

A Statistical Analysis of Patrons' In-Library Information Consumption Behaviors of Print Materials

Jiebei Luo

Patrons' information consumption behaviors of print materials are influenced by their content and tangible characteristics. This quantitative study incorporates several content-based factors widely adopted in collection management practices as well as two physical dimension-related factors, including book size and number of pages, and applies a series of regression analyses to evaluate the relative impact of these factors on patrons' checkout versus in-library use decisions. The results suggest that shelf time and book size exhibit consistent and significant impact on patrons' use of print materials. These findings can shed further light on collection management practices and policies in academic libraries striving to provide patrons more user-centered collections and space.

Introduction

Academic libraries are changing their role from "book warehouse" to "information commons" that seeks to facilitate student learning in all aspects.¹ With an overwhelming amount of information available in digital format, librarians and library practitioners have witnessed a continuous change in the usage of library resources and space. Academic libraries now provide services to a more digital-oriented patron base, who searches information online, conducts work and collaborates on electronic devices, and uses the library as a learning space. A growing number of academic libraries have developed awareness toward redesigning their library space with a user-centered concept to meet this evolving trend.² Yet many library practitioners often find their visions hindered by their "long-term assets" in the sense that a great portion of a library's physical space, operational budget, and staff time is still devoted to its print book collections. Echoing the advocacy for providing more user-centered collections and learning space, this study seeks to understand how patrons' information consumption of print materials in the library has correspondingly changed with their increased exposure to the digital content.

We posit that library users may demonstrate different usage patterns when they are interacting with local collections as opposed to electronic resources. This study aims to contribute to the understanding of patrons' information usage behaviors of print books within the library space in the presence of abundant digital content, with a particular focus on in-library usage (in other words, on-site usage without items being checked out) versus checkout usage (that

*Jiebei Luo is Data Services Librarian in the Thomas P. O'Neill, Jr. Library at Boston College; email: luojx@bc.edu.
©2021 by Jiebei Luo.

is, usage outside the library with items being checked out). When assessing and developing policy for local collections in the library, our view is that, rather than employing an aggregate analysis approach that does not differentiate usage data between print materials and other resources, especially those in digital format, collection management should instead seek a more user-centered strategy that targets particular characteristics featured in their print collections. To this end, our study specifically develops a comprehensive analytical model that examines local collection usage from the perspective of academic library patrons.

In the context of this study, the local collections refer to print books in circulation, which do not include other content formats such as print serials, CDs/DVDs, and oversized folios. Our model employs both traditional quantitative criteria adopted in previous collection management studies and physical characteristics of print books that may potentially influence patrons' information usage behaviors. It is hypothesized that, when patrons are in the process of using local collections, their information consumption behaviors are influenced by not only the content-based factors but also the tangible physical features of print materials. These tangible features and their potential impact on patrons' usage patterns have been largely neglected in previous collection usage studies and in turn motivate us to reconsider our existing collection management practices to foster more user-centered library services. From the perspective of daily operations in most academic libraries, local collections present an inherently different inventory stock compared to electronic resources—while local collections are regarded as a more stable possession since they tend to remain unchanged once they reside in the library, access to electronic resources can be more dynamic in the sense that academic libraries are often in constant negotiations with various vendors for different databases, “big deals,” or resource sharing consortia.

Literature Review

Academic libraries have long been interested in applying a wide range of quantitative analyses of collection usage patterns. Findings on this subject help reveal library resource utilization and provide a key criterion for acquisition and weeding practices. Among studies during the predigital era, circulation record is widely adopted as a primary indicator for collection usage. As books can be used without leaving the library space, the in-library usage data proxied by browsing record is another important circulation usage indicator considered in past collection usage assessment research.³ A number of prior studies examine collection usage by using one of the usage indicators (such as circulation) to predict the other (such as in-library use), and this strand of early studies generally agrees that in-library usage is positively correlated with and can be predicated by circulation usage.⁴ Specifically, in a seminal study, Fussier and Simon randomly placed questionnaires in selected books during a 6-month study period and concluded that low book circulation usage can help predict low in-library usage. In a highly cited work based on field evidence from the University of Pittsburgh, Galvin and Kent captured usage data by placing a selected sample of books on librarians' tables during 30 randomly chosen days, and they found that the in-library usage highly correlates with the external usage and that the circulation usage is sufficient to predict whether a monograph will ever be used.⁵

On the other hand, studies relying on random sampling during a limited study period have also been challenged by other researchers who pointed out that the results obtained under such methodology may not be able to represent usage patterns of the entire collection. Later work based on more comprehensive sampling has in fact reached different conclusions. For

instance, Selth, Koller, and Briscoe pointed out that more than 30 percent of the monographs and 25 percent of the serial volumes see only a single type of usage (either circulation or in-library use), which does not allow for usage predictions, implying that deselection may incorrectly remove a large number of recently used items if the weeding criterion focuses only on a single type of usage data as suggested by earlier studies.⁶ Similarly, Borkowski and Murdo as well as Hayes disagreed with the deselection practices derived from earlier studies that weeding and off-storage decisions can be safely made solely based on the corresponding circulation data.⁷

