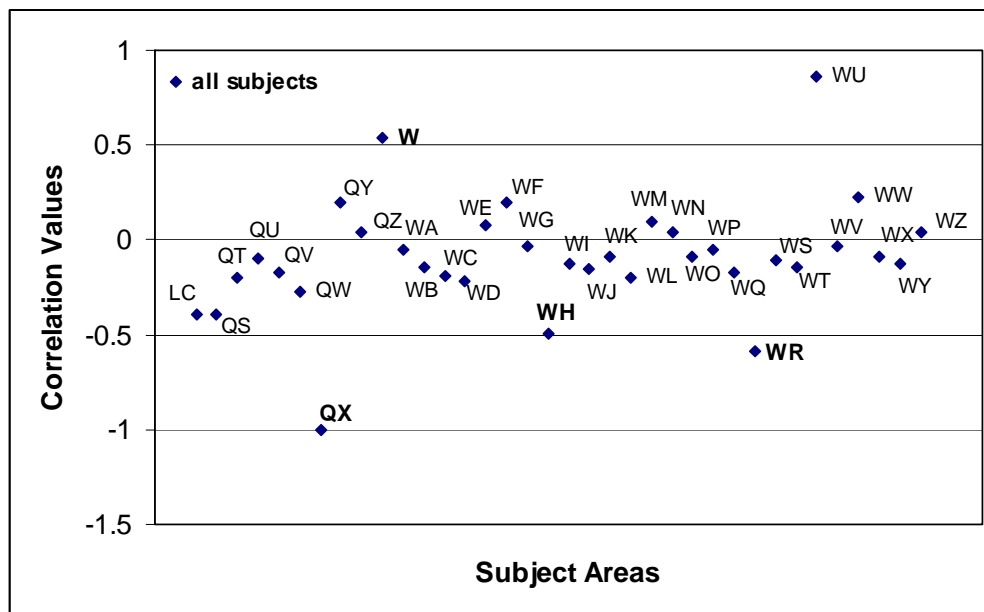
As an initial analysis, two-tailed Pearson correlations were performed on the expenditures and circulations across the entire samples and then across the individual expenditures and circulations of monographs within each subject area. The intent was to find out whether there is a correlation between both variables in this sample. The p value

Table 2. Purchases, Expenditures, and Circulation Data

Subject	# of Items	% of Items	Expen. in dollars	% of Expen.	# of Loans	% of Loans
LC books	18	1.319%	2487	1.913%	170	3.741%
QS – Human anatomy	35	2.564%	2097.2	1.613%	261	5.744%
QT – Physiology	21	1.538%	2391.4	1.840%	56	1.232%
QU – Biochemistry	14	1.026%	2088.7	1.607%	46	1.012%
QV – Pharmacology	27	1.978%	2055.36	1.581%	122	2.685%
QW – Microbio. & Immun.	20	1.465%	1719	1.322%	41	0.902%
QX – Parasitology	2	0.147%	241	0.185%	4	0.088%
QY – Clinical Pathology	16	1.172%	1378.38	1.060%	32	0.704%
QZ – Pathology	78	5.714%	8329.17	6.407%	235	5.172%
W – Health Professions	64	4.689%	3456.51	2.659%	211	4.643%
WA – Public Health	61	4.469%	3222	2.479%	206	4.533%
WB – Practice of Med	81	5.934%	5259.01	4.045%	284	6.250%
WC – Commun. Diseases	19	1.392%	1604	1.234%	52	1.144%
WD – Dis. of Systemic, Metabolic, or Env. Origin	14	1.026%	1534.42	1.180%	20	0.440%
WE – Musculosk. Dis.	117	8.571%	14748.77	11.345%	352	7.746%
WF – Respiratory Dis	31	2.271%	3665.69	2.820%	103	2.267%
WG – Cardio. System	109	7.985%	10816.57	8.321%	435	9.573%
WH – Hemic & Lymph. Sys	20	1.465%	2783.62	2.141%	52	1.144%
WI – Digestive System	29	2.125%	4576.85	3.521%	68	1.496%
WJ – Urogenital System	22	1.612%	2822.17	2.171%	56	1.232%
WK – Endocrine System	10	0.733%	842.14	0.648%	30	0.660%
WL – Nervous System	90	6.593%	10655.9	8.197%	323	7.108%
WM – Psychiatry	68	4.982%	4889.3	3.761%	165	3.631%
WN – Rad./Diag. Imaging	31	2.271%	2992.85	2.302%	159	3.499%
WO – Surgery	45	3.297%	5379.4	4.138%	126	2.773%
WP – Gynecology	25	1.832%	3321.94	2.555%	64	1.408%
WQ – Obstetrics	21	1.538%	2027.36	1.560%	49	1.078%
WR – Dermatology	12	0.879%	1274.5	0.980%	36	0.792%
WS – Pediatrics	74	5.421%	6733.11	5.179%	214	4.710%
WT – Ger./Chronic Dis.	20	1.465%	1156.89	0.890%	39	0.858%
WU – Dentistry/Oral Surg.	3	0.220%	378	0.291%	9	0.198%
WV - Otolaryngology	8	0.586%	956.56	0.736%	5	0.110%
WX - Ophthalmology	25	1.832%	3147.69	2.421%	96	2.113%
WX –Hospitals/Other Health Facillities	18	1.319%	2949.72	2.269%	127	2.795%
WY – Nursing	110	8.059%	5809.52	4.469%	280	6.162%
WZ – History of Medicine	7	0.513%	205.37	0.158%	16	0.352%
TOTALS	1365	100%	129997.60	100%	4544	100%
Mean	37.92	2.778%	3611.03	2.778%	126.22	2.778%
Standard Deviation	32.58	2.387%	3217.141	2.475%	112.0327	2.466%
Maximum Value	117	8.571%	14748.77	11.345%	280	6.162%
Minimum Value	2	0.147%	205.37	0.158%	4	0.088%

