Citations and Circulation Counts: Data Sources for Monograph Deselection in Research Library Collections

Bruce White

Studies of data-driven deselection overwhelmingly emphasise the importance of circulation counts and date-of-last-use in the weeding process. When applied to research collections, however, this approach fails to take account of highly influential and significant titles that have not been of interest to large numbers of borrowers but that have been highly cited in the literature. It also assumes that past borrowing activity is a reliable indicator of future usage. This study examines the correlations between past and future usage and between borrowing and citation of monographs and concludes that both data elements should be used when deselecting research monographs.

niversity libraries in the second decade of the twenty-first century find themselves facing a number of critical challenges as they negotiate changing environmental demands, new technologies that complicate their mission as much as they facilitate it, and an ongoing struggle to maintain their relevance and centrality within their host institutions. The teaching-research nexus remains at the core of what universities do; consequently, library collections need to meet the needs of a wide range of constituents, from first-degree students through postgraduate and doctoral students to world-class researchers working at the very highest level. Library collections aim to satisfy the information demands of all of these groups through the purchase of large journal packages, through extensive monograph collections at all levels, and through the provision of support facilities and services geared to their common need to access, manage, and use the best and most appropriate information. This study will look at how this complex set of demands is played out in one particular sphere, the decision process around deselecting dated print monographs or relegating them to storage collections to maintain current collections of high-value items for the support of teaching and research.

Libraries of all kinds periodically remove books from their shelves and dispose of them or move them into "just in case" storage collections, and this is as true of university libraries as it is of other types. The drivers of this activity are the need to conserve building space as new acquisitions fill existing shelves to capacity,[1] the

decline in circulation of print books as e-books become more ubiquitous,[2] the growth of other demands on library value space from new technologies, student activities, or organizational needs,[3] and concerns about dated and obsolete materials crowding out more recent works.[4] If deselection is done well, students are presented with strongly focused collections of current material, while researchers have access to core collections that represent the depth and breadth of their fields and that make a visit to the library shelves worthwhile.

The demands placed on the academic researcher have grown considerably in recent decades. In the past, research may have been seen by some academic staff as being something of an optional extra, or a long-term goal with less immediate urgency than teaching, but there is a growing demand to produce evidence of solid peer-reviewed research publications. Promotion and tenure at American universities typically requires evidence of publishing in reputable journals,[5] while academic researchers in other jurisdictions are subject to a range of more centralized research evaluation regimes: for example, the Research Excellence Framework (REF) in the United Kingdom, Excellence in Research for Australia (ERA), and Performance Based Research Funding (PBRF) in New Zealand. Doctoral programs are a further driver of research, with students expecting to be supervised and mentored by active researchers with international credentials backed by credible publications. Doctoral students are themselves researchers and account for a significant proportion of the research activity and outputs of their host institutions.[6] Library collection management activities must therefore take cognizance of the need to support intensive high-level research, and this applies as much to deselection as it does to other collection management processes. Over time, a library collection is the sum of items acquired minus those removed, and a core research collection must give its users access to both the latest research and to items of enduring relevance and worth.

There are a number of methods that can be used, separately or in concert, to decide which books should be removed from a collection and which retained. Publication date will generally be an important factor, and libraries will not normally choose to consider books for deselection within the first ten years of their date of publication; beyond that, however, there are (broadly speaking) two approaches to decision making—the use of expert knowledge in some form or the use of data. Expert knowledge may be solicited from subject experts about which works have enduring merit and which are not worth retaining, or it may be located from lists of high-value titles such as *Choice Reviews Online*. Data, on the other hand, generally emanate from users, specifically borrowers, and normally consist of both total circulation counts and dates of last use, although it can include numbers of citations to specific works or authors as well. From the initial work by Trueswell,[7] circulation and date-of-last-use data have dominated studies of monograph deselection and continue to do so.[8] The rationale for this is clear and, on the surface at least, compelling. The Pittsburgh study in 1976 pointed to a strong consensus that "past use is the best indicator of future use of material."[9] and this was endorsed by Dinkins in 2003: "At the root of all library acquisitions decisions is the goal to add to the collection those materials that will meet users' needs; compiling usage statistics is an effective way of assessing previous decisions and allows extrapolation for the future."[10] Added to this, the ease of collection of circulation and date-of-last-use data through Library Management Systems[11] makes circulation data–driven deselection an obvious choice for libraries looking to cull collections of unnecessary items.

