Effect of Serial Characteristics and Library Environment on Serial Collection Decision in an Academic Health Science Library

의학분야 학술잡지 선택에 영향을 미치는 요인 연구

Giyeong Kim
gkim@scils.rutgers.edu
School of Communication, Information, and Library Studies, Rutgers University.
~~4 Huntington St. New Brunswick, NJ 08901, USA~~

Abstract

Since the beginning of discussions on serial collection management, as budgets have waxed and waned over the ensuing decades, a number of key variables affecting selection/deselection have emerged but without the framework of a coherent and accepted theoretical model. This study is an effort to identify variables which affect the serial collection decision with special attention to selection/deselection in the context of an academic health science library. Based on results from correlation analyses and logistic regression analyses, the serial collection decision can be explained and predicted using various combinations of a reduced set of objective variables. Applications of the results to libraries are discussed, and further research is proposed.

초록

학술잡지 구입 예산의 구입비용의 -상승에 따른 압력으로 지난 수십년간 학술잡지의 선택에 영향을 미치는 요인들에 대한 연구가 활발히 진행되어 왔지만, 학술잡지의 선택에 대한 만족할만한 이론적 틀이 제시되지 못하였다. 이에 따라 본 연구에서는 의학도서관에서 의학분야의 학술잡지의 선택에 영향을 미치는 요인들을 확인하여 이러한 이론적 틀을 제시할 수 있는 근거를 마련코자 한다. 본

연구는 상관관계 분석과 로지스틱회귀분석을 통해 학술잡지선택의 분산을 설명하고, 나아가 예측하는 통계적 모델들을 여러 변수조합을 이용해 제시한다. 또한 이러한 모델의 실제적 적용과 향후 연구방향을 논의한다.

Key words: serial collection decision, selection/deselection, hHealth sScience, lLogistic rRegression

학술잡지 구독결정, 의학 도서관, 장서선택 요인

1. Introduction

The evaluation of each academic periodical for collection management in a particular library needs to be done continuously to track commitments and expenditures as serial prices increase beyond the ability of library budgets to expand at the same rate as the serial price spiral. Evaluation criteria have been developed in past research to guide serials' selection and/or deselection, although there is no common agreement on any particular criterion as essential to making this decision. Some criteria identified as important are: use, impact factor, faculty and staffs judgments, price, number of subscriptions, number of times the title is indexed/abstracted, and so on (Nisonger ,–1998). However, no overall model has been identified for the evaluation of serials which includes both print and electronic serials. Also, models developed for specific libraries have some difficulties when applied to other collection environments. Reports of serial collection management projects have warned that their results may not apply to other library environments (Lightman & Manilov, 2000; Schmidt & Davis, 1994).

The purpose of this study is to explain integrated print and online serial title decision-making by identifying those variables which determine how essential particular

serials are for inclusion in the library's collection. Doing this offers the promise of a flexible, practical model of serial collection decision which could be applied to different library environments. The focus of this study is to examine serials subscription prioritization through an examination of serials characteristics, library environment, and previous use. Each serial title is prioritized based on its characteristics, the library's environment, and its previous use. The subject area used in this investigation is medical science, and the objective library is an academic health science library. There is some consistency in the overall subject matter but there are still numerous aspects of health sciences to consider this a research study in a slightly controlled environment.

2. Previous Research

Serial collection management, especially selection/deselection, is a process to predict future usefulness of materials by identifying characteristics associated with serials and with the library environment. Equally important are the past uses of the collection. This approach is depicted by matching the characteristics of each periodical with its local library environment (See Figure 1). Shown here are the overlapping and intersecting considerations for periodical collection management when periodical characteristics interact with library environments.

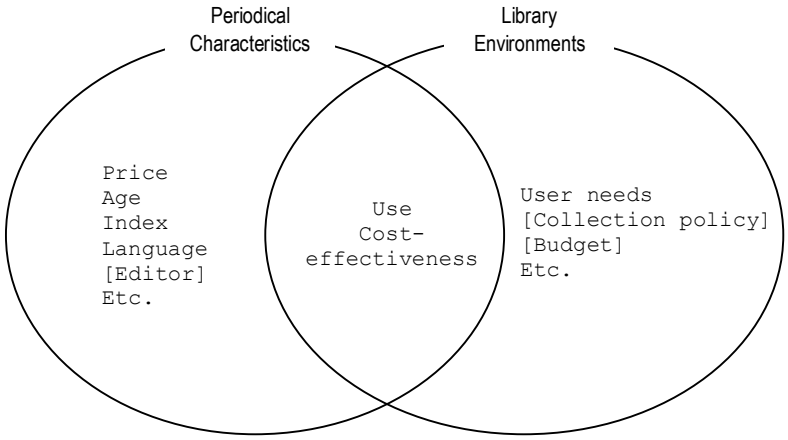


Figure 1. Considerations for Periodical Collection management

Periodical characteristics and the library environment have usually been discussed as criteria, factors, and variables in periodical collection management research. Unfortunately, the characteristics and the environment tend to be discussed arbitrarily. Nisonger (1998) listed 25 characteristics as criteria for periodical micro-evaluation without indicating any structure among them. Evans lists five approaches to collection management and selection: "cost, citation analysis, worth or use, polling [by experts], and core lists" (Evans, 2005, p. 130). Each of these areas have been defined by different variables in different studies, leading to the lack of have a formative base from which to create an evidence-based decision.