returned for the sum of expenditures and circulations was 0.836 and was significant at the .01 level. Therefore, there is a correlation between circulations and expenditures in the overall sample. However, correlations performed on monographs within each subject area returned p values that were significant at the .05 level for only 5 subject areas. They were QS ($p = -.397$), QX ($p = 1.00$), W ($p = 0.537$), WH ($p = -.497$), and WR ($p = -.583$). A list of the calculated p values are listed in Table 3. Therefore, there were 31 subjects for which there was no significant correlation between expenditures and circulation. These findings mirror those of Crotts (1999) in that he found fewer than 30 percent of subjects in his study where there was a correlation between circulation and expenditure and there were both positive and negative correlations across subjects (Crotts, 1999, p. 263). These results show that a simple correlation does not show the entire story and that a more refined, subject-specific analysis is necessary.

Table 3. Correlation Values Calculated for All and Individual Subject Areas. Data in bold are statistically significant at the .05 level.



Subject	p value	Subject	p value
all subjects	0.836	WI	-0.129
LC items	-0.397	WJ	-0.15
QS	-0.397	WK	-0.084
QT	-0.2	WL	-0.196
QU	-0.095	WM	0.098
QV	-0.174	WN	0.043
QW	-0.271	WO	-0.087
QX	-1	WP	-0.055
QY	0.197	WQ	-0.173
QZ	0.038	WR	-0.583
W	0.537	WS	-0.104
WA	-0.048	WT	-0.141
WB	-0.146	WU	0.866
WC	-0.193	WV	-0.031
WD	-0.218	WW	0.224
WE	0.074	WX	-0.091
WF	0.196	WY	-0.125
WG	-0.033		

The first part of the analysis was calculation of the ACU for each subject area.

This involved computing the average cost of items in each subject by dividing the total

expenditure in each subject by the number of items purchased in the subject. This calculation revealed a substantial variation in the cost of usage for different subjects. The lower the value of the ACU, the lower the cost of use for materials in a subject and indicates levels of circulation that reduce the effective cost to the library for monographs in a specific subject area. The results are shown in Table 4. The standard deviation for the ACU statistic (30.44) for the entire dataset was quite large relative to the mean (30.44). In particular, the ACU values for WZ – History of Medicine (12.84), WU – Dentistry/Oral Surgery (42.00), QX – Parasitology (60.25), and WV – Otolaryngology (191.31) are of concern due to the small number of items purchased in each subject (2 QX, 3 WU, 7, WZ, 8 WV). Therefore, these subjects were excluded from the final analysis.

The resulting mean (34.33) and standard deviation (15.43) from excluding these subjects further reinforced exclusion of those data points. Of the remaining 32 subjects, 17 (53.13%) returned ACU values below the mean with the lowest value returned by QS – Human Anatomy (8.04) and the highest value returned by WD – Disorders of Systemic, Metabolic, or Environmental Origin (76.72).

The second part of the analysis involved the calculation of ECU for each subject. This involved first computing the expected cost of items in each subject by multiplying the average cost of an item in the entire sample (95.24) and the number of items in the subject. This number was then divided by the circulations of items in the subject. The results of the calculations are shown in Table 5. As with the ACU statistic, the values returned for QX (47.62), WU (31.75), WV (152.38) and WZ (41.67) were omitted from the final analysis. The resulting mean (32.38) and standard deviation (11.33) from excluding these values again reinforced exclusion of these subjects. Of the remaining 32

subjects, 17 (53.12%) returned ECU values below the mean with the lowest value returned by the LC books (10.08) and the highest value returned by WD – Disorders of Systemic, Metabolic, or Environmental Origin (4.09).

