

Contextualizing and Interpreting Cost per Use for Electronic Journals

Matthew Harrington
Virginia Tech, mdharrin@ncsu.edu

Connie Stovall
Virginia Tech, cjstovall@ua.edu

Follow this and additional works at: <http://docs.lib.purdue.edu/charleston>

An indexed, print copy of the Proceedings is also available for purchase at: <http://www.thepress.purdue.edu/series/charleston>.

You may also be interested in the new series, Charleston Insights in Library, Archival, and Information Sciences. Find out more at: <http://www.thepress.purdue.edu/series/charleston-insights-library-archival-and-information-sciences>.

Matthew Harrington and Connie Stovall, "Contextualizing and Interpreting Cost per Use for Electronic Journals" (2011). *Proceedings of the Charleston Library Conference*.
<http://dx.doi.org/10.5703/1288284314928>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

Contextualizing and Interpreting Cost per Use for Electronic Journals

Matthew Harrington, Collection Management, Virginia Tech

Connie Stovall, Collection Management, Virginia Tech

Abstract:

Cost and usage have been consistent elements among both serials decision databases and commercial decision support systems, and the cost per use calculation has become a well-established criterion for assessing electronic subscriptions. However, it is just a numerical value until it can be plotted along several axes related to its components. Mapping these calculated values within and across platforms and subjects allows them to be read through multiple contexts to define what is relatively “high” or “low,” and establishing the relative averages and benchmarks within these multiple contexts informs the difficult serials decisions often faced. This presentation looks at ways in which cost per use, as well as other cost and use calculations, has been incorporated into Virginia Tech’s relational database for serials decisions in order to arrive at an understanding of what those values mean in both the immediate context and the larger picture.

Introduction

As resource allocations continue to shrink, academic libraries increasingly rely on statistical data to assist with tough decisions about serials subscriptions, and ultimately, when faced with dwindling funding and the prospect of subscribing to one journal title or package over another, collection managers attempt to determine how valuable a resource will be to the institution’s students and researchers. Commonly, collection managers analyze usage data and calculate cost per use. However, the utility of such calculations are diminished when the relationship between those calculations and the “value” of a journal is unclear. To more accurately assess what usage statistics really mean and how they relate to cost, more contextualized input is needed, and to aid in this task, a relational database has been constructed at Virginia Tech. This database pulls data together in order to establish baselines, highlight outliers, and examine what this data is capable of telling us, and the results of this analysis have become a critical component in the process of assigning value to serials at Virginia Tech.

Use and Exchange Values for Library Resources

When it comes to the concept of value, few have contributed as much as Karl Marx. Marx asserted that value was dual in nature, meaning that things, or commodities, can be expressed both by their use-value and their exchange value. Exchange value and use-value are independent measurements, with the former reflecting what one would trade for the commodity and use-value defined by what that

commodity is worth to the one using it (Marx 1906, 42-3). Aalto Repo builds upon these same ideas in his examination of the value of information. However, when information is thrown into the mix, use values are more relevant to information need, and exchange values relate to information sold as a commodity (375). This distinction between information and information products is somewhat lessened in the context of the university’s online resources, in which the ability to retrieve necessary research is almost directly related to subscription costs for those resources.

Library usage is basically a quantitative expression of use value. It is tied to a very specific definition of use and therefore cannot be compared beyond the “use” it measures. Gate counts, database searches, article downloads, reference questions, and shelving statistics constitute different and incomparable quantitative translations of a “use.” At the fundamental level of assessment, what was used, how often it was used, why it was used and how it was used all contribute to our understanding of patron behavior and the ability of collections to address the needs of those patrons. However, no single measurement answers all use-related questions, and therefore, no single measurement of “use value” exists. Rather than dismiss the shortcomings of any single measurement, understanding what each measurement says and does not say about the “value” of a collection is a more practical approach to usage assessment.