Central to much of this research as been a concern with identifying and measuring the use of a serial title.Use can be defined as an output of past periodical collection management and it has been considered the most important predictor of future use. It has been measured in various ways, such as the *sweep method*, *patron record method*, and *slip method* (Farrington, 1997; Nisonger, 1998) for in-house use, and *document delivery*

request analysis for use of external materials. Fussler and Simon (1969) identify numerous problems with defining browsing and determining what actually constitutes the use of a printed source. Also, they note a number of weaknesses in any method selected, such as how time-consuming and laborious it would be to collect data, how user-dependent a specific data point might be, and, how to assess the absence of items in a collection being evaluated. It is recognized that each method has its own strengths and weaknesses.

It has been claimed that *citation analysis* is a reliable predictor of the potential use of a periodical (Pan , 1978) based on the assumption that "the cited work has been somehow used by the author(s) of the citing work" (Nisonger, 1998). There are various types of citation studies: use of dissertations produced in a university to determine which sources were used in the conduct of original research (Buchanan & Herubel, 1994 and Herubel, 1991); studies looking at aggregate data use citation indexes, such as Science Citation Index (Hughes, 1995; Lascar & Mendelsohn, 2001; Lightman & Manilov, 2000; and Loughner, 1996); citations studies based on students' use of source material when completing term assignments (Magrill & St. Clair, 1990); students' research papers references (Sylvia & Lesher, 1995 and Joswick, 1994); and, studies of textbooks (Currie, 1989) as sources. Some weaknesses in using citation data are: the lack of a citation index which includes all publications and the inaccuracy of the citation information such as incomplete citation and typographical errors (St.Clair & Magrill, 1990; Pandit, 1993; and Sweetland,1989).

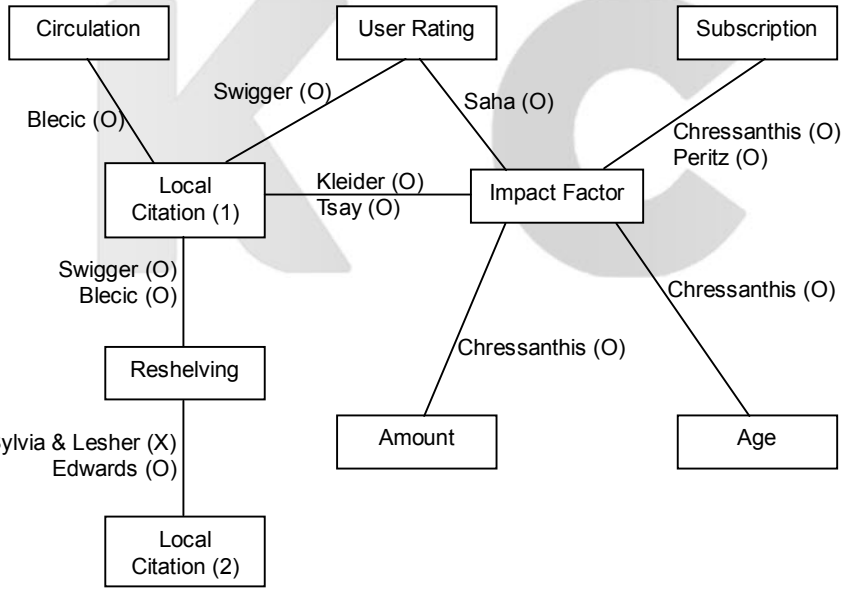
Cost is another important element which has a direct bearing on the development and continuation of the periodical collection. Usually, cost has been estimated from the subscription price (Chrzastowski, 1991; Klassen, 2001; Milne & Tiffany, 1991).

Authority judgment by subject specialists, especially as it relates to a discipline's core journal list has been considered as an important criterion in journal collection management. Usually, the lists are developed within a specific subject area by a group of librarians and subject specialists, such as Magazines for Libraries (LaGuardia, 2003) and the Brandon & Hill list (1997). Different professional associations can list core journals within their discipline and library holdings may reflect accreditation recommendations.

There are other considerations pertinent to the development of collection evaluation models focusing on periodicals. They include journal history (Chressanthis & Chressanthis, 1993), number of subscriptions (Chressanthis & Chressanthis, 1993; Peritz, 1995; De Marchi & Rocchi, 2001), peer review, and subject area (Virgo, 1977). Specific information on how these link to collection management is presently lacking, but these variables do share an important common link with others included in the model here: they are measurable. Being measurable, these variables have been investigated regarding their relationship to journal importance and quality and it appears that these could then be linked to collection issues. Note that not all studies using measurable variables have been cast within the framework of an overall theory.