Table 4. Calculation of ACU statistic w/ mean and standard deviation

Mean (ACU): 39.03
St. Dev. (ACU): 30.44

**Excluding QX, WU, WV,
& WZ**
Mean (ACU): 34.33
St. Dev. (ACU): 15.43

Subject	Total expen. in dollars	Total circs. of items	ACU \$/circ
QS	2097.2	261	8.04
WZ	205.37	16	12.84
LC books	2487	170	14.63
WA	3222	206	15.64
W	3456.51	211	16.38
QV	2055.36	122	16.85
WB	5259.01	284	18.52
WN	2992.85	159	18.82
WY	5809.52	280	20.75
WX	2949.72	127	23.23
WG	10816.57	435	24.87
WK	842.14	30	28.07
WM	4889.3	165	29.63
WT	1156.89	39	29.66
WC	1604	52	30.85
WS	6733.11	214	31.46
WW	3147.69	96	32.79
WL	10655.9	323	32.99
WR	1274.5	36	35.4
QZ	8329.17	235	35.44
WF	3665.69	103	35.59
WQ	2027.36	49	41.37
WE	14748.77	352	41.9
QW	1719	41	41.93
WU	378	9	42
WO	5379.4	126	42.69
QT	2391.4	56	42.7
QY	1378.38	32	43.07
QU	2088.7	46	45.41
WJ	2822.17	56	50.4
WP	3321.94	64	51.91
WH	2783.62	52	53.53
QX	241	4	60.25
WI	4576.85	68	67.31
WD	1534.42	20	76.72
WV	956.56	5	191.31

Table 5. Calculation of ECU statistic w/ mean and standard deviation and sample means

Mean (ECU): 36.38
St. Dev. (ECU): 22.75

**Excluding QX, WU, WV,
& WZ**

Mean (ECU): 32.38
St. Dev. (ECU): 11.33

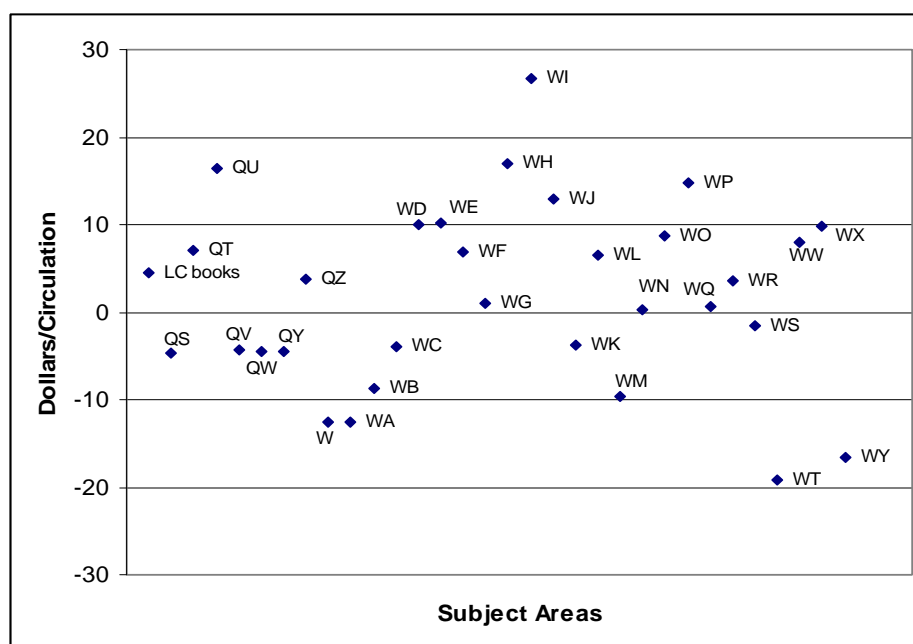
\$129997.60/1376 items
= \$95.24 mean cost/item

Subject	Exp. expen. in dollars	Total circs. of items	ECU \$/circ
LC books	1714.25	170	10.08
QS	3333.27	261	12.77
WX	1714.25	127	13.50
WN	2952.33	159	18.57
QV	2571.38	122	21.08
WG	10380.76	435	23.86
WW	2380.91	96	24.80
WL	8571.27	323	26.54
WB	7714.14	284	27.16
WA	5809.41	206	28.20
WF	2952.33	103	28.66
W	6095.12	211	28.89
QU	1333.31	46	28.99
QZ	7428.43	235	31.61
WE	11142.65	352	31.66
WK	952.36	30	31.75
WR	1142.84	36	31.75
WU	285.71	9	31.75
WS	7047.49	214	32.93
WO	4285.63	126	34.01
WC	1809.49	52	34.80
QT	1999.96	56	35.71
WH	1904.73	52	36.63
WP	2380.91	64	37.20
WY	10475.99	280	37.41
WJ	2095.2	56	37.41
WM	6476.07	165	39.25
WI	2761.85	68	40.62
WQ	1999.96	49	40.82
WZ	666.65	16	41.67
QW	1904.73	41	46.46
QX	190.47	4	47.62
QY	1523.78	32	47.62
WT	1904.73	39	48.84
WD	1333.31	20	66.67
WV	761.89	5	152.38

The final step in the data analysis involved a comparison of the results of the ACU calculations with those of the ECU calculations. The ACU values were subtracted from the ECU values and the results are listed in Table 6. When combined with the

analysis of the ACU values relative to the mean ACU value, these results produced 4 sets of subject areas for discussion. See Table 7.