COUNTER-compliant journal usage reports count the number of article downloads by title. This measure does not address how articles were used after being downloaded, and it does not answer questions concerning how a particular article was chosen. It simply counts what was downloaded. In a 2008 survey of academic libraries, this report was “the usage statistic identified as most useful by survey respondents” (Baker and Read 2008, 53). Article downloads are fairly large scale, uniform expressions of interest in journal articles, and their usage allows us to examine what is being requested, even if we do not know how it was later used. Usage is also tied into the educational aspect of the university. Students rarely publish and therefore that “use” is unrepresented in citation factors, which are grounded almost solely in the university as a research institution. Impact factors and local citation statistics show what has been incorporated into scholarship. However, academia, in both its educational and research bases, assumes a constant engagement with scholarship, regardless of whether those scholarly works find their way into a published list of citations. Knowing what scholars view on a grand scale certainly sheds light on assessing electronic journal collections and provides a fairly stable foundation for other projects, which might examine through focus groups or surveys how those articles were then used.

The exchange value of an electronic resource is typically expressed as a subscription cost, and comparisons of resources can only be made in terms of exchange value, primarily because these values exist within a system of currency and are, for the most part, determined by the larger market. However, resources, and more specifically in the context of this paper, journal titles, vary to such a degree that comparing costs alone is problematic. A large disparity exists in subscription costs across disciplines, and numerous other factors, such as the number of articles published each year, affect the price set for a given title. The cost per use calculation attempts to address this issue.

Interpreting Cost/Use

The components of the cost per use calculation require a few assumptions. For example, each element is associated with a different definition of time. For journal subscriptions, cost corresponds to

the subscription year, and refers to a specific year or years of publication. Use, on the other hand, refers to the period in which an article from that title was downloaded, regardless of publication year. Therefore, downloads counted on a JR1 are likely associated with costs from multiple years. But for practical reasons, it is assumed that for currently subscribed titles, the usage of that title over one year is roughly equal to the overall usage of a single publication year. The more recent JR5, which reports usage by publication year, may allow this assumption to be tested once it gains more popularity. However, the JR5s collected at Virginia Tech over the past year indicate a significant portion of usage can be attributed to the most recent year, and CIBER Research Limited found, in their Stage 2 Final Report on the Journal Usage Factor, that “[e]ven with a 6-month window around a sixth of lifetime use would be captured” (23).

A second assumption relies on the validity of this quotient as a measure of value. The cost of a resource and its use by students and researchers are two of the primary concerns when assessing an academic resource. Balancing the demand evident by use with the limitations imposed by the budget requires a calculation that can address both issues because use and exchange values are not necessarily mutually determined. Cost per use calculations attempt to incorporate use value into the exchange value by way of a ratio, thereby allowing both elements to factor into a calculated “value.” In this sense, the cost per use calculation is an intermediary value mutually determined by both the use value and exchange value. It reconciles the two relative values.

Cost per use is in no way the end of the serials decision process, nor is it a replacement for the qualitative expertise of trained librarians. However, it is a tool that will raise flags on certain titles that fall outside the typical ratio of cost. For this reason, it is a time-saver rather than a decision-maker. This calculation allows high cost titles with high use to be grouped closer to the cluster of titles around a typical value, and it leaves low use, high cost titles standing out from the “norm.” To establish a “norm” and to analyze cost per use calculations across the entire current electronic journals collec-

tion at Virginia Tech, a MS Access database has been developed.

The Serials Review Database at Virginia Tech connects journal usage reports collected from more than 100 serials vendors to the subscription costs attached to order records in the library's catalog. This is a complex connection because a relatively stable relational database must adapt to the instability of academic resources. In addition to the title changes, platform changes, and constant resource additions, it must also be able to historicize data in order to present longitudinal reports. The goal of the database is to present consolidated reports based on a varied set of criteria, which might include timeframe, subject area, platform, or budgetary fund code. Without describing the database in too much detail, a series of queries, built upon each other, run to generate the final report. This report pulls together usage from multiple years, tying the usage to the appropriate subscription costs each year and calculating a cost per use for each subscription. The reports produced from this database have played an essential role in the serials decisions over the past few years. The data is fairly comprehensive, incorporates multiple years, and can be updated and manipulated easily.