Despite different views among the collection usage studies in the predigital era, Slote provided a comprehensive quantitative analysis framework for evaluating collection usage and offered practical guidelines for deselection.⁸ According to his research, shelf time (in other words, the duration that a print book stays on the shelf since its last checkout) appears to be the most effective factor indicating not only the popularity of the book but also its likelihood of being used in the future. Besides shelf time, factors including book age, subject, duplication, and language are also recommended as effective indicators in predicting the collection usage that also require further customized analysis for more accurate predictions.⁹ Stemming from Slote's work, similar analyses have been conducted across different academic libraries of various scales. For example, Soma and Sjoberg included contributions from multiple library departments and faculties' input as the weeding criteria; Wilde and Allison shared their experience of applying statistical analysis of the collection usage data to make collection development decisions; and Tucker employed usage data of local print serials to examine electronic resource collection development decisions at the University of Nevada, Las Vegas.¹⁰

As information seekers enter the digital era with more scholarly resources becoming available in electronic format, a growing number of libraries have adopted an aggregate approach that jointly assesses usage data of local collections with other electronic resources within the same analysis framework. For instance, Knievel, Wicht, and Connaway used the circulation and interlibrary loan data of the English-language monograph collection at the University of Colorado Boulder and provided strategies for remote storage, preservation, and collection development by comparing the usage data of local collections and interlibrary request data for each subject.¹¹ Knowlton introduced a two-stage collection assessment strategy. By adopting a collection assessment tool called "percentage of expected use" (PEU), his study calculated the PEU within print and electronic books and compared the relative PEU distance within each format to determine if patrons have strong preference for one over another.¹² By analyzing the book titles available in both print and electronic formats, Yuan, Van Ballegoie, and Robertson claimed that the usage of both print and digital versions would be high for popular items in the collection and low for unpopular items, and preferences for print books do exist in some subjects, such as American history.¹³ Also, according to the collection study at Texas A&M University, Kimball, Ives, and Jackson found that top science e-books see significantly higher usage than their print counterparts.¹⁴

A number of studies have further devoted their focus on the usage preference differences across disciplines and among different user groups. In particular, studies conducted by Moore as well as Kimball, Ives, and Jackson pointed out that patrons' usage preference is strongly related to the topic of the book, patrons' academic status, and, most importantly, the purpose of the book.¹⁵ The case study conducted by the Utrecht University Library revealed that researchers in humanities acquire and use paper format materials a lot more than other disciplines, while other case studies showed that most users prefer using academic resources electronically.¹⁶ When comparing usage preference for print books versus their digital coun-

terparts, both faculty and students were found to prefer digital format for searching and scanning information and paper format for perusing a large portion of the book.¹⁷ Ithaka S+R US Faculty Survey in 2015 also revealed that preferences for using print materials over their digital counterparts may still prevail among faculty and researchers.¹⁸ Among different user groups, several studies such as Dewan and Silas Marques found that students prefer reading literature in paper format; in contrast, the study by the Utrecht University Library pointed out that the majority of their BA and MA students rarely use paperback materials on bookshelves due to the tight deadlines that they constantly have for their assignments, while most faculty members would in fact search and use print materials during their library visits.¹⁹

Our study is motivated by and seeks to bridge two major gaps in existing literature. First, prior research has largely neglected the impact of tangible characteristics of print materials from the perspective of patrons' information-seeking behaviors. When patrons are physically interacting with library collections during their information-seeking process, we posit that the tangible characteristics of a book, mainly the physical dimension and the number of pages, can potentially lead to different preferences for in-house or checkout usage. Next, recent assessment studies on collection usage patterns tend to pool usage data of print materials with electronic resources (including ILL items) within the same discipline and seek to predict the usage of one format based on the other.²⁰ A potential issue with such aggregate approach, which also helps explain the large discrepancy in findings among recent related studies, is that print collections and digital collections are inherently different from each other, and the differences in content format between print and digital collections can lead to patrons' different information-seeking behaviors, such as browsing book titles displayed on bookshelves versus skimming through search results from one's laptop. In addition, from collection management's point of view, due to the stable nature of local print materials and the dynamic composition of electronic resources, academic libraries typically devote different departments, staff, and workflow to handle these two types of resources in their daily operations.

Differing from previous collection assessment studies that compare collections in different formats, our study therefore proposes an alternative approach, akin to methodologies adopted in the predigital era, that performs a modular analysis to assess the usage patterns based on a single content format (that is, print books). We aim to understand how our local collections are used by patrons through a user-centered approach by incorporating not only several widely adopted traditional collection analysis factors (such as subject, book age, and shelf time), but also the physical features associated with a print book (for instance, size and number of pages). We hypothesize that these factors may affect patrons' information-consumption behaviors during their information-seeking process within the library space.²¹

By exploring the potential factors in explaining the usage gap of titles in local collections, we summarize the purposes of our study as follows:

1. Examine how patrons' print book usage patterns are affected by the content- and time-related factors (such as subject, book age, shelf time) as well as tangible characteristics (like book size and number of pages);
2. Provide statistical evidence to facilitate routine collection management for subject librarians, and shed light on off-site storage options and space modification possibilities for library administrators;
3. Draw practical implications for a wide range of academic libraries, especially those engaged in active assessment, to make evidence-based decisions on library daily operations.

Study Design, Data, and Methodology

The collection usage data for this study are collected from the Lucy Scribner Library at Skidmore College. Skidmore College is a four-year, private, coeducational liberal arts college located in upstate New York. As the only library on campus, the Lucy Scribner Library provides reference services to students, faculty, and its community members on course-related questions, research projects, and other general inquiries. Currently, the library hosts about 300,000 items locally, including print books, journals, folios, DVDs, CDs, and other items. The library system collects two types of local collection usage data, including circulation volume (historical circulation records by item) and browsing volume (historical records of the items taken away from the shelf and used in the library without being checked out). When examining these two sets of data, we found various usage gaps between in-house use (browsing volume) and circulation volume of the print books. Specifically, when comparing the browsing counts and checkout counts of the same item, the browsing counts are significantly higher than the checkout counts for certain types of books in our collections. In other words, we found that some books are used much more frequently in the library but with very limited checkout usage, and such differences are thought to have reflected patrons' usage preferences when they were interacting with local collections.