Table 6. Difference between ACU and ECU scores. Negative values indicate lower average cost per monograph in a subject relative to mean cost of the entire dataset.



Subject	ACU - ECU	Subject	ACU - ECU
WZ	-28.83	QZ	3.83
WT	-19.18	LC books	4.55
WY	-16.67	WL	6.45
WA	-12.56	WF	6.93
W	-12.51	QT	6.99
WM	-9.62	WW	7.99
WB	-8.64	WO	8.68
QS	-4.74	WX	9.73
QY	-4.54	WD	10.06
QW	-4.53	WE	10.24
QV	-4.23	WU	10.25
WC	-3.95	QX	12.63
WK	-3.67	WJ	12.98
WS	-1.47	WP	14.70
WN	0.25	QU	16.42
WQ	0.56	WH	16.90
WG	1.00	WI	26.69
WR	3.66	WV	38.93

The first set consisted of subjects where the ACU value was less than the mean value (34.33) and the value of ACU – ECU was negative. Subjects in this set were characterized as providing the most value for allocated dollars as 1) the cost of use was lower than the average and 2) costs of use were lower than the estimate predicted by the population level statistic, indicating that the materials in these subjects are, on average, less expensive than the average book purchased for the collection. Subjects that fell in this category are listed in the bottom left side of Table 7.

The second set consisted of subjects where the ACU value was less than the mean value (34.33) and the value of ACU – ECU was positive. Subjects in this set were characterized as also providing value for allocated dollars as 1) the cost of use was lower than average in light of 2) costs of use were higher than the estimate predicted by the population level statistic, indicating that the materials in these subjects are, on average, more expensive than the average book purchased for the collection. Subjects that fell in this category are listed in the top left side of Table 7.

Table 7. Breakdown of Subjects Areas by Average Cost and Rates of Use

		Subjects with higher average costs and higher average rates of use		Subjects with higher average costs and lower average rates of use			
		Subject	ACU – ECU \$/use	ACU \$/use	Subject	ACU – ECU \$/use	ACU \$/use
ECU – ACU value positive	LC Books		4.55	14.63	WR	3.66	35.40
	WN		0.25	18.82	QZ	3.83	35.44
	WX		9.73	23.23	WF	6.93	35.59
	WG		1.00	24.87	WQ	0.56	41.37
	WW		7.99	32.79	WE	10.24	41.90
	WL		6.45	32.99	WO	8.68	42.69
					QT	6.99	42.70
					QU	16.42	45.41
					WJ	12.98	50.40
					WP	14.70	51.91
					WH	16.90	53.53
					WI	26.69	67.31
				WD	10.06	76.72	
		Subjects with lower average costs and higher than average rates of use		Subjects with lower average costs and lower than average rates of use			
		Subject	ACU – ECU \$/use	ACU \$/use	Subject	ACU – ECU \$/use	ACU \$/use
ECU – ACU value negative	QS		-4.74	8.04	QW	-4.53	41.93
	WZ		-28.83	12.84	QY	-4.54	43.07
	WA		-12.56	15.64			
	W		-12.51	16.38			
	QV		-4.23	16.85			
	WB		-8.64	18.52			
	WY		-16.67	20.75			
	WK		-3.67	28.07			
	WM		-9.62	29.63			
	WT		-19.18	29.66			
	WC		-3.95	30.85			
	WS		-1.47	31.46			
ACU value lower than mean				ACU value higher than mean			

The third set consisted of subjects where the ACU value was greater than the mean value (34.33) and the value of ACU – ECU was negative. Subjects in this set were characterized as providing some value for allocated dollars as 1) costs of use were lower than the estimate predicted by the population level statistic, indicating that the materials

in these subjects are, on average, less expensive than the average book purchased for the collection even though 2) the cost of use was higher than average. The subjects that fell in this category are listed in the bottom right side of Table 7.

The fourth set consisted of the remaining 9 subjects (excluding QX, WU, and WV) that returned an ACU value greater than the mean value (34.33) and the value of ACU – ECU was positive. Subjects in this set were characterized as the most expensive allocation subjects as 1) the cost of use was higher than average and 2) costs of use were higher than the estimate predicted by the population level statistic, indicating that the materials in these subjects are, on average, more expensive than the average book purchased for the collection. The subjects that fell in this category are listed in the top right side of Table 7.