Contextualizing Cost/Use

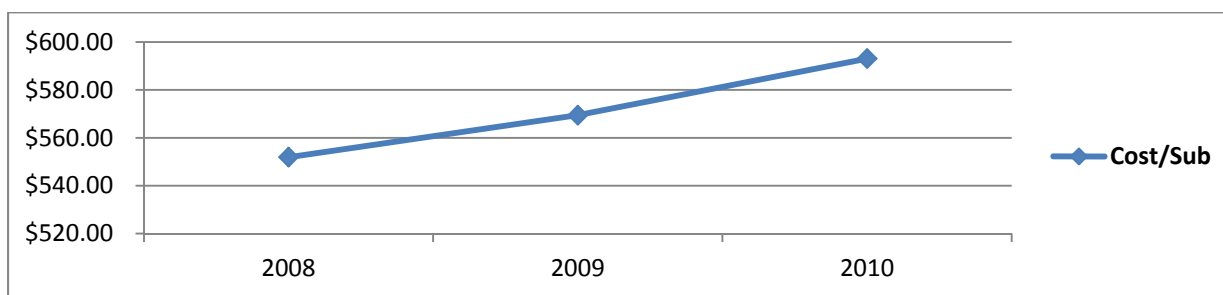


Fig. 1 Calculated cost per subscription from 2008-2010

These findings only reinforce what has already been demonstrated at other university libraries. However, from this three-year sample, we were able to uncover additional information. Usage graphed by semester and by month show distinct patterns across time. Electronic usage of e-journals is clearly higher during the Spring semester at Virginia Tech, and the two peak periods for usage are April and

A query was recently created to include all subscriptions that had remained in our online journal collection from 2008 through 2010 and had available usage data for all three years. The goal of this project was to establish a stable sample of subscriptions to analyze within several different contexts. The results were limited to non-consortial titles. Though this cut out more than half our online journal collection, comparing consortial costs to non-consortial costs is problematic for obvious reasons. What remained included 4,224 subscriptions with 4,060 unique journal titles. Although this represented just 18% of our total e-journal collection in 2009, it made up 50% of all e-journal usage collected through COUNTER JR1s that year and approximately 70% of the budget spent on e-journals, or roughly \$2.4 million.

The results from this analysis point to several trends in our collection. Not surprisingly, the combined cost of these 4,224 subscriptions rose more than 7% from 2008 to 2010, or about 3.66% annually. To make matters worse, the serials budget dropped 3% from fiscal year 2007/2008 to 2009/2010. The budget has since risen back to the 2007/2008 level, but inflation is still outpacing the budget's small growth, prompting the need to continually scrutinize the value of online journal packages and subscriptions.

October. Figures 2 and 3 show that usage is not constant throughout the year, and because the cost per use calculation is based so heavily on usage, these graphs suggest usage is best analyzed annually in order to capture all the highs and lows and establish a baseline value for additional calculations. CIBER Research also pointed out a wide variation of usage at the monthly level in their analysis (5).