More formally, we express the usage gap of a single item (U_{BC}) using the following equation:

$$U_{BC} = U_B - U_C \quad (1)$$

where U_B represents the browsing counts of a title and U_C represents the checkout counts of the same title. Thus, when U_{BC} grows, the usage preference is toward browsing; and when U_{BC} decreases, the usage preference is toward checkout.

The Lucy Scribner Library launched the Voyager ILS in 1998. As of April 30, 2018, 280,535 titles were locally hosted for circulation, of which about 187,000 items were owned prior to 1998. Data used in this study are extracted from the Bibliography and Circulation tables in Voyager—the Bibliography table contains bibliographic records of all items in the library database, while the Circulation table records the checkout and in-house use history. The two tables are mapped through the ITEM_ID field available in both tables. In particular, in-house usage is collected by scanning the books returned to shelf carts available at multiple locations on each floor before the daily reshelving. The library started collecting in-house usage data since the launch of Voyager. As the usage preference (U_{BC}) is calculated by the difference between circulation and in-house usage data under the same title, only books purchased after 1998 ($n = 93,260$) that have a full track record of checkout and browsing history are adopted in this study to ensure accuracy.

As we aim to understand the usage preference from the perspective of patrons, our analysis model includes two tangible indicators of a book, including the number of pages (*Page*) and size (*Dimension*), collected via subfields \$a—Extent (R) and \$c—Dimensions (R), respectively, under field 300—Physical Description (R) of WorldCat MARC record; two continuous time-related factors, including book age ($Year_age = 2018 - \text{Year of Acquisition}$) and shelf time ($Year_shelf = 2018 - \text{Year of Last Checkout}$); and one content-based variable LCC_group that differentiates four major academic disciplines (arts, humanities, sciences, and social sciences) by dividing the titles into the four disciplines according to the Library of Congress Classification. It is also worth noting that the variable *Year_shelf* consists of titles with actual checkout dates ($n = 56,362$) ranging from 0 (that is, last checked out in 2018) to 19 (that is, last checked out in 1999), and the remaining titles without checkout history are imputed with a value of 20 in the final dataset for the regression analysis.

Records with missing values and extremely small/large page and dimension values (that is, smaller than 0.1 percentile and larger than 99.9 percentile) are removed from the merged dataset; a title with multiple copies in the collection is regarded as the same book ($n = 427$, or 0.48% of total valid titles), and usage records of all duplicates are combined; different editions of the same title are regarded as different books. After the data cleaning process, 87,483 valid titles (93.9% of the raw data) are retained for the regression analysis. Table 1 summarizes the details of the variables employed in our regression analysis.

Variable	Description	Calculated by
U_{BC}	the usage difference between browsing and checking out by title	$U_B - U_C$
U_B	historical browse counts (January 1, 1999–April 30, 2018)	Circulation
U_C	historical charge counts (January 1, 1999–April 30, 2018)	Circulation
Page	worldcat physical description 1: page number of the book	Bibliography
Dimension	worldcat physical description 2: length of the book	Bibliography
Year_age	the time the book has been in the collection	2018—year acquired
Year_shelf	the time the book has not been checked out	2018—year last checkout
LCC_group	Library of Congress Classification: M, N, TR, TT	LCC_group = arts
	Library of Congress Classification main classes: B, C, D, E, F, P	LCC_group = humanities
	Library of Congress Classification main classes: A, Z	LCC_group = others
	Library of Congress Classification main classes: GA-GE, HA, Q, R, S, TA-TP, TS, U, V	LCC_group = sciences
	Library of Congress Classification main classes: GF-GV, HB-HX, J, K, L, TX	LCC_group = social sciences

Table 2 presents the descriptive statistics of our main variables of interest. The mean value of historical checkout counts (U_C) is about twice as much as that of historical browsing counts (U_B), suggesting that checkout usage is generally far more frequent than in-library usage in our local collections. The larger standard deviation of historical checkouts ($S.D._{UC} = 3.01$) indicates that the variation in historical checkout counts by title is also larger than historical browsing counts in the library ($S.D._{UB} = 2.19$). The book usage gap (U_{BC}) ranges from -48 to 40 , which indicates that usage preferences (either in-library or checkout usage) do exist in our local collections. With regard to the composition of the collections, books in the humanities account for the largest proportion (46%); books in the arts, sciences, and social sciences account for 19, 10, and 25 percent of the sample, respectively.