DISCUSSION

Past criteria for collection development at the Duke University Medical Center Library has included input from the following resources: collection reviews involving input from library users, reviews of authorized lists of core titles in specific disciplines such as Doody's and Brandon Hill, statistics of online content use, and journal impact factors to evaluate collection development activities. At present, this library is exploring the use of acquisitions and circulation data gathered from the integrated library system to feed into an evaluation of the monograph collection development process.

The results of the analysis indicate that with regards to some subjects, this library has done well in allocating budgetary resources from the standpoint of cost per circulation. For 17 subject areas, the library has allocated funding in the last 3 fiscal years such that the cost per circulation is below the average cost of use for all subjects. Of these

17 subject areas, 12 indicated lower per book cost than the average expenditure for monographs purchased during this time period. Further, 2 of the 14 subject areas with above average cost per circulation are subjects that indicated lower per book cost than the average expenditure for monographs purchased.

The subject group that should be further examined is the list of 13 subject areas that indicated higher per book cost than the average monograph purchase and reflected a higher cost per use than the average. The subjects in this group accounted for 34.57 percent of expenditures for the three year period reviewed yet contributed only 27.79 percent of loans. See Table 8 for calculations. Of particular interest, 4 of the top 10 subjects in terms of percent of budget allocated are included in this category. However, the total allocation in these subject areas equals only 34.57 percent of expenditures. See Table 1 for percentages. In light of these finding, the subject areas in this group should undergo further review to identify whether collection development in these areas should be revised or shifted to other parts of the collection. The lack of usage in these subjects may be due to either lack of subject interest in the institution or materials purchased in these subjects may not be appropriate for the user base. Further, an implication for the high amount of budget allocation in QZ and WE is that these subjects constitute core subject areas that the library supports. Therefore, reviewing the allocations in those particular subjects is suggested to more adequately support ongoing research and scholarship at the institution.

Table 8. Details of Allocations in Subjects Meriting Further Consideration

Subject	# of Items	Expenditure in Dollars	# of Loans
QT – Physiology	21	2391.40	56
QU – Biochemistry	14	2088.70	46
QZ – Pathology	78	8329.17	235
WE – Musculoskeletal System	117	14748.77	352
WF – Respiratory System	31	3665.69	103
WD – Disorders of Systemic, Metabolic, or Environmental Origin	14	1534.42	20
WH – Hemic & Lymphatic Systems	20	2783.62	52
WI – Digestive System	29	4576.85	68
WJ – Urogenital System	22	2822.17	56
WO - Surgery	45	5379.40	126
WP – Gynecology	25	3321.94	64
WQ – Obstetrics	21	2027.36	49
WR – Dermatology	12	1274.50	36
Totals	449	54943.99	1263
Totals from Sample	1365	129997.60	4544
Percent of Total Acquisitions	32.89% of items	34.57% of expend.	27.79% of loans

If the library is considering re-allocation of funding going forward, subject areas demonstrating strong usage in relation to cost should be considered. The subject areas with ACU Values below the mean (34.33) are ranked in order of ACU in Table 9. Of particular note, 6 of the top 10 subject areas, in terms of total expenditures, are included in this list. These subjects are WG, WS, WY, WB, WM and W. In addition, the total budget allocation for these subject areas is 53.982 percent. These two points indicate that the library is maintaining a strong emphasis on subject areas that maintain strong user demand for those materials. Further, if these subject areas are core disciplines of the departments supported by the library, an argument can be made that the library is doing a good job of acquiring materials that are in demand by library users.

Table 9. Allocations in Subjects with ACU Values Below the Mean (34.33)

Subject	ACU	# of Items	Expenditure in Dollars	# of Loans
QS – Human anatomy	8.04	35	2097.2	261
LC Books	14.63	18	2487.00	170
WA – Public Health	15.64	61	3222	206
W – Health Professions	16.38	64	3456.51	211
QV - Pharmacology	16.85	27	2055.36	122
WB – Practice of Medicine	18.52	81	5259.01	284
WN – Radiology/Diagnostic Imaging	18.82	31	2992.85	159
WY – Nursing	20.75	110	5809.52	280
WX – Hospitals & Other Health Facilities	23.23	18	2949.72	127
WG – Cardiovascular System	24.87	109	10816.57	435
WK – Endocrine System	28.07	10	842.14	30
WM – Psychiatry	29.63	68	4889.3	165
WT – Geriatrics/Chronic Disease	29.66	20	1156.89	39
WC – Communicable Diseases	30.85	19	1604	52
WS – Pediatrics	31.46	74	6733.11	214
WW - Ophthalmology	32.79	25	3147.69	96
WL – Nervous System	32.99	90	10655.9	323
Totals		860	70174.77	3174
Totals from Sample		1365	129997.60	4544
Percent of Total Acquisitions		63.000% of items	53.982% of expend.	69.850% of loans