In previous studies, the relationships among variables have been an interesting sub-area for both collection management and research productivity issues. For collection management, the interest is mainly on the relationship among variables as they are used to define use. Some researchers have done this with data from a citation index (Swigger, K. & Adeline, W. 1991; Blecic, D. 1999), others with data from primary materials--such as journal articles produced by a university's faculty and the dissertations and theses written by its students (Sylvia, M. & Lesher, M. 1995; Edwards, S. 1999).

For research productivity, the focus is often on how periodical characteristics affect periodical quality. Researchers have usually set "Impact factor" as an indicator of quality. Impact factor is defined as the normalized number of citations per article given to a periodical during a specific time period (Institute for Scientific Information, 2000). Impact Factor is then linked to other periodical characteristics, such as periodical age, periodical density (number of pages per volume), and number of subscribers (Chressanthis & Chressanthis 1993; Peritz 1995). Some researchers have been interested in whether Impact Factor is used as a variable for collection management, and in its relationship to local citation data (Kreider 1999; Tsay 1998). These relationships are



Notes. 1. Signs with lines are authors and results (O = significant, X = insignificant)
2. Local citation (1): Data from citation index
3. Local citation (2): Data from texts, such as dissertations.
4. Amount: The number of pages in a volume
5. Age: The age of a serial title
6. Subscription: The number of subscriptions

Figure 2. Relationships among the variables related to the serials management based on the previous research

illustrated in <Figure 2.>

From these relationships, we might be able to infer that some variables can be estimated from other variables based on the strength of their mutual bivariate relationships. Of concern when constructing such models is the overlap or redundancy among the variables of interest.

3. Methodology

The unit of analysis is a serial title. Medical Science is set as the subject area under investigation because it is one of the larger areas covered in Science Citation Index (SCI). Medical science is also selected since it represents a focused collection which is a more stable environment. The population is the serial titles assigned as titles in Medical Science and its related area in Ulrich's Periodical Directory (UPD), which is well known as one of the exhaustive databases for periodicals. The data have been gathered from bibliographic databases, such as UPD, Journal Citation Report (JCR), and SCI, and from a medical school library's online catalog.

The independent variables (IV) used in the study include the following: impact factor (IF), local citation (LCT), local publication (LPB), subscription price (PRC), frequency (FQ), number of articles per year (AY), number of indexes (IDX), number of subscriptions (SUB), publication history (PH), the local subscription history (LSH), relative history of local subscription (LHR), English (ENG). All IVs are ratio, except for ENG, which is binary. The dependent variable (DV), which represents the serial collection decision, is holding status (HS), which is also binary. The variables' definition is shown in <Table 1.>

Table 1. Variables and their characteristics

Variable	Description	Source	Charact eristics	Variable Definition	Cate gory
IF	Impact factor	JCR	Ratio	a measure of the frequency with which the average article in a journal has been cited in a particular year	IV
LCT	Local citation	SCI	Ratio	number of citations linked back to people in an institution which can include a local library	IV
LPB	Local Publication	SCI	Ratio	number of publications linked back to people in an institution which can include a local library	IV
PRC	Subscription price	UPD	Ratio	price of institutional subscription for a year in US Dollar	IV
FQ	Frequency	UPD	Ratio	number of publications of a serial title in a year	IV
AY	Number of articles per year	JCR	Ratio	number of articles in a year	IV
IDX	Number of indexes	UPD	Ratio	number of indexing and abstracting services, which include particular serial titles	IV
SUB	Number of subscription	UPD	Ratio	number of paid subscriptions of a periodical title	IV
PH	Publication history	UPD	Ratio	number of years that a serial title has been published	IV
LSH	Local subscription history	Library	Ratio	number of years that a serial title has been subscribed in a library	IV
LHR	Relative local history of subscription	Library	Ratio	ratio between the publication history and the local subscription history	IV
ENG	English	UPD	Nominal	language of content. The values are 1 (English) and 0 (non-English).	IV
HS	Holding Status	Library	Nominal	current status of a serial title in the library. The values are 1 (held) and 0 (not-held)	DV

cf. SCI: Science Citation Index
JCR: Journal Citation Report
UPD: Ulrich's Periodical Directory

The sample is randomly selected from the population of source periodicals which exceed over 20,000 titles. The number of selected sample is 188 titles. The selected sample titles are searched in a library's online catalog. The library system in University of Medicine and Dentistry in New Jersey (UMDNJ) is selected as a subject library. It is one

of the largest academic health science libraries for a free standing medical, dental, and health institution.