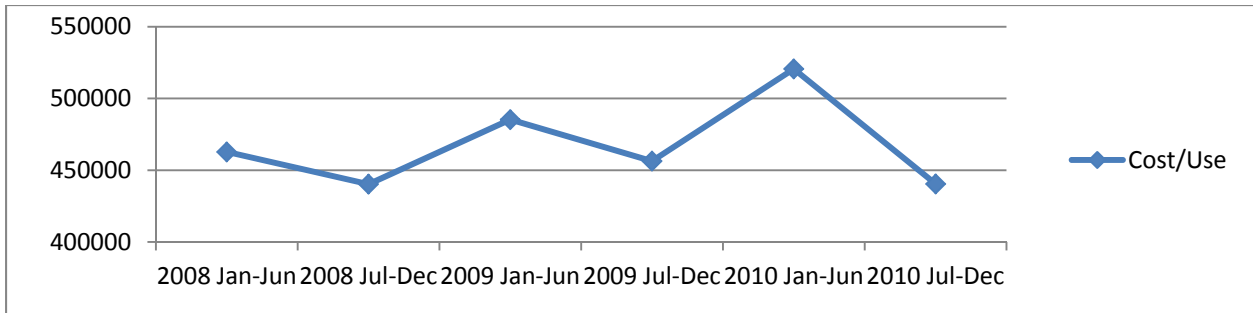


Fig. 2 Usage totaled by semester

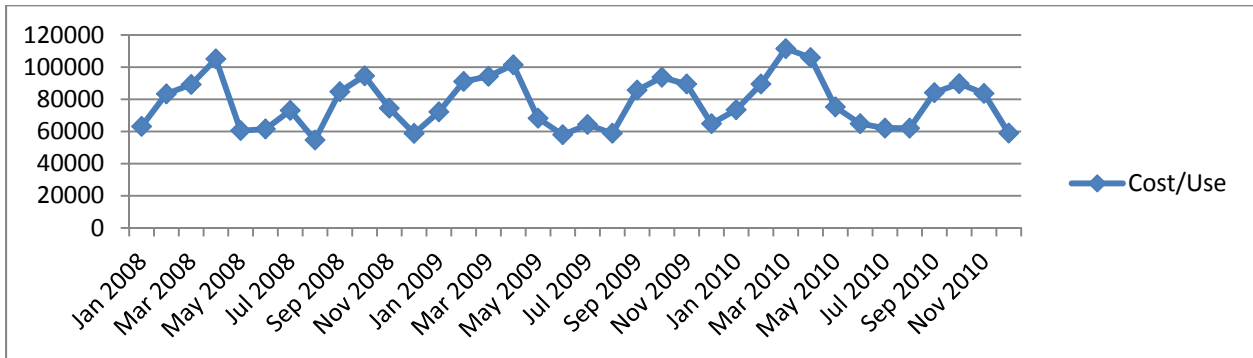


Fig. 3 Usage totaled by month

At this point, it is important to clarify how the baseline is calculated. The total cost per use for these 4,224 subscriptions is slightly higher than \$2.50, if the total cost is divided by the total usage. The median value is close to \$4.00 per use. However, the average cost per use for one of these journals is nearly 10 times as high as the total cost per use. Averages are skewed when there are several values existing at

distant extremities. The recent study on Usage Factor concluded similarly while analyzing usage with a wide range of values between the highest and the lowest counts, and they recommended basing a Usage Factor off median values for this reason (CIBER, 5). Figure 4 shows the total and median cost per use values appear to be more stable and a better representation of a baseline value.

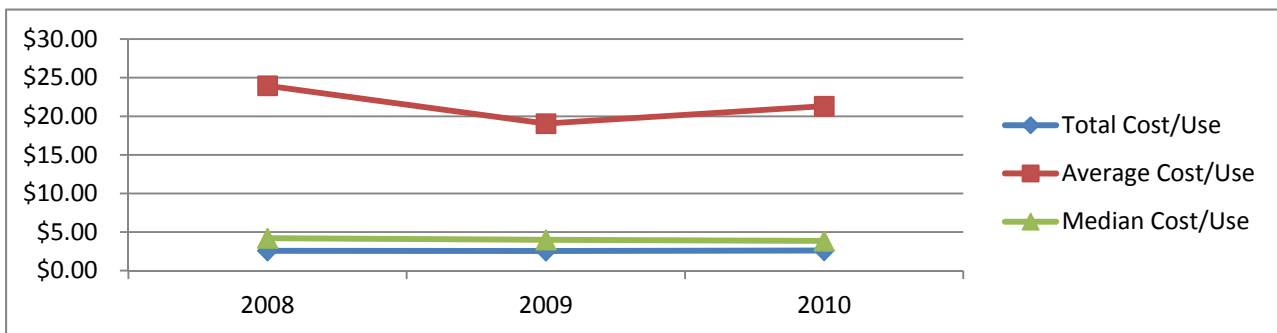


Fig. 4 Comparison of the total, average and median cost per use values

To establish a disciplinary context for the cost per use calculation, the three-year values from the sample were graphed according to subject classification. The subjects were limited to those with at least 10 subscriptions classified under each heading. When

graphing the cost per title for each of the subject areas, the upper call number ranges, i.e. the science and technology subject areas, clearly show an increased cost per title. Physics, plant culture and aeronautics all peak above \$1,300 per title. The Q call

numbers through the TS call numbers are all well above the \$500 per title line, and below Q, only 1 classification (oceanography) costs more than \$600

per title. This disparity between STEM programs and the humanities and social sciences makes it difficult to compare titles based solely on cost.