To study the factors that can help explain the book usage differences, we use a standard linear regression model to estimate the difference between in-house and checkout usage under the same title. The regression model can be expressed by the following equation:

$$U_{BC} = \beta_0 + \beta_1 \text{Page} + \beta_2 \text{Dimension} + \beta_3 \text{Year_age} + \beta_4 \text{Year_shelf} + \varepsilon. \quad (2)$$

In this model, U_{BC} is the book usage gap as defined earlier. The intercept is represented by β_0 ; β_1 is the set of coefficients that can affect the usage gap difference, including our main variables of interest such as number of pages and book size. The coefficients β_1 and β_2 then help capture

TABLE 2
Descriptive Statistics

Variable	Mean	SD	Min	Max
U_{BC}	0.94	2.94	-48	40
U_B	1.08	2.19	0	48
U_C	2.01	3.01	0	60
Page	299.88	160.57	9	2,471
Dimension	24.22	2.60	40	40
Year_age	11.02	5.48	0	19
Year_shelf	11.18	7.44	0	20
LCC_arts	0.19	0.39	0	1
LCC_humanities	0.46	0.50	0	1
LCC_sciences	0.10	0.30	0	1
LCC_social sciences	0.25	0.43	0	1
LCC_others	0.01	0.09	0	1

how the number of pages and the dimension of a book can affect patrons when they need to decide whether to use the book in the library or check it out. Meanwhile, our model also includes traditional time factors such as book age (β_3) and shelf time (β_4) that librarians widely adopt in the weeding process. We perform the regression analysis separately for each academic discipline, including the arts (*LCC_arts*), humanities (*LCC_humanities*), sciences (*LCC_sciences*), and social sciences (*LCC_social sciences*), to further investigate how our factors of interest may influence the book usage patterns within each discipline.

Baseline Results and Discussions

In this section, we discuss the baseline findings from our regression analysis. We first proceed by estimating our model using the full sample. We then explore the potential discipline-specific heterogeneous impact by extending our analysis to subsamples based on different academic disciplines, including arts, humanities, sciences, and social sciences.

Regression Analyses Involving All Titles

We first present in table 3 the regression estimates based on the full sample containing all titles and disciplines. We use the Library of Congress top-level classification as a control variable, and the results from table 3 indicate that our variables of interest all exert statistically significant impact on patrons' usage preferences. Specifically, the positive coefficients on the variables *Dimension* and *Year_shelf* ($(\hat{\beta}_{dimension} = 0.188, \hat{\beta}_{year_shelf} = 0.156; p < 0.01)$) indicate that patrons' usage preferences move toward the browsing side with respect to an increase in the book's spine length or the time span that the book remains on the shelf. In particular, the coefficient on *Dimension* suggests that when the book size increases by 1 cm in spine length, usage preference moves by 0.188 usage count toward the browsing side. Similarly, a book's usage preference moves by 0.156 usage count toward the browsing side as its shelf time increases by one year.

On the other hand, the regression results of the variables *Page* and *Year_age* demonstrate their different influences on patrons' book usage behaviors. First, the negative coefficients on both variables ($(\hat{\beta}_{page} = -0.000137, \hat{\beta}_{age} = -0.0295)$) suggest that, as the number of pages or the time span of a book being held in the library increases,

TABLE 3
Regression Results for All Titles

Variables	U_{BC}
Page	-0.000137** (5.63e-05)
Dimension	0.188*** (0.00349)
Year_age	-0.0295*** (0.00166)
Year_shelf	0.156*** (0.00121)
Constant	-7.093*** (0.0901)
Discipline controls	YES
Observations	86,740
R-squared	0.190
Standard errors in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$	

usage preferences move toward the checkout rather than the browsing side. In other words, the negative sign suggests that patrons tend to check out a book with more pages or when the book is more aged. However, compared to the impact of variables *Dimension* and *Year_shelf*, the small coefficient on variable *Page* ($\hat{\beta}_{page} = -0.000137$) suggests that there is only 0.0137 usage count movement to the checkout side per every 100-page increase. With regard to the coefficient on the variable *Year_age*, contradicting to our hypothesis that the longer a book stays in local collections, the more likely that patrons tend to use the book as a quick reference in the library, the relevant regression coefficient actually suggests a 0.0295 usage count movement toward the checkout side with the book aging by an additional year.

Regression Analyses by Disciplines

Since book usage preferences can differ across different disciplines (for example, scholars in the humanities tend to consult materials in their original form in order to have direct interactions with the materials),²² we further extend our analysis to subsamples based on four different academic disciplines, including arts, humanities, sciences, and social sciences. As shown in table 4, the regression results within four major disciplines indicate that variables *Dimension* and *Year_shelf* remain statistically significantly positive and consistent with the all-title regression results. Results in table 4 further suggest that, despite their different research interests, patrons across different disciplines tend to use the book within the library as a book's dimension or shelf time increases. Meanwhile, coefficient on the variable *Year_shelf* continues to exhibit small magnitude (ranging from 0.119 to 0.172 across different specifications), while the magnitude of the variable *Dimension* varies to a greater range (from 0.0543 to 0.221). This suggests that time on the shelf has similar impact on patrons' book use regardless of the academic disciplines, while the dimension of a book has the greatest impact on patrons using art books ($\hat{\beta}_{dimension_arts} = 0.221$) and smallest impact on social science book users ($\hat{\beta}_{dimension_social\ sci} =$

VARIABLES	LCC_arts	LCC_humanities	LCC_sciences	LCC_social sciences
Page	0.00136*** (0.000211)	-0.000214*** (6.97e-05)	0.000747*** (0.000148)	-0.000382*** (0.000113)
Dimension	0.221*** (0.00882)	0.0897*** (0.00585)	0.121*** (0.0127)	0.0543*** (0.00958)
Year_age	0.132*** (0.00553)	-0.0578*** (0.00217)	-0.0364*** (0.00489)	-0.0606*** (0.00291)
Year_shelf	0.119*** (0.00397)	0.172*** (0.00160)	0.152*** (0.00361)	0.168*** (0.00210)
Constant	-8.899*** (0.254)	-4.490*** (0.140)	-5.412*** (0.310)	-3.562*** (0.228)
Observations	14,522	40,864	8,688	21,953
R-squared	0.131	0.240	0.186	0.242
Standard errors in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$				

0.0543). Specifically, usage preference moves to the browsing side by 0.221 usage count for art books and only 0.054 usage count for social science books as the length of the book spine increases by 1 cm.