All variables are easily obtained online from the sources. However, even if they are obtained electronically, the transformation of the source data into a usable form for evaluation purposes have not been done easily in previous studies, especially in local citation studies (Klassen, 2001; Lightman & Manilov, 2000; Loughner, 1996), in which data are usually transformed from the source using word processors and/or spread sheet programs. The integration of data from different sources for the analyses requires time-consuming and laborious work with the stipulation that each transformation needs to be accurately transcribed. In this study, the transformation is done with a group of small text manipulation programs using Perl, a programming language, and MySQL, a database management system, in order to facilitate data gathering and placement of data in appropriate formats for analysis. This development might be applicable to other periodical evaluation projects and thus, could be useful beyond the current investigation.

After data integration, the effect of identified variables on the serial collection decision is investigated. First, correlation analyses between Independent Variables (IVs) are computed. The purposes of the correlation analyses are: 1) to confirm the relationships between IVs, which have been studied in other investigations (and cited above); and 2) to confirm the possibility if the serial collection decision is explained by various combinations of a smaller number of variables. Based on the results from the correlation analyses, the effects of the variables on the serial collection decision are investigated using logistic regression (LR). The reasons for using LR are: 1) DV is categorical and IVs are both continuous and categorical; and 2) LR has fewer

assumptions than methods such as discriminant analysis. A usual assumption such as normal distributions of variables is not realistic here and collected data can be expected to be non-normal, thus reaffirming the need to use LR. In this study, non-normally distributed variables are not transformed in order to obtain normal distribution, because the scores in the data are real and objective, and because data transformations provide, in any way, later difficulties when interpreting results. LR is conducted several times to measure for the effects of all IVs as well as some combinations of IVs.

4. Data Manipulation

Some variables are manipulated or transformed to correspond to the definition for a particular variable. The price in United State Dollars for an institutional subscription is used in this study. When a serial title has a multiple subscription prices policy, the price for library or institutional subscriptions in United States is selected. The prices in currencies other than US Dollar are calculated into US Dollars using the most recent currency rate at the time of data analysis. For the data collected, there are 31 currencies in use and these are given in US dollars. Language information is gathered from UPD. When a serial title has multiple languages in its content and its abstracts, the first major content language is selected. In the gathered data, there are 27 languages. This language information is then transformed to binary values: English and non-English, as defined in <Table 1.>.

Table 2 shows the statistical characteristics of the samples. As shown in the table, the distributions of most variables are not normal, except for frequency (FQ) and relative local subscription history (LHR). Three variables contain a large number of missing

values: impact factor (IF), number of articles per year (AY), and number of subscriptions (SUB). Missing values in IF and AY are derived from the different number of titles included in UPD and JCR databases. JCR includes about 7,000 titles, and UPD has about 250,000 titles. The information on both variables is taken from JCR.This means that the titles missed in JCR do not have enough IF to be selected in JCR.

Table 2. Characteristics of original data

Variable	Valid N	Missed N (%)	Mean	SD	Skewness	Kurtosis
IF	82	106 (56%)	1.77	1.74	2.34	6.45
LCT	188	0 (0%)	2.11	7.89	7.88	69.05
PRC	187	1 (.5%)	408.77	791.16	4.66	25.84
FQ	171	17 (9%)	6.67	4.34	.97	1.02
AY	82	106 (56%)	131.32	118.83	2.36	7.18
IDX	181	7 (4%)	13.22	14.17	1.40	2.96
SUB	111	77 (41%)	4588.69	8782.39	5.78	39.75
PH	188	0 (0%)	42.11	27.86	.91	1.34
LSH	188	0 (0%)	12.60	19.64	1.73	2.66
LHR	188	5 (3%)	.30	.40	.83	-1.04
LPB	187	1 (.5%)	.10	.45	7.73	75.49
ENG	188	0 (0%)				
HS	188	0 (0%)				

Note: ENG and HS are categorical, and their descriptive statistics are not calculated

However, the missing values for AY seem to be random. So these variable scores are filled in with estimations using multiple regression analysis with correlated variables with AY. The results from bivariate correlation between the variables shows that frequency (FQ) and local citation (LCT) are highly correlated with AY (Pearson's r > .50), at the same time the effect size between them is not high ($r = .24$, $p < .01$, $r^2 = .057$). Using both variables, the missing values in AY can be predicted using following formula;

$$AY = -25.83 + 16.41\ FQ + 4.19\ LCT\ (r = .76,\ r^2 = .57,\ p < .001)$$

The missing values in SUB are also missing in UPD. These missing values appear to

be random error and not systematic. The results from the bivariate correlation shows that only two variables, AY and LCT, are correlated with SUB at higher than .45 as the *r* coefficient score; all the remaining variables have coefficients less than .20, and most of these are not statistically significant. Using two variables, the missing values in SUB can be estimated as follows;

$$SUB = 1739.38 + 3.20\ AY + 413.75\ LCT \quad (r = .84,\ r^2 = .70,\ p < .001)$$

This approach could be termed recoding to the regressed mean for missing data where values are approximated on the regression line using two variables with stronger correlations. Such a method has the extra validity of linking to a known variable rather than recoding to a constant such as a mean or median score. After transforming the data, the following descriptive measures emerge (as shown in Table 3).