By contrast, the use of citations as a data source for deselection decisions is much less frequently mentioned and appears to have aroused little interest or support. In *Fundamentals of Collection Development Management*, Johnson wrote that "Citation studies assume that the more frequently cited publications are the more valuable, will continue to

be used heavily, and, consequently, are more important to have in the library collections. Data from citation studies can guide cancellation and retention decisions." However, she regarded them primarily as a guide for decisions on journals and was unconvinced of their value in relation to monographs: "citation analysis is time consuming and labor intensive. Important materials for consultation or background work may not be cited frequently."[12] The Association of Research Libraries *SPEC Kit 337: Print Retention Decision Making* noted that "Citation counts are often used as a criterion" in decisions on deselection; but no specific questions on the topic were asked in the survey of American academic libraries—and the technique barely rates a mention in the summary of results—while the use of circulation data is noted in nearly all of the reports.[13] A number of studies on the use of citation data in collection management relates to the analysis of citations by staff or students of a specific institution,[14] and these provide useful insights into the value of the collection in specific cases; but the absence of a standard methodology makes comparisons between these studies difficult.[15] Detailed local studies of this kind are labor-intensive; while they can reveal rich insights into the use of library collections by scholars, they are essentially retrospective and not well adapted to the needs of a deselection exercise, which needs to be able to make determinations about all of the books in a collection, most of which will not have been cited by local authors.

The use of citation data from citation databases (as opposed to local sources of references) for the purposes of collection management is less well developed. As Enger noted, "While citation analysis has been used extensively to manage journal collections, it has not been used to develop book collections. It is, however, one measurable way to effectively manage them. Using citation analysis to develop core book collections in academic libraries is discipline centered and goes beyond the walls of individual libraries to include material discussed by scholars in the academic literature."[16] Enger did not specifically deal with deselection, but he used a citation-based method to identify key authors in specific subject areas and then looked at the holdings of the books they had authored. This information was used to identify and select titles suitable for core research collections based on author reputation, and the later circulation data from titles selected by this method were compared to titles selected by conventional methods. The only author to have directly suggested the use of database-derived citation data for the purposes of deselection was Burdick, who used Science Citation Index (SCI) as a "safety net" for circulation figures in the course of a deselection exercise on a range of low-use science books more than ten years old.[17] Due to the technical limitations of SCI at that time, her sample was small (79 titles) and she referred to it as a "pre-experimental design," but she made the interesting suggestion that "a correlation of citation frequency with circulation statistics would provide evidence of the value of citations in predicting use." She went on, however, to suggest that citation counts had a value in the deselection process independent of usage considerations and that "citation frequency can be used to supplement other measures such as faculty input and frequency of use." This study directly addresses the question of whether citation rates correlate with, and are predictive of, circulation counts while endorsing Burdick's view that circulation counts alone are localized and contingent and therefore need to be triangulated against external citation data to achieve "a less parochial perspective to long-range decisions for collection development in research libraries." If there is a low correlation between circulation and citation rates, they are measuring different properties; and, if it is further accepted that citation is a measure of scholarly value, then its circulation count will be only a partial measure of a book's value within a collection. On the other hand, if there is a high positive correlation, then the use of circulation counts on their own will ensure that a circulation-based method of deselection is an effective way of maintaining and enhancing the scholarly value of a research collection.

The Research Questions

This study looks at four questions relating to books published at least fifteen years before the present:

1. Are historic monograph circulation counts a robust predictor of future circulation counts?
2. Are historic monograph citation counts a robust indicator of future citation counts?
3. Do local-institution monograph citation counts mirror total citation counts?
4. Do monograph circulation counts correlate with, and can they be used as a proxy for, citation counts?