The regression coefficients on the variables *Page* and *Year_age* clearly demonstrate heterogeneity in usage preferences for print books across different disciplines. As shown in table 4, the larger coefficient value of books in arts indicates that patrons tend to use art books in the library to a greater extent when the books become thicker ($\hat{\beta}_{page_arts} = 0.0013$: in other words, 0.13 usage count per 100-page increase) or older ($\hat{\beta}_{age} = 0.13$: in other words, 0.13 usage count per centimeter increase in thickness). As for books in the other three disciplines, except for books in sciences that demonstrate the same trend as art books but with smaller magnitude ($\hat{\beta}_{page_sciences} = 0.0007$: in other words, 0.07 usage count per 100-page increase), coefficient values for the books in humanities and social sciences actually suggest that patrons tend to check out books when they become thicker or older.

Potential Mechanisms and Applications

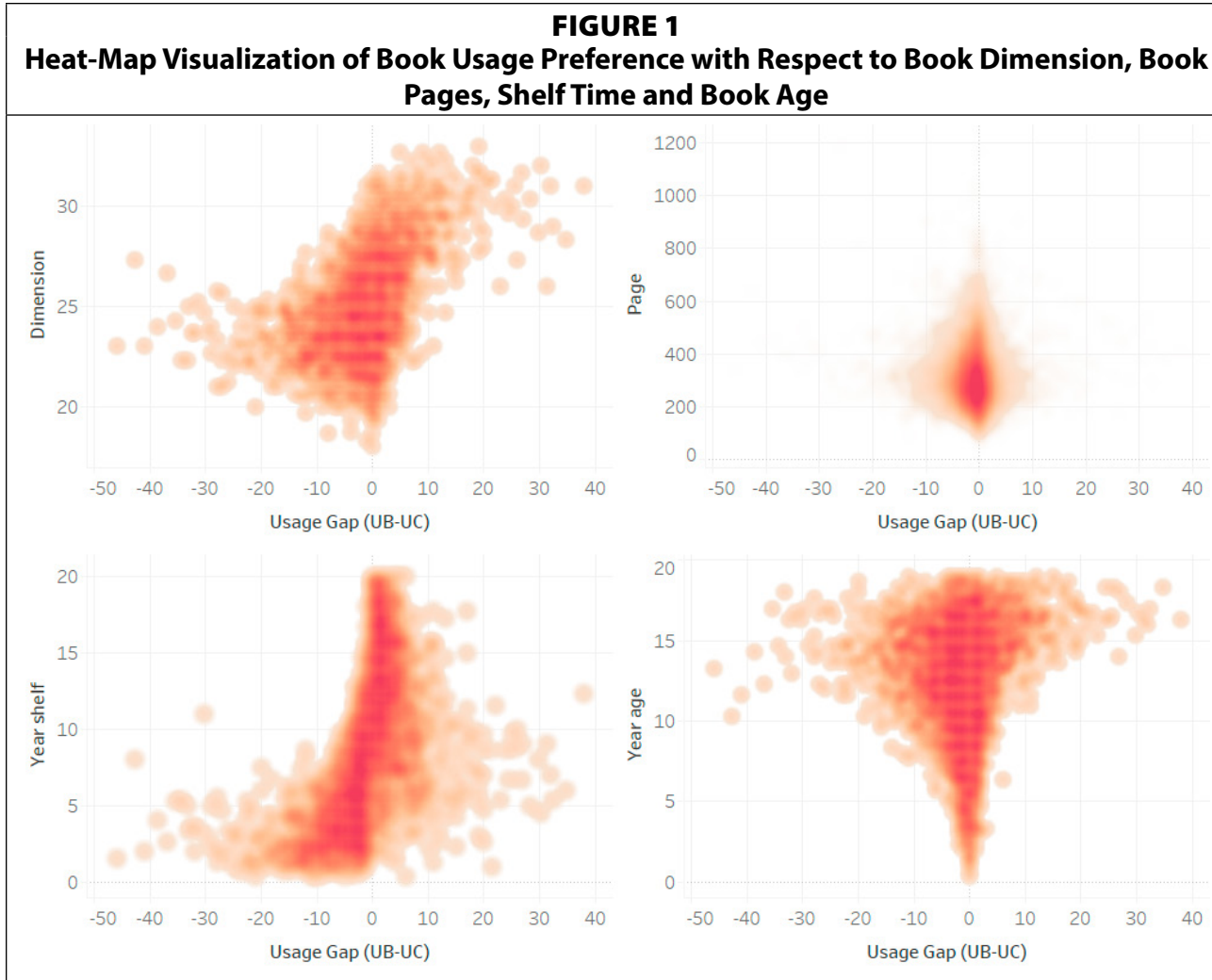
Overall, our empirical analyses deliver two major findings that may provide important implications to academic libraries with weeding needs. First, we find that, across all disciplines, a book's physical dimension and its shelf time can impose a significant impact on patrons' preferences for using a book in the library space. Second, results from the within-discipline regressions indicate that the impact of page counts and book age can be more discipline-specific with users of art books demonstrating the most salient responses to these factors. Next, we explore the potential mechanisms behind these findings and outline further applications.

Preference for In-House Use

In our full sample analyses involving all titles, our results suggest that an increase of a print book's dimension and shelf time will significantly raise the likelihood of users using the material in the library space. We visualize these findings in four heat maps in figure 1. Here, the heat maps of the variables *Dimension* and *Year_shelf* clearly indicate a growing trend of usage preference with respect to the increase in a book's size and time span of nonusage. On the other hand, the usage preference does not demonstrate as clear a usage pattern with respect to the change in a book's time span in library collections or the number of pages.