Table 3. Characteristics of data after data manipulation (Continuous variables)

Variables	N	Mean	SD	Skewness	Kurtosis
IF	188	.77	1.45	3.13	12.30
FQ	170	6.71	4.32	.97	1.02
AY	170	94.04	101.33	2.55	9.71
SUB	171	3672.94	7176.81	7.19	61.58
IDX	181	13.22	14.16	1.40	2.96
PRC	188	406.60	789.60	4.67	25.96
LPB	187	.10	.45	7.73	75.49
LCT	188	2.11	7.89	7.88	69.05
PH	188	42.11	27.86	.91	1.34
LSH	188	12.60	19.64	1.73	2.66
LHR	187	.30	.40	.83	-1.05

5. Results

5.1 Correlation Analyses between IVs

First, Pearson product moment correlations are computed to show the relationships among all pairs of variables and this is provided in Table 4. These results show that most of the variables are correlated with each other. This is most probably derived from the fact that most variables can be viewed as functionally related to 'use'. The only indicator of use among the gathered variables is local citation. However, impact factor is a ratio between the number of articles and the number of citations in a journal for a one year average and this, in fact, is another form of use. For frequency and the number of articles per year, it can be inferred that the amount of publication variable increases as use increases. As serial titles use increases, then there is likely to be concomitant increase in the number of indices listing those titles. However, some variables do not appear to be related to the function of use and these are frequency, subscription price, publication history, and local publication.

Table 4. Multiple correlation with all continuous variables

	IF	FQ	AY	SUB	IDX	PRC	LPB	LCT	PH	LSH
FQ	.380**									
AY	.396**	.781**								
SUB	.163*	.200**	.291**							
IDX	.699**	.530**	.564**	.207**						
PRC	.414**	.403**	.233**	-.046	.455**					
LPB	.192**	.150	.427**	.140	.246**	.082				
LCT	.463**	.237**	.507**	.467**	.581**	.105	.451**			
PH	-.043	.148	.282**	.207**	.189*	-.113	-.013	.300**		
LSH	.279**	.196*	.322**	.123	.515**	.217**	.106	.427**	.370**	
LHR	.336**	.180*	.184*	-.025	.494**	.308**	.129	.225**	-.001	.806**

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

These results indicate a large number of significantly correlated pairs of variables and this does reaffirm previous research, in which variables in serial collection management are related to each other as described in the previous sections of this paper. Also, they

show the possibility that the serial collection decision can be explained using various combinations of a smaller number of variables due to the expected overlap in variance among the variables.

5.2 Logistic Regression

A binary logistic regression analysis was done to examine the effects of the variables on the subscription decision (holding status, HS) by the library.

The results are;

$$\chi^2_{\text{Model}}\,(12) = 180.15,\ p < .001$$

$$-2LL = 23.193,\ \textit{Nagelkerke}\ R^2 = .94$$

Surprisingly, 160 out of 162 cases are correctly predicted by the results (98.8%). The results show that 12 variables are good predictors of the inclusion of each serial title in the library. However, not all variables contribute to the subscription decision evenly based on the correlations among the variables, and it is possible that another logistic regression with fewer variables predicts the membership of each serial title into the groups of subscribed titles and non-subscribed titles based on the results from the correlation analyses. To investigate what combinations of variables can predict the membership as well as the combination of all variables, R^2 of all possible combinations are examined using StatGraphics. Table 5 shows R^2 for each variable and Table 6 shows R^2 for each combination of variables which have higher scores than other combinations. Note that R^2 in the table is not the same as *Nagelkerke* R^2.

From Table 5, we can identify the most important variable related to a library's

decision to subscribe to a serial publication: the relative history of a subscription (LHR), which is a ratio between publication history and local subscription history. The number of indexes (IDX) is also an important variable in terms of local subscription. Impact factor, language, and subscription price are additionally significant variables related to the decisions. However, since there are inter-correlational relationships among the variables, all variables do not contribute to explaining the local subscription decision independently. For instance, even though the two highest variables are LHR and LSH, one of them should be ignored in terms of its contribution to explaining local subscription, because the two measures are highly inter-correlated ($r = .81$, $p < .001$), thus, indicating a problem of multicollinearity.

Table 5. R^2 and Adjusted R^2 for each variable

Variable	R^2	Adjusted R^2
LHR	73.66	73.49
LSH	56.91	56.65
IDX	42.27	41.91
IF	22.84	22.36
ENG	17.08	16.56
PRC	12.04	11.49
FQ	11.69	11.14
AY	11.68	11.13
LCT	11.36	10.81
LPB	5.04	4.45
PH	1.28	.66
SUB	.62	.00

Table 6 shows the combination of variables in their contribution to increasing effect size, that is, the increase in R^2 on explaining local subscription. The first content row shows R^2 for the combination of all variables, which is not far from other variables. As expected, highly correlated variables are not included in one combination simultaneously, such as the pair of LHR and LSH. The combination for the highest R^2 is

FQ*IDX*LHR*LPB, which yields an effect size or R^2 = 79.99%, which is quite close to the combination achieved using all variables. The differences in R^2 among the combinations does not appear to be meaningful. This means there are multiple alternatives for the model of membership prediction of serial titles in the library. In the 12 combinations, IDX and LHR are used in all combinations, FQ in 8 combinations, LPB in 5, and ENG, PH, SUB, and AY are in two combinations, respectively. The results of logistic regression analyses with selected combinations are shown in Table 7.