The Context

Massey University is a research university in the North Island of New Zealand. From its beginnings as an agricultural college in the Manawatū district in the 1920s, it became a university in 1964 offering a broad range of subjects both on campus and by distance, and it has subsequently opened campuses in the cities of Auckland and Wellington. With a current enrollment of more than 1,000 PhD students and an annual research output of around 1,400 published journal articles per year (Scopus database), it supports research programs over many disciplines in the sciences, social sciences, humanities, creative arts, and business with a strong emphasis on life sciences, environmental science, agriculture, and food production. The library has developed in line with international trends, adopting the VTLS integrated library system (ILS) in the 1980s and then switching to Innovative in the 1990s. Up until the mid-1990s, most book selection was carried out by academic staff, but this has largely been taken over by liaison librarians working through automated recommendation systems. Obviously there is a strong emphasis on New Zealand material; in the early 1990s most of this was transferred from the general shelves to a separate New Zealand collection, a development that has been significant for the present study. The collections are organized by the Dewey system.

Methodology

Two ranges of books in the main lending collection of the Manawatū Library were chosen for the survey: 591 (Specific Topics in Zoology) and 324 (The Political Process). As noted, because of the existence of the New Zealand collection, specifically local titles were not included in these ranges. Because most New Zealand animal species are endemic (occurring in this geographic al region only), they attract relatively little research from outside the country—with the result that specifically New Zealand zoological titles are cited well less than average, while the books themselves are borrowed much more than average. This would have skewed the results if they had been included in the sample. The same applies to the books on politics: local titles are of intense local interest and are borrowed heavily, while they attract little research attention by international standards. The result of this exclusion is that the two samples are relatively generic and represent more or less typical samples of what would be found in university libraries of comparable size.

The following information on books with a publication date of 2000 or earlier was downloaded from the ILS to Excel spreadsheets:

- Call numbers
- Authors and editors (MARC tags 100 and 700)
- Titles and subtitles (MARC tags 245a and 245b)
- Publication dates
- Circulation counts
- Dates of last check-in

Where there were multiple copies of titles, circulation counts were automatically totaled and the most recent check-in date was assigned to the title, so that each title had a single circulation count and last check-in date. The ILS stores initial checkouts and renewals separately; renewals were deliberately omitted from the counts as they do not represent additional instances of usage. Circulation count and last-use dates had been carried forward from VTLS to Innovative, so this information was available from 1985.

Scopus was used as the source of citations rather than Web of Science or Google Scholar. The ability to carry out a search of the references attached to Scopus records for character strings and proximate terms means that it can be reliably searched for citations of any document with a unique title and first author, regardless of document type and with no intervening protocols; these searches can then be scripted into a spreadsheet of titles and authors and executed with a single click. Web of Science–cited reference searches cannot be automated, making its use impractical for any but small-scale studies and not scalable from research projects to actual implementations. Google Scholar could potentially be used for studies of this kind; however, as it lacks an Application Programming Interface (API), it has to be searched manually—with the result that repetitive searches tend to evoke its anti-bot protection. In any case, its use in this instance would have raised questions about the accuracy of citations and required further manual checking, whereas Scopus could be relied on to provide a valid sample of citing literature. It is acknowledged that, in using a predominantly article-based citation database, the study focuses largely on citations of books in journal articles, although it is noted that Scopus claimed coverage of the citations of more than 75,000 books at the date of the study.[18] The important point, however, is that the purpose of the present study is not to identify *all* of the citations of the books in question, but rather to take a large sample to compare citations of books within distinct subject areas. In 2011, Kousha, Thelwal, and Rezaie[19] found that both Google Scholar and Google Books respectively returned 3.2 and 1.4 times as many citations to books as Scopus, but their analysis was carried out for the purposes of research assessment in which case finding as many citations as possible was essential, whereas the purpose of the present study is to identify the quality of comparative citedness in relation to individual titles rather than to quantify all citations. Their analysis indicated a significant degree of correlation between comparative rates of citedness between all three sources and would suggest that Scopus was a sufficiently sensitive instrument to be used in the present study.