Through further regression analyses within each of the major academic disciplines, patrons' preferences for in-house use are found to significantly grow with book size across all disciplines. Therefore, we can further conclude that the dimension (spine length) of a book is a good physical feature to consider in the deselecting process. From patrons' perspective, given that it takes more effort for patrons to carry books when they are bulky, it is reasonable to conclude that patrons tend to look for the needed information in the library space without checking out when the books become more difficult to carry.

Similarly, we also observe the significant impact of shelf time on book usage preference through the within-discipline regressions. Therefore, we can conclude that the longer a book stays on the shelf since last checkout, the more likely that patrons would prefer to use the book in the library. On the other hand, given that the variable *Year_shelf* is widely considered as an indicator of whether a book is still applicable to current research needs of the collection users, we can further infer that a book's demand for in-library use (as opposed to checkout) increases as the book becomes unable to meet collection users' current research needs.



Books in Arts

The usage preference for art books demonstrates the most salient response to the four factors in our test. We propose several potential explanations. First, compared to books in other disciplines, a higher portion of books in arts contains rich graphical content and is printed on high-quality paper. As an art book increases its pages or dimension, its weight can grow at a faster rate than books in other disciplines, which subsequently increases the difficulty for patrons to carry outside the library. In addition, the fact that art books tend to contain relatively less textual content allows patrons to read through or locate the target information within a shorter period of time, which can also decrease the likelihood of checking out the book.

Moreover, it is also worth discussing art book users' different responses to the two time-related variables. Based on the regression results shown in table 4, we find that, compared to books in the other three disciplines, art book users are more sensitive to the book's age while responding similarly to shelf time. In our study, book age is considered a measure of the shift in popular topics and publishing format over time (for instance, more graphical elements have been supplemented to textual content due to the convenience provided by digital technology), while shelf time indicates whether the book still possesses relevance to patrons' ongoing research interests. It can therefore be inferred that our art book users

demonstrate a stronger preference toward checking out books with newer topics or being more visually attractive, and they respond similarly with regard to the book's relevance to their research interests.

Impact of Number of Pages

Different from the consistent impact of book dimension, we find that a book's page count exhibits minor and heterogeneous impact on patrons' usage preferences across disciplines. As indicated by its subfield name in MARC record ("Extent"), the number of pages not only indicates the bulkiness of a book but also the amount of content in the book, which may help explain the heterogeneous responses across different disciplines.

Nevertheless, our findings on book pages offer several implications that subject librarians may find helpful in their weeding process. For books with less textual content, like books in arts, that are relatively easy for users to locate the target content through browsing, the impact of book pages thus emerges when patrons need to decide whether it is necessary to check out the book. Such decision-making process will reasonably lead to the most significant usage preference change in those books with less textual content or more inconvenience to carry. On the other hand, with regard to the books with content that patrons tend to digest entirely and progressively, such as novels, the increase in a book's extent can push the book usage preference toward checking out. This also helps explain the relatively inconsistent and insignificant results of books in humanities and social sciences, within which textual content generally takes the majority of the collection holdings. One may further infer that the number of pages is not a universally reliable indicator to be considered in the weeding process, given that books in social sciences and humanities account for 71 percent of the sample collection in this study.

Book Age in Weeding

Academic librarians traditionally adopt book age as a measure for a book's obsolescence or relevance to the current research interest of their community. Yet, consistent with Slote's finding, our statistical evidence also suggests that the time span of a book held in local collections is not a good criterion for making weeding decisions.²³ Similar to the variable *Page*, the coefficients on *Year_age* obtained from the within-discipline regression analyses also demonstrate an inconsistent impact of a book's age on patrons' usage preferences. Although a significant preference for using books outside the library is captured in the full sample involving all titles, the results by different disciplines are much more heterogeneous. Specifically, for older books in our local collections, users tend to use those from the arts in the library while checking out books from other disciplines.

Based on the subject-specific impact of book age, our empirical findings further suggest that librarians should not simply weed books based on a single cutoff, such as removing all books that were published before 1950. Instead, they should incorporate other time-related factors such as shelf time to make a comprehensive decision.

Potential Applications

Collection management understandably requires comprehensive and collaborative inputs from librarians, staff, and student workers. As a result, library staff would benefit from a few easy-to-apply guidelines in their daily work if regular practices can be drawn and streamlined

through a better understanding of users' information consumption behaviors in light of our findings. Based on patrons' discipline-specific responses to the content- and time-related factors as well as tangible characteristics of a book, table 5 summarizes their impact into three categories. Subject librarians can thus take advantage of this summary and selectively incorporate relevant factors into their collection management to build a more user-centered and potentially space-efficient collection.

	Page	Dimension	Year_age	Year_shelf
Arts	high	high	high	high
Humanities	low	median	low	high
Sciences	median	median	low	high
Social Sciences	low	median	low	high

An increasing number of academic libraries have recently started to adopt the concept of operation planning in their collection management. Some examples include that the University Library at Indiana University—Purdue University Indianapolis conducted a systemic study of their shelf occupancy to inform the decision-making process for their space renovation project; or the Kansas State University Library built high-density storage facilities in an existing library building to ease the tension between collections and services.²⁴ We thus suggest a more patron-centered strategy to be integrated into collection management that not only regards books as physical items but also differentiates the deselection criteria based on patrons' usage preferences. A possible scenario can be that when academic libraries are conducting deselection activities for high-density or off-site storage, they may apply a more aggressive threshold for books in the humanities and social sciences than books in the arts, given that patrons will be more likely to use books in these two disciplines outside the library.