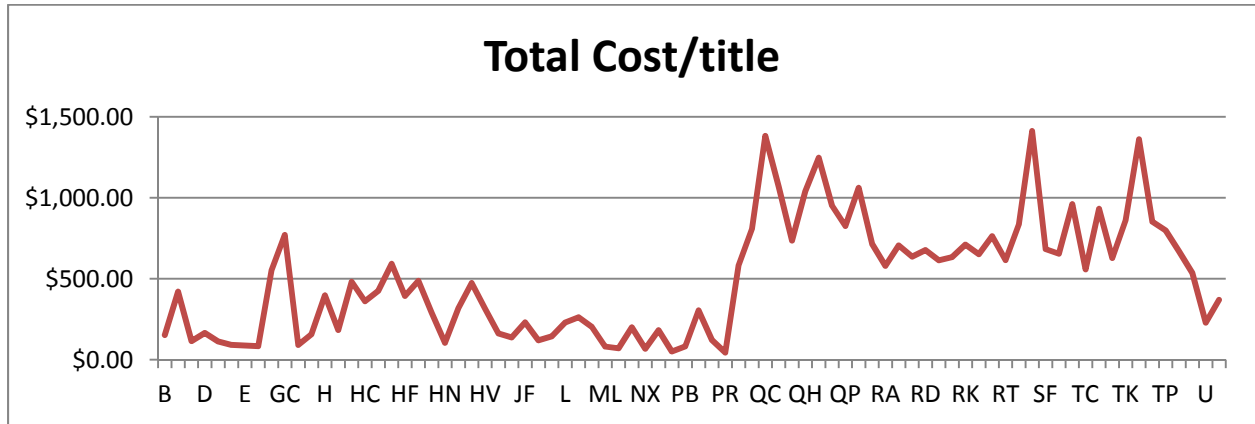


Fig. 5 Cost/title by subject (limited to 80 subjects containing at least 10 subscriptions each)

Cost per use calculations do not show the same lopsided graph. There is variation across subject classifications, but aside from the medical fields, the graph is much more balanced between the disciplines. The medical fields, of course, stand out with the highest cost per use values, three of which reach above \$10.00 per use. But from this analysis, it is easier to compare values across the board in

order to determine what subject areas have higher cost per use values. Oceanography still has a prominent peak, but linguistics and finance also stand out as high cost per use areas of scholarship, and when comparing individual titles, this variation, which is still fairly significant, can provide a valuable context for examining the relative “value” of that title within that subject area.

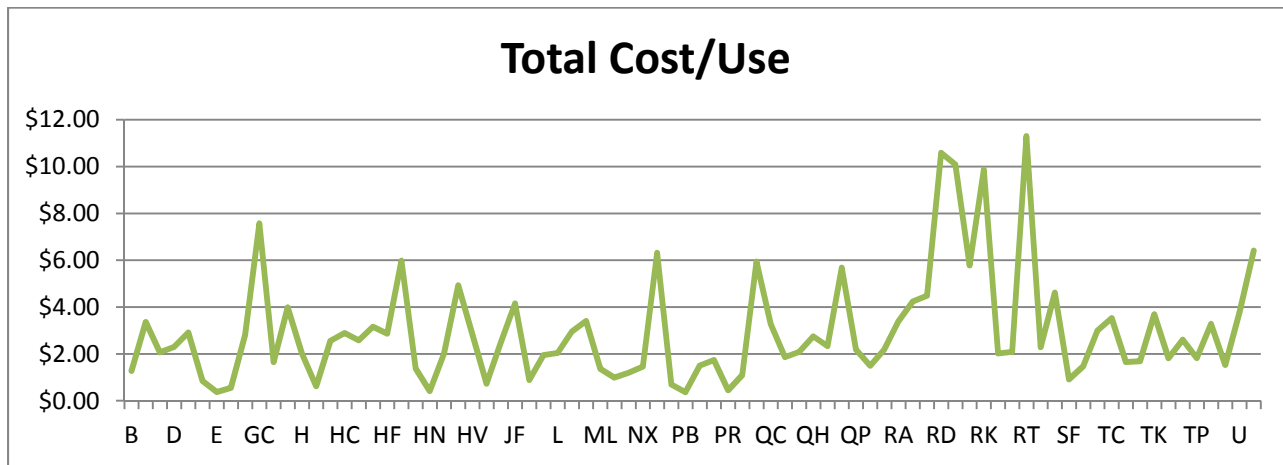


Fig. 6 Cost per use by subject (limited to 80 subjects containing at least 10 subscriptions each)

Subject headings do not provide the only context. Most of the analyses over the last two years at Virginia Tech have specifically looked at the value of journal packages, which often span individual disciplines even when broadly focused on a subject. In

2009, 51% of the serials budget, excluding what was spent on databases, paid for just five journal packages, so we analyzed these five large packages and calculated cost per use and cost per title in order to begin making comparisons.