The results of the analysis of acquisitions and circulation data present an active use of library materials acquired between July 2004 and June 2007 by this particular library. This is particularly evident by the subject areas included in Table 9. In view of these findings, it is recommended that at the very least, allocations for materials in these subjects be sustained. Unlike Britten (1990), who states, “those areas that are deviating from the average in a positive way should be ‘rewarded’ with enlargement,” (Ochola, 2002, p. 11) considerations for inflation and budget unpredictability are cause for guarded optimism in light of these findings. Further, unlike Crotts, who suggests, “decreasing funds allocated to books with low ratios and shifting them upwards to subjects with high circulation in relation to expenditure” (Crotts, 1999, p. 267), the recommendation in this

situation is to further review allocations in those subjects with high costs of use.

Additional information would be required to determine if less allocation in these subjects is merited or if future funds should be allocated with more information regarding research and clinical needs.

STUDY LIMITATIONS

In light of the information obtained from the acquisitions and circulation data, there are a number of limitations to this study that should be discussed. Several limitations are related to the use of circulation data to determine levels of use of library materials. Lancaster (1982) enumerated on an ideological discussion regarding the limitations of using circulation data. He described studies that use circulation data as focusing on the demands of users rather than the needs of users. As a result, he argued, “they tend to focus only on the expressed needs of those people who are currently active users of a library” (Lancaster, 1982, p. 39). In this situation, Lancaster made a valid argument in that there is no way to quantify what the proportion of user needs for materials exist outside of the visible and recorded transactions conducted by the library. This particular study only considers circulations of materials, i.e. the system logs the loan of an item to a patron record. The integrated library system also allows for the capture and review of events such as in-house use statistics (that are collected twice a year in two week blocks) as well as hold requests placed on items. These datasets may provide a more complete picture of material use in the analysis, however, even these logged events do not capture all intentions to use library materials on the part of library users.

Kraemer also discussed this issue and stated, “the unique nature of monographic purchases poses substantial challenges to the proposition of basing future monographic purchases on the usage of monographs purchased in the recent past” (Kramer, 2001, p. 37). He noted that factors including the changing number of monographs made available in specific disciplines as well as how usage is counted can also impact the collection of usage statistics. Another factor that has received attention but has been left unresolved is the potential for using circulation statistics to predict future use. Day (1995) cites two studies that looked at past circulation history to predict future use, yet no follow-up studies have been conducted to test the hypotheses that past use informs of future use. However, he uses the argument that, “high performing areas will continue to perform well unless there is a major change in teaching patterns” (Day, 1995, p. 157). Despite the limitations that have been discussed, the American Library Association has included book circulation as one measure for Measuring Academic Library Performance in its (MALP) manual (Chen, 1997, p. 74). Therefore, circulation data has a track record of use to assess collection development for academic libraries.

Limitations of this study related to the research design are related to the selected sample, the quality of the data, and a reliance on consistent circulation status of materials. With regards to the sample, this study considered only physical items purchased between July 2004 and June 2007 that circulate in the general collection. Therefore, this sample excludes items such as electronic resources, course reserves and reference materials, materials for staff use, and materials for the leisure reading collection that are purchased using the same fiscal budget. Implications for this study are that the expenditures for excluded items constitute a proportion of the budget that should be considered to properly

evaluate proportions of the budget allocated to specific subjects. However, collecting circulation information for these materials would be an exercise in fruitlessness as only the leisure reading materials are available for circulation. In addition, leisure reading materials are a more heterogeneous collection of items than the LC classified portion of the sample used in this study and as such, constitute a unique collection of monographs for which no catalog classification is provided.

The most prominent limitation in this study is the quality of the data that was gathered from the integrated library system. The acquisitions data extracted included detailed information related to each step in the purchase and processing of ordered materials. As a result, the data included between 3 and 7 records for one purchased item. In addition, changes in library staff and workflow resulted in inconsistent acquisitions data entry. To remedy this problem, the cataloging records that were retrieved to provide circulation data were used in the process that vetted the acquisitions information. However, 17 records were excluded from data analysis at the end of this process as appropriate information could not be obtained. In many respects, the challenges of using integrated library systems data to inform collection development decisions in this study reflect the same points that Knutter, Hawkes, Carrigan, Casserly & Ciliberti discussed in relation to the processing, packaging, and analysis of information captured from computer systems.