Table 6. R^2 and AdjustedR^2 for combinations of variables

R^2	Adjusted R^2	Combination of variables
81.11	79.45	All 12 Variables
80.49	79.99	FQ*IDX*LHR*LPB
80.58	79.96	ENG*FQ*IDX*LHR*LPB
80.57	79.94	FQ*IDX*LHR*LPB*PH
80.54	79.91	AY*FQ*IDX*LHR*LPB
80.21	79.83	FQ*IDX*LHR
80.32	79.81	ENG*FQ*IDX*LHR
80.27	79.77	FQ*IDX*LHR*PH
80.26	79.76	FQ*IDX*LHR*SUB
79.96	79.58	AY*IDX*LHR
79.96	79.58	IDX*LHR*LPB
79.66	79.40	IDX*LHR
79.75	79.36	IDX*LHR*SUB

Table 7. Results from a series of logistic regression with various combinations of IVs and with local subscription as the DV

Combination of IVs	Model			-2LL	Nagelkerke R^2	% of Correctly predicted cases
	d.f	χ^2	*p*			
FQ*IDX*LHR*LPB	4	171.43	> .001	31.92	.91	96.3
FQ*IDX*LHR	3	171.76	> .001	32.36	.91	96.3
AY*IDX*LHR	3	171.23	> .001	32.89	.91	96.9
IDX*LHR*LPB	3	177.03	> .001	42.17	.89	95.0
IDX*LHR	2	177.07	> .001	42.84	.89	95.0

The most important factor in terms of subscription decision is local history, which means how long the title has been previously subscribed to in a local library. The fact that

relative local subscription history, which is the ratio between publication history and local subscription history of a serial title, shows the library tends to maintain subscriptions of subscribed titles, and tend to keep complete volumes of serial titles. In other words, a library's long term commitment to a particular title outweighs other considerations in determining continued ownership.

The number of index services which include the serial title (IDX) is the third influencing variable on local subscription, and it is the next highest contributor to explaining the effect size after the local history factor. Impact factor (IF) is the fourth variable contributing to the local subscription; however it co-varies with IDX. The correlation between IF and IDX is .699 ($p < .001$). This is the reason that IF is not considered in the series of logistic regression analyses. In comparison with IDX, IF has its own drawbacks. IF is gathered from Journal Citation Reports, which includes about 7,000 serial titles, however IDX is gathered from Ulrich's Periodical Directory, which includes 250,000 serial titles. This means IF is more likely to be missed than IDX for each serial title, and IDX seems to be a stronger variable than IF based on comprehensiveness of its coverage and availability. .

<Table 8> shows the coefficients when a combination is selected to predict the membership of serial titles into the subscription group. For example, in the combination of FQ*IDX*LHR, the following formula is used;

$$P=\frac{1}{1+e^{-Z}}$$

Where,

$Z = -8.58 + .13\,(FQ) + .18\,(IDX) + 9.48\,(LHR)$

Table 8. Coefficients for each combination of variables

Combination of Variables	Coefficients					
	Constant	FQ	IDX	LHR	LPB	AY
FQ*IDX*LHR*LPB	-8.61	.14	.17	9.45	.90	
FQ*IDX*LHR	-8.58	.13	.18	9.48		
AY*IDX*LHR	-8.16		.19	9.28		.003
IDX*LHR*LPB	-7.97		.20	8.48	.95	
IDX*LHR	-7.98		.21	8.53		

This is quite interesting, because local citation, which is the only variable for the concept of *local use*, shows that it does not contribute to the decision on serial subscription in the library, even though there are some variables which represent *use*. But the use which does relate to this is *global use*. Furthermore, local subscription history (LSH and LHR) and local publication (LPB) contribute highly to the decision, as does local environment.

≤Table 9≥ shows the comparison when missing values and incorrectly grouped cases are included among the combinations of variables. The table shows that the incorrectly grouped cases are similar to each other. In cases of IDX*LHR*LPB and IDX*LHR, the numbers of incorrectly grouped cases increase a bit , but these could be derived from the differences due to missing cases. In fact, cases 134, 135, and 136, which are incorrectly grouped by those two combinations, are missing data found in other combinations. Case 77, 84, 110, 111, and 126 are always grouped incorrectly, but case 137 is only grouped incorrectly in some combinations. Even within the combinations, in which the cases are grouped correctly, the residuals of the cases are not small. For example, in the combination of AY*IDX*LHR, the residual for case 137 is .37, which possible range is $-.50 < residual < .50$. This means the results from any one combination can be similar to another combination. A more important consideration appears to be the

number of missing cases. Therefore, when a library gathers a variety of variables in terms of serial collection management, the best combinations of variables for serial subscription may be the one with the least number of missing cases.