Conclusions, Limitations, and Future Research

Employing checkout and in-library use data from an academic library during a span of 20 years, this evidence-based statistical analysis delivers several key findings, including that a book's dimension and time on shelf present significant and consistent impact on patrons' print book usage preferences across all disciplines, and books in arts in general demonstrate the most salient responses. Our study also presents important potential applications for future collection management practices. With a user-centered concept in mind, this study bridges significant gaps in the literature by demonstrating the importance of tangible characteristics in affecting print book usage patterns and by making a methodological contribution through a modular analysis of usage patterns over a single content format. Despite our focus on print materials, our study design and findings could still be of great relevance to today's academic libraries with resources being increasingly accessible in digital format.

It is worth noting several caveats in our study design and findings. First, despite the large sample size compared to previous related studies, our dataset still may not fully capture a complete book usage landscape in the Scribner Library. As explained in the Study, Data and Methodology section, the browsing usage data were not collected until the launch of Voyager

ILS, and the circulation data before 1998 was manually but partially entered into Voyager. Thus, our findings could present a selection issue especially for books in the humanities that tend to see a much longer shelf time.

Additionally, studies analyzing browsing data could suffer from the issue of unobserved browsing uses. When users are interacting with print books, they may only take a glimpse of the book and put it back on the shelf upon finishing. In this case, no relevant browsing uses are captured within the library system. Another issue with our browsing data is that, different from the circulation records, Voyager only records the browsing counts of a book without recording the date information of each browsing. This means that the shelf time is calculated by the latest checkout.

We recognize that patrons in different libraries may demonstrate different book usage patterns due to the types of institutions, student body compositions, research focus, library allocations on campus, library collections, and so on. Nevertheless, this study aims to provide a novel user-centered approach to collection management that will motivate future related evidence-based research in this area. It would be interesting to see whether the conclusions and implications obtained in this study can be generalized if studies with similar design are conducted in other academic libraries with a longer period of usage data or different institutional features.

Acknowledgment

I am grateful to the editor, Wendi Kaspar, and anonymous reviewers for their insightful comments and suggestions that helped greatly improve the quality of this manuscript. I thank discussants from the University of Texas Libraries for helpful feedback. A significant portion of this research was carried out during my prior academic affiliation with Skidmore College.

APPENDIX. Additional Tables

TABLE A1	
Regression Results for All Titles, Excluding Imputed Non-Use Titles	
VARIABLES	U_{bc} (Usage Gap)
PAGE	-0.000266*** (5.78e-05)
DIMEN	0.131*** (0.00366)
Year_age	-0.115*** (0.00218)
Year_shelf	0.202*** (0.00244)
Constant	-4.784*** (0.0939)
Discipline controls	YES
Observations	54,928
R-squared	0.133
Standard errors in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$	

TABLE A2				
Regression Results by Discipline, Excluding Imputed Non-Use Titles				
VARIABLES	LCC_arts	LCC_humanities	LCC_science	LCC_social science
PAGE	0.00145*** (0.000300)	-0.000355*** (0.000102)	0.000875*** (0.000245)	-0.000743*** (0.000173)
DIMEN	0.288*** (0.0123)	0.120*** (0.00887)	0.177*** (0.0206)	0.0667*** (0.0143)
Year_age	0.109*** (0.00973)	-0.176*** (0.00409)	-0.145*** (0.0104)	-0.170*** (0.00563)
Year_shelf	0.163*** (0.0122)	0.281*** (0.00460)	0.236*** (0.0109)	0.249*** (0.00597)
Constant	-10.66*** (0.353)	-4.325*** (0.211)	-6.019*** (0.500)	-2.918*** (0.341)
Observations	9,925	26,460	5,232	14,213
R-squared	0.110	0.138	0.100	0.117
Standard errors in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$				