Reliance on consistent circulation status of materials is another limitation of this study. Library collections inevitably contain materials that change circulation status. Course reserve items and reference materials may be moved into the general collection when an updated edition is acquired. Books may be placed on course reserve or checked

out by one user for periods far exceeding allotted borrowing periods. This study does not account for items that, during the period from July 2004 to June 2007, changed status. Circulation status changes can influence circulation rates of materials. Therefore an analysis of cost per use of materials should give consideration to such changes.

CONCLUSIONS AND FUTURE WORK

This analysis presented in this paper provides a method for generating a snapshot of cost per use information for monographs purchased for an academic health sciences library. The intent is for this data analysis to be used as part of a collection review process to assess collection development activities and move forward with informed decision making. Information that provides not only cost per use information but also information regarding relative cost of materials across subjects has the potential for allowing libraries to adjust budget allocations to support disciplines with high levels of use and audit disciplines with low levels of use. Looking ahead, there are several directions to go from this analysis.

This study brings to forefront the issue of data quality for statistics gathered from integrated library systems. Future studies seeking to utilize data from automated library systems will need to confront this issue and identify a set of guidelines or best practices to ensure that the data used for analysis may be appropriately used for decision making. Any system that relies on human input and interaction will involve a degree of error inherent in the data collected. Therefore, future studies will also be well served by addressing error rates and incorporate measures of data accuracy to provide more reliable data analysis.

Extensions of this particular study may take one of several forms. The first may be to include hold requests, in-house statistic information, as well as interlibrary loan information. Including these types of use statistics will allow for a more complete snapshot of monograph collection use beyond that contained in this analysis. Another extension may be to conduct several follow-up studies to track changes in collection use over a larger period of time. Extending the scope of time may also allow for the inclusion of inflation in the price of books into calculations of cost per use to allow for richer information to develop allocation forecasts for future expenditures. A third extension to this study would be to conduct a follow-up study to assess cost of use in the future following collection development changes implemented in light of recommendations stemming from this analysis.

This analysis may also contribute to future research on data mining of acquisitions and circulation data from integrated library systems. Studies utilizing data mining for collection development decision making have focused on relative use of collection by academic departments and analysis of the subjects used by individuals within those departments to drive budget allocations. This analysis does not focus on circulations at that level of granularity but does focus more attention on the actual expenditures for monographs in various subjects. As interdisciplinary research becomes more prominent in academia, analyses utilizing department and subject utilization will become more valid. Within the context of special libraries, such as academic health science libraries, there is less flexibility in terms of branching out from a core set of disciplines. Therefore, data mining analyses within the context of health science libraries may be better served by circulation statistics to drive budget allocations via algorithms and search agents.

REFERENCES

- Agee, Jim (2005). Collection Evaluation: A Foundation for Collection Development.
Collection Building, 24, 92-95
- Adkins, Stephen (1996). Mining Automated Systems for Collection Management.
Library Administration & Management, 10, 16-19
- Aguilar, William (1986), The Application of Relative Use and Interlibrary Loan Demand
in Collection Development. *Collection Management*, 8, 15-23
- Banerjee, Kyle (1998). Is Data Mining Right for Your Library. *Computers in Libraries*,
18, 28-31
- Baskerville, Richard, & Dulipovici, Alina. (2006). The Theoretical Foundations of
Knowledge Management. *Knowledge Management Research and Practice*, 4, 83-
105
- Blake, Julie & Schleper, Susan P. (2004). From Data to Decisions: Using Surveys and
Statistics to Make Collection Development Decisions. *Library Collections,
Acquisitions & Technical Services*, 28, 460-464

- Britten, William A. (1990). A Use Statistic For Collection Management: The 80/20 Rule Revisited. *Library Acquisitions: Practice & Theory*, 14, 183-189
- Carrigan, Dennis P. (1996). Data-Guided Collection Development: A Promise Unfulfilled. *College & Research Libraries*, 57, 429-437
- Casserly, Mary F. & Ciliberti, Anne C. (1997). Collection Management and Integrated Library Systems. In G. E. Gorman & Ruth H. Miller (Eds.) *Collection Management for the 21st Century : A Handbook for Librarians*, Greenwood Library Management Collection, (pp. 58-80), Westport, CT: Greenwood Press
- Chaudhry, Abdus Sattar (1993). Automation Systems as Tools of Use Studies and Management Information. *IFLA Journal*, 19, 397-409
- Chen, Tser-yieth (1997). A Measurement of the Resource Utilization Efficiency of University Libraries. *International Journal of Production Economics*, 53, 71-80
- Cortes, Edwin. (1983). *Library Automation and Management Information Systems*. *Journal of Library Administration*, 4, 21-33
- Crotts, Joe (1999). Subject Usage and Funding of Library Monographs. *College & Research Libraries*, 60, 261-273