Table 9. Comparisons of the combinations of variables with cases

Combination of Variables	Number of Missing Cases			IDs of Incorrectly grouped cases
	subscribed	non-subscribed	total	
FQ*IDX*LHR*LPB	8	18	26	77, 84, 110, 111, 126, 137
FQ*IDX*LHR	8	17	25	77, 84, 110, 111, 126, 137
AY*IDX*LHR	8	17	25	77, 84, 110, 111, 126
IDX*LHR*LPB	6	3	9	77, 84, 110, 111, 126, 134, 135, 136, 137
IDX*LHR	6	2	8	77, 84, 110, 111, 126, 134, 135, 136, 137

How are the incorrectly classified cases interpreted? The group of incorrectly classified cases also consider those cases where residuals are higher than others, such as more than .3, because 'incorrectly classified' means 'high residual', or even approaching .5. There are two ways to interpret such cases: 1) there are other variables needed to classify the serial titles more accurately. In fact, the overall effect size (R^2) of all 12 variables to the local subscription is 81.11%, and there is almost 19% of unexplained variance when considering local subscription. Therefore, we might conclude that there are other variables which explain the remaining 19% of variance. 2) The cases are possible candidates for changing their subscription status in the library. It is possible that the 19% of unexplained variance is due to random error.

In order to show which approach will yield a more accurate model, a follow-up investigation could be conducted with a librarian in a serials management department of a library, and each title could be checked in terms of local subscription. Also, there could

be a concomitant investigation of variables which are moderately correlated with local subscription.

6. Discussion

The statistical analyses, especially correlation analyses, provide us with considerations to understand variable relationships—especially in terms of variable overlap and variable contributions to depict the final decision model. The significant relationships between periodical subscription price and other variables, such as impact factor, publication frequency, number of articles in a year, number of indexing services, local subscription history, and relative local subscription history (refer to Table 4), shows that the price is not the only variable affecting the cost factor variable. Price may affect other factors, such as quality, amount, and local history, because it links to and is partly derived from a publisher's price policy. It might be expected that price determines other aspects of a periodical title but this is an untested assumption and it would need confirmation prior to being employed as a marker for other variables. Also, the number of indexing services is significantly correlated with almost all other variables (refer to Table 4), because when a periodical title has been selected in an indexing service, the periodical title's respective variables may have determined how many different indexing and abstracting services will include it. Therefore, when the expected model is reconstructed or corrected, the relationships among variables and factors from the results of the pilot study should be considered cautiously.

In this study, the budget for the periodical collection's current subscriptions is not considered, because the selection and deselection process is prior to the budget process.

That is, the selection/deselection process can be viewed within a timeline to determine each title's subscription, and the results from this process can be an effective instrument to substantiate the budget allocations to specific serial titles. When there is an expected budget for a periodical, it is easily applied to the multivariate functions presented here. The functions can be expressed as a ratio which shows the degree (probability) of membership of a periodical to a group, so that the periodicals can be ranked by that membership function. Therefore, the range of each group can be adjusted based on the expected budget. It should be noted that extreme budget cost scores will influence selection but that the preponderance of titles will be considered independent of this factor. Secondly, an analysis by budget implies a different unit of analysis for this study which focuses on the title itself with the library as variable. To emphasize budget, the library may need to be the unit of analysis. Nonetheless, that said, the cost of periodicals is included in the beginning, exploratory model. Lastly, budget importance wanes when journals are aggregated since individual costs are often not considered as a separate decision point.

The investigation identifies promising variables in serial collection management which might be used to predict the likelihood of specific serials being included in a library's subscription. The variables should show some level of predictive power when placing periodicals into a context with groups of titles held in a library. The predicting functions for the membership of serials in identifiable groups can be estimated using statistical functions with importance in flexibility in the number of variables. Also, it confirms that the variables proposed in the methodology section are sufficient to explain the variability in the DV, periodical subscription in the library. This methodological

approach plausibly assumes that such analyses can support the subscription decision in a library. Also, this shows the possibility of building a model to explain serial subscription decisions using serial characteristics and local environments. Further, it suggests that the results obtained might apply beyond one library.

References

Blecic, Deborah. (1999) "Measurements of journal use: an analysis of the correlations between three methods." *Bulletin of the Medical Library Association, 87*(1) :-20-25.

Brandon, Alfred N. & Hill, Dorothy R. (1997). "Selected list of books and journals for the small medical library." *Bulletin of the Medical Library Association, 85*(2) : 111-135.

Buchanan, A. L., & Herubel, J. P. V. M. (1994). "Profiling PhD dissertation bibliographies: serials and collection development in political science." *Behavioral and Social Sciences Librarian, 13*(1) :-1-10.

Chressanthis, George A. & Chressanthis, June D. (1993). "The relationship between manuscript submission fees and journal quality." *The Serials Librarian, 24*(1) :-71-86.