Serials Budget

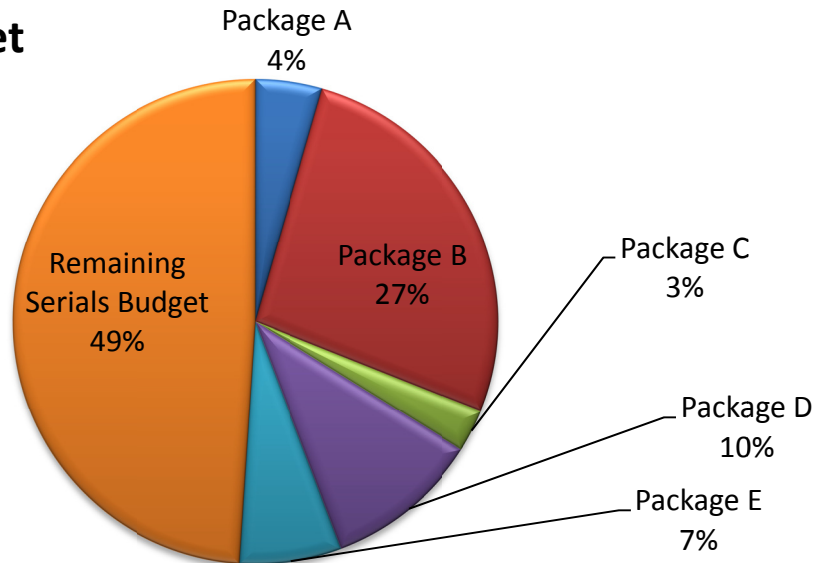
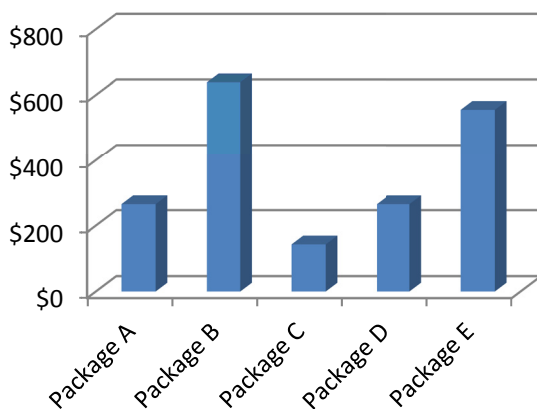


Fig. 7 Percentages of the serials budget spent on 5 journal packages in 2009

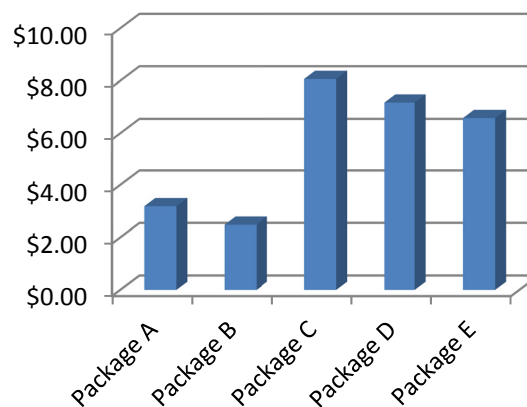
Although Package C had the lowest cost per title, we found it had the highest cost per use among these journal packages. The cost per title graph illustrates the gulf between the cost of subscriptions in a largely STEM-oriented package such as Package B and Package C, which focused on social sciences and humanities. These two packages mark the high and low within the comparison. However, the cost

per use graph indicates those roles are reversed when usage is incorporated. In other words, in their respective disciplinary contexts, the cost difference between the two packages made sense, but when usage is considered, which should more evenly distribute values across disciplines, Package C stood out. Making this comparison led to a deeper analysis of Package C.

Cost/Title



Cost/Use



Figs. 8-9 Cost/Use and cost/title by 5 largest platforms using 2009 data

After examining usage within the package, it was clear that zero-use items were common, and the bulk of usage was divided among a few titles. Another cross-package comparison was made, which looked at the percentage of high and low use titles

within each package. The results roughly matched the cost per use comparison, indicating that the “value” of Package C may be less than a strictly title-by-title subscription to the most heavily used titles within the package.