Notes

1. Barbara Fister, "Repositioning Library Space," in *New Roles for the Road Ahead: Essays Commissioned for ACRL'S 75th Anniversary* (Chicago, IL: Association of College and Research Libraries, 2015), 43–45.
2. "Standards for Libraries in Higher Education," Association of College & Research Libraries (ACRL) (last revised 2018), available online at <https://www.ala.org/acrl/standards/standardslibraries> [accessed 21 June 2019].
3. Herman Howe Fussler and Julian Lincoln Simon, *Patterns in the Use of Books in Large Research Libraries* (Chicago, IL: The University of Chicago Press, 1961).
4. Fussler and Simon, *Patterns in the Use of Books in Large Research Libraries*; William E. McGrath, "Correlating the Subjects of Books Taken out of and Books Used within an Open-stack Library," *College & Research Libraries* 32, no. 4 (1971): 280–85; Anthony Hindle and Michael K. Buckland, "In-library Book Usage in Relation to Circulation," *Collection Management* 2, no. 4 (1979): 265–77.
5. Fussler and Simon, *Patterns in the Use of Books in Large Research Libraries*; Thomas J. Galvin and Allen Kent, "Use of a University Library Collection: A Progress Report on a Pittsburgh Study," *Library Journal* 102, no. 20 (1977): 2317–20.
6. Jeff Selth, Nancy Koller, and Peter Briscoe, "The Use of Books within the Library," *College & Research Libraries* 53, no. 3 (1992): 197–205.
7. Casimir Borkowski and Murdo J. MacLeod, "Report on the Kent Study of Library Use: A University of Pittsburgh Reply," *Library Acquisitions: Practice & Theory* 3 no. 3/4 (1979): 125–51, [https://doi.org/10.1016/0364-6408\(79\)90011-5](https://doi.org/10.1016/0364-6408(79)90011-5); Robert M. Hayes, "The Distribution of Use of Library Materials: Analysis of Data from the University of Pittsburgh," *Library Research* 3, no. 3 (1981): 215–60.
8. Stanley J. Slote, *Weeding Library Collections: Library Weeding Methods* (Englewood, CO: Libraries Unlimited, 1997).
9. Slote, *Weeding Library Collections: Library Weeding Methods*.
10. Amy K. Soma and Lisa M. Sjoberg, "More than Just Low-hanging Fruit: A Collaborative Approach to Weeding in Academic Libraries," *Collection Management* 36, no. 1 (2010): 17–28; Michelle Wilde and Allison Level, "How to Drink from a Fire Hose without Drowning: Collection Assessment in a Numbers-driven Environment," *Collection Management* 36, no. 4 (2011): 217–36; Cory Tucker, "Collection Assessment of Monograph Purchases at the University of Nevada, Las Vegas Libraries," *Collection Management* 34, no. 3 (2011): 157–81.
11. Jennifer E. Knievel, Heather Wicht, and Lynn Silipigni Connaway, "Use of Circulation Statistics and Interlibrary Loan Data in Collection Management," *College & Research Libraries* 67, no. 1 (2006): 35–49.
12. Steven A. Knowlton, "A Two-step Model for Assessing Relative Interest in E-books Compared to Print," *College & Research Libraries* 77, no. 1 (2016): 20–33.
13. Weijing Yuan, Marlene Van Ballegoie, and Jennifer L. Robertson, "Ebooks versus Print Books: Format Preferences in an Academic Library," *Collection Management* 43, no.1 (2018): 28–48.
14. Rusty Kimball, Gary Ives, and Kathy Jackson, "Comparative Usage of Science E-book and Print Collections at Texas A&M University Libraries," *Collection Management* 35, no. 1 (2009): 15–28.
15. Kate B. Moore, "Are We There Yet? Moving to an E-only Collection Development Policy for Books," *Serials Librarian* 68, no. 1/4 (2015): 127–36, <https://doi.org/10.1080/0361526X.2015.1016836>; Kimball, Ives, and Jackson, "Comparative Usage of Science E-book and Print Collections at Texas A&M University Libraries," 15–28.
16. Coen Wilders, "Predicting the Role of Library Bookshelves in 2025," *Journal of Academic Librarianship* 43, no. 5 (2017): 384–91; Kelsey Corlett-Rivera and Timothy Hackman, "E-book Use and Attitudes in the Humanities, Social Sciences, and Education," *portal: Libraries and the Academy* 14, no. 2 (2014): 255–86, <https://doi.org/10.1353/pla.2014.0008>; James Bierman, Lina Ortega, and Karen Rupp-Serrano, "E-book Usage in Pure and Applied Sciences," *Science & Technology Libraries* 29, no.1/2 (2010): 69–91; Helen Belefant-Miller and Donald W. King, "How, What, and Why Science Faculty Read," *Science & Technology Libraries* 19, no. 2 (2000): 91–112.
17. David Nicholas et al., "UK Scholarly E-book Usage: A Landmark Survey," *Aslib Proceedings: New Information Perspectives* 60, no. 4 (2000): 311–34; Suzanne M. Ward, Robert S. Freeman, and Judith M. Nixon, *Academic E-books: Publishers, Librarians, and Users* (West Lafayette, IN: Purdue University Press, 2015).
18. Christine Wolff, Alisa B. Rod, and Roger C. Schonfeld, "Ithaka S+R US Faculty Survey 2015," Ithaka S+R (2016), <https://doi.org/10.18665/sr.277685>.
19. Pauline Dewan, "Are Books Becoming Extinct in Academic Libraries?" *New Library World* 113, no. 1/2 (2012): 27–37; Silas Marques De Oliveira, "E-textbooks Usage by Students at Andrews University," *Library Management* 33, no. 8/9 (2012): 536–60; Wilders, "Predicting the Role of Library Bookshelves in 2025," 384–91.
20. Knievel, Wicht, and Connaway, "Use of Circulation Statistics and Interlibrary Loan Data in Collection Management," 35–49; Kimball, Ives, and Jackson, "Comparative Usage of Science E-book and Print Collections at Texas A&M University Libraries," 15–28; Knowlton, "A Two-step Model for Assessing Relative Interest in E-

books Compared to Print," 20–33; Yuan, Van Ballegoie, and Robertson, "Ebooks versus Print Books," 28–48.

21. Slote, *Weeding Library Collections: Library Weeding Methods*.

22. Sue Stone, "Humanities Scholars: Information Needs and Uses," *Journal of Documentation* 38, no. 4 (1982): 292–313.

23. Slote, *Weeding Library Collections: Library Weeding Methods*.

24. William H. Weare Jr, Paul Moffett, and John P. Cooper, "Preparing for Renovation: Estimating Shelf Occupancy to Inform Decision Making Regarding the Redevelopment of Library Space," *Collection Management* 41, no. 3 (2016): 168–81; Michelle R. Turvey-Welch and Roberta L. Johnson, "Repurposing and Renovating Space for High-Density Storage: Kansas State University Libraries' Answer to Additional Collections Space," *Collection Management* 42, no. 2 (2017): 109–19.