Chrzastowski, Tina E. (1991). "Journal collection cost-effectiveness in an academic chemistry library: results of a cost/use survey at the University of Illinois at Urbana-Champaign." *Collection Management, 14*(1/2) :-85-98.

Currie, William W. (1989). "Evaluating the collection of a two-year branch campus by using textbook citations." *Community and Junior College Libraries, 6*(2) :-75-79.

De Marchi, Mario & Rocchi, Maurizio. (2001). "The editorial policies of scientific journals: testing an impact factor model." *Scientometrics, 51*(2) :-395-404.

Edwards, S. (1999). "Citation analysis as a collection development tool: a bibliometric study of polymer science theses and dissertations." *Serials Review, 25*(1) :-11-20.

Evans, G. Edward. (2005). *Developing Library and Information Center Collections.* 5th Ed. Englewood, CO: Libraries Unlimited, Inc.

Farrington, Jean W. (1997). Serials Management in Academic Libraries: A Guide to

서식 있음: 글꼴: 기울임꼴

Issues and Practices. Westport, CT: Greenwood Press.

Fussler, Herman H. & Simon, Julian L. (1969). Patterns in the Books in Large Research Libraries. Chicago, IL: University of Chicago Press.

Herubel Jean Pierre, V. M. (1991). Philosophy dissertation bibliographies and citations in serials evaluation. *Serials Librarian, 20*(2/3), 65-73.

Hughes, Janet. (1995). Use of faculty publication lists and ISI citation data to identify a core list of journals with local importance. *Library Acquisition: Practice & Theory, 19*(4). 403-413.

Institute for Scientific Information. (2000). *JCR Glossary*. Retrieved Jan. 15, 2004, from the World Wide Web: http://jcr4.isiknowledge.com/www/help/hjcrgls2.htm

Joswick, K. E. (1994). Library materials use by college freshmen: a citation analysis of composition papers. *College and Undergraduate Libraries. 1* (1). 43-66.

Klassen, T. W. (2001). Dollars with sense: measuring serials usage using faculty cited journal data. *The Bottom Line: Managing Library Finances, 14*(1), 37-43.

Kreider, Janice. (1999). The correlation of local citation data with citation data from journal citation report. *Library Resources & Technical Services, 43*(2). 67-77.

LaGuadia, Cheryl Ed., (2003) Magazines for Libraries. 12th ed. New Providence, NJ: Bowker.

Lascar, Claudia & Mendelshon, Loren D. (2001) An analysis of journal use by structural biologists with applications for journal collection development decisions. *College & Research Libraries, 62*(5). 422-433.

Lightman, H., & Manilov, S. (2000). A simple method for evaluating a journal collection: a case study of Northwestern University Library's Economics collection. *The Journal of Academic Librarianship, 26*(3), 183-190.

Loughner, W. (1996). Scientific journal usage in a large university library: a local citation analysis. *The Serials Librarian, 29*(3/4), 79-88.

Magrill, R. M., & St. Clair, G. (1990). Undergraduate term paper citation patterns by disciplines and level of course. *Collection Management, 12*(3/4), 25-56.

Milne, D., & Tiffany, B. (1991). A cost-per-use method for evaluating the cost-effectiveness of serials: a detailed discussion of methodology. *Serials Review, 17*(2), 7-20.

Nisonger, Thomas E. (1998). *Management of Serials in Libraries*. Englewood, CO:

서식 있음: 영어(미국)

Libraries Unlimited, Inc.

Pan, Elizabeth. (1978). Journal citation as a predictor of journal usage in libraries. *Collection Management, 2*(1). 29-38.

Pandit, Idrisa. (1993). Citation errors in library literatures: A study of five Library Science journals. *Library & Information Science Research, 15*(2). 185-198.

Peritz, B. C. (1995). On the association between journal circulation and impact factor. *Journal of Information Science, 21*(1), 63-67.

Schmidt, D., & Davis, E. B. (1994). Biology Journal use at an academic library: a comparison of use studies. *Serials Review, 20*(2), 45-62.

St. Clare, Gloriana, & Magrill, Rose Mary. (1990). Incomplete citations in undergraduate term papers from four campuses. RQ, 75-81.

Sweetland, James H. (1989). Errors in bibliographic citations: a continuing problem. *Library Quarterly. 59* (4). 291-304.

Swigger, K., & Adeline, W. (1991). The use of citation data to evaluate serials subscriptions in an academic library. *Serials Review, 17*(2), 41-46.

Sylvia, M. & Lesher, M. (1995). What journals do psychology graduate students need? A citation analysis of thesis references. *College and Research Libraries. 56* (4) 313-318

Tsay, Ming-yueh. (1998) The relationship between journal use in a medical library and citation use. *Bulletin of the Medical Library Association, 86*(1). 31-39.

Virgo, J. A. (1977). A statistical procedure for evaluating the importance of scientific papers. *The Library Quarterly, 47*(4), 415-430.