High and Low Usage Counts

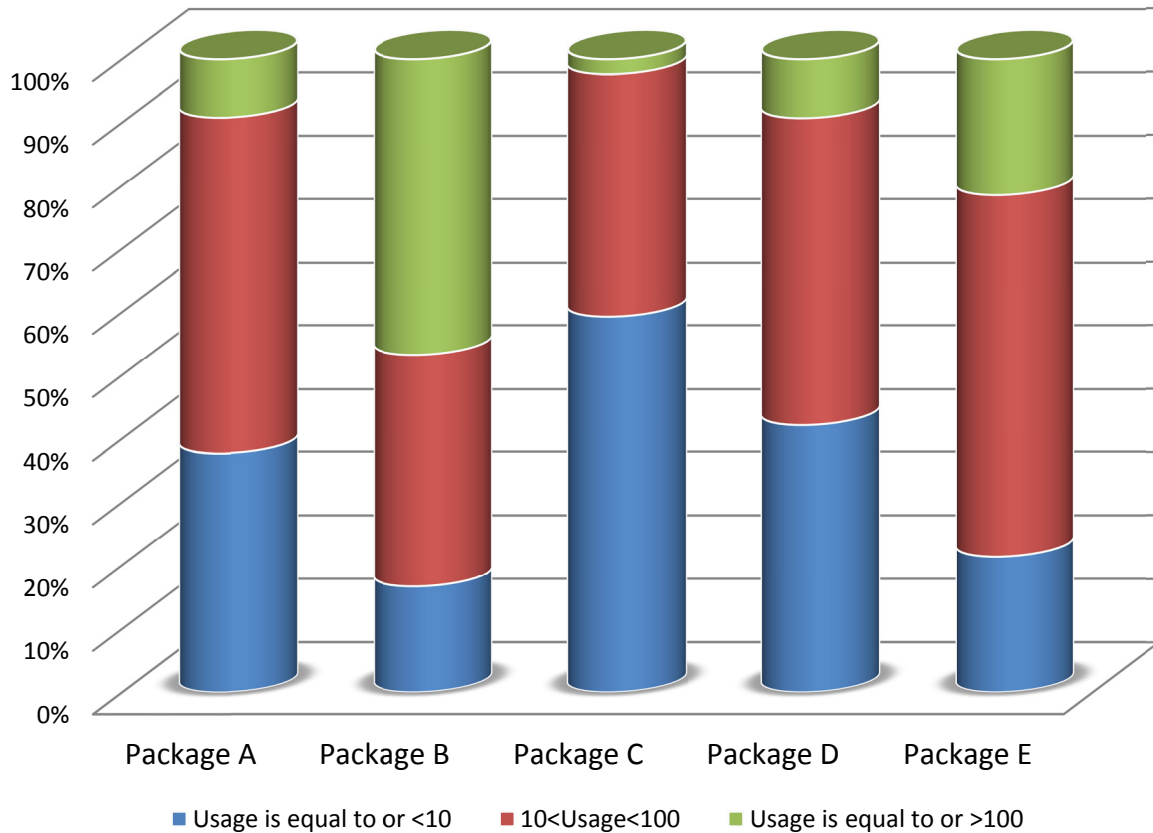


Fig. 10 Percentages of high and low journal titles by package

Conclusion

Several theories attempt to explain how we determine value, and most would agree that making such a determination is multifaceted and that a certain degree of subjectivity exists in any attempt to determine value. Valuing information adds an additional complexity. However, collection managers must constantly make judgments based on value. Simply comparing title costs presents a skewed perspective on library collections across the board, and calculating return on investment or contingent valuation can be impractical as an ongoing metric tool for assessing collections. Cost per use calculations, on the other hand, are fairly easy to compile because the data has become so readily available. But before they are uncritically accepted or too casually dismissed, a deeper understanding about what they are, what they can tell us, and where their limits lie is needed.

Bibliography

- Baker, Gayle, and Eleanor J. Read. (2008). "Vendor-Supplied Usage Data for Electronic Resources: A Survey of Academic Libraries." *Learned Publishing* 21: 48-57.
- CIBER Research Limited. (2011). "The Journal Usage Factor: Exploratory Data Analysis." http://www.projectcounter.org/documents/CIBER_final_report_July.pdf
- Marx, Karl. (1906). *Capital: A Critique of Political Economy*. Translated by Samuel Moore and Edward Aveling. New York: Random House, Inc.
- Repo, Aatto J. (1986). "The Dual Approach to the Value of Information: An Appraisal of Use and Exchange Values." *Information Processing & Management* 22: 373-83.