- Cullen, Rowenda. (1992). A Bottom-Up Approach from Down-Under: Management Information in Your Automated Library System. *Journal of Academic Librarianship*, 18, 152-157
- Day, Mike & Revill, Don (1995). Toward the Active Collection: The Use of Circulation Analyses in Collection Development. *Journal of Librarianship and Information Science*, 27, 149-157
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD Process for Extracting Useful Knowledge from Volumes of Data. *Communications of the ACM*, 39, 27-34
- Gleeson, Michael E. & Ottensmann, John R. (1993). Using Data from Computerized Circulation and Cataloging Systems for Management Decision Making in Public Libraries. *Journal of the American Society for Information Science*, 44, 94-100
- Guenther, Kim (2000). Applying Data Mining Principles to Library Data Collection. *Computers in Libraries*, 20, 60-63
- Hawks, Carol Pitts (1988). Management Information Gleaned from Automated Library Systems. *Information Technology & Libraries*, 7, 131-138

Hawks, Carol Pitts (1992). In Support of Collection Assessment: The Role of Automation in the Acquisitions and Serials Departments. *Journal of Library Administration*, 17, 13-30

Jenks, George M. (1976). Circulation and Its Relationship to the Book Collection and Academic Departments. *College & Research Libraries*, 37, 145-152

Kao, S. -C., Chang, H. -C., & Lin, C. -H. (2003). Decision Support for the Academic Library Acquisition Budget Allocation Via Circulation Database Mining. *Information Processing & Management*, 39, 133-147

Knieval, Jennifer E., Wicht, Heather, & Silipigni Connaway, Lynn (2006). Use of Circulation Statistics and Interlibrary Loan Data in Collection Management. *College & Research Libraries*, 67, 35-49

Kraemer, Alfred B. (2001). Evaluating Usage of Monographs: Is It Feasible and Worthwhile? *Collection Management*, 26, 35-46

Lancaster, F. W. (1982). Evaluating Collections by Their Use. *Collection Management*, 4, 15-43

- Littman, Justin & Silipigni Connaway, Lynn (2004). A Circulation Analysis of Print Books and E-Books in an Academic Research Library. *Library Resources & Technical Services*, 48, 256-262
- Malhotra, Yogesh. (1998). Knowledge Management, Knowledge Organizations & Knowledge Workers: A View from the Front Lines [WWW document]. URL: <http://www.brint.com/interview/maeil.htm>
- Morse, Philip P. (1968). *Library Effectiveness: A Systems Approach*. Cambridge, MA: MIT Press
- Nicholson, Scott (2003). The Bibliomining Process: Data Warehousing and Data Mining for Library Decision Making. *Information Technology & Libraries*, 22, 146-151
- Nicholson, Scott & Stanton, Jeffrey (2004). Gaining Strategic Advantage Through Bibliomining: Data Mining for Management Decisions in Corporate, Special, Digital, and Traditional Libraries. In Hamid R. Nemanti & Christopher D. Barko (Eds.) *Organizational Data Mining : Leveraging Enterprises Data Resources for Optimal Performance.*, (pp. 247-262), Hershey, PA: Idea Group Publishing
- Nicholson, Scott (2006). The Basis for Bibliomining: Frameworks for Bringing Together Usage-Based Data Mining and Bibliometrics Through Data Warehousing in Digital Library Services. *Information Processing & Management*, 42, 785-804

- Nutter, S. K. (1987). Online Systems and the Management of Collections: Use and Implications. *Advances in Library Automation Networking*, 1, 125-149
- Ochola, John N. (2002). Use of Circulation Statistics and Interlibrary Loan Data in Collection Development. *Collection Management*, 27, 1-13
- Peters, Thomas (1996). Using Transaction Log Analysis for Library Management Information. *Library Administration & Management*, 10, 20-25
- Reed-Scott, Jutta (1989). Information Technologies and Collection Development. *Collection Building*, 9, 47-51
- Simmons, P. (1970). Improving Collections Through Computer Analysis of Circulation Records in a University Library. *Proceedings of the American Society for Information Science*, 7, 59-63
- Stoller, Michael (2006). A Decade of ARL Collection Development: A Look at the Data. *Collection Building*, 25, 45-51
- Webster, Judy (1993). Allocating Library Acquisitions Budgets in an Era of Declining or Static Funding. *Journal of Library Administration*, 19, 57-74

- Wise, Kenneth & Perushek, D. E. (2000). Goal Programming as a Solution Technique for the Acquisitions Allocation Problem. *Library & Information Science Research*, 22, 165-183
- Wu, C. -H. (2003). Data Mining Applied to Material Acquisition Budget Allocation for Libraries: Design and Development. *Expert Systems with Applications*, 25, 401-411
- Wu, C. -H., Lee, T. -Z., & Kao, S. -C. (2004). Knowledge Discovery Applied to Material Acquisitions for Libraries. *Information Processing & Management*, 40, 709-725