The books fell into two categories, authored and edited (including conferences), and produced two distinct types of citation; authored books are cited by author(s) and title while, in edited books, it is generally a chapter that is being cited—although the full title of the book is also part of the citation. Although editor names are usually included in chapter citations, these proved not to be searchable. After some experimentation, a script was arrived at: the titles of authored books were string-searched within sixty characters of the name (without initials) of the first author, the intervening characters allowing for the names of additional authors, while for edited books the title only was searched for. Subtitles presented another issue, as many books have subtitles that may or may not be included in citations; for example, Stephen Jay Gould's *Bully for brontosaurus: Further reflections in natural history* received only 25 citations under its full title, whereas a further 233 documents had simply cited it as *Bully for brontosaurus*. The initial search was carried out for both title and subtitle as a single string and if necessary a "switch" was activated to exclude the subtitle. Once the search had been run, the Scopus "secondary documents" were examined to ensure that the search had correctly identified the book in question. In some cases, it was difficult to isolate citations of a unique title through this method; for example, the search REF ("animal

behaviour" w/60 scott) found references to both John Paul Scott's *Animal behavior* and Graeme Scott's *Essential animal behavior.* In the animal sciences, many edited books have very generic titles like *Comparative socioecology*, and it proved impossible to differentiate their citations from references to journal articles containing the same phrase. In the 324 range, it was found that many book titles ended with date ranges that citing documents tended to omit: the book *Tory and Whig the struggle in the constituencies, 1701–1715* was cited only 7 times under its full title but 12 times as *Tory and Whig the struggle in the constituencies*. When there was any ambiguity, titles were excluded from the study, which amounted to nearly 11 percent of titles in the 591 range and 4 percent in the 324 range; it is to be expected that this figure will vary according to discipline norms. Finally, Scopus citation searching does not readily distinguish between editions; but this was not considered to be a problem in this context, as citations to different editions of a book are all citations of that specific book and therefore valid indicators of the extent of its impact.

In addition to the initial search for all citations of a title, two further searches were run: the first for citations in documents published 2010 or later and the second for citations in documents with Massey University affiliations. Because it is not possible to isolate secondary documents from Scopus AND searches, the decision to include or exclude titles was made before these searches were run. Once this process was completed, the spreadsheets contained the following data for each title: publication date, total circulation count, date of last check-in, total number citations, number of citations from 2010 on, and number of citations in Massey-affiliated documents. From this, two further data elements could be automatically derived: the number of months since the last check-in date and the number of citations from pre-2010 publications. All data was captured in July 2015.

The following data were compiled:

TABLE 1		
Summary of Variables		
Key Variable	591 Zoology (n = 864) Mean/Standard deviation	324 Politics (n = 403) Mean/Standard deviation
Circulation Count	10.6/11.5 Range 0–78	7.3/16.2 Range 0–249
Months Since Check–in	123.2/85.2 Range 0–366	133.9/101.9 Range 0–366
Total Citations	167.4/425.2 Range 0–6076	48.7/121.5 Range 0–1374
Citations Since 2010	41.7/134.5 Range 0–2186	19.1/51.5 Range 0–686
Pre-2010 Citations	125.8/297.3 Range 0–4041	29.5/72.8 Range 0–688
Massey Citations	0.3/1.3 Range 0–14	0.04/0.2 Range 0–2

Results

As table 1 shows, both circulation and citation data are highly skewed. While some books had never been borrowed and many had only a few circulations, one title in the 324 range had been borrowed 249 times. This was even more pronounced in the case of citations, with many books having few or no citations while a small number had

been cited thousands of times, with the result that the standard deviation for the 591 range was more than double the mean. This confirmed what was already suspected, however: that both circulation and citation do not occur in a regular fashion but are characterized by the presence of "celebrity titles" that fall well outside discipline norms. It is the purpose of the study to determine the extent to which these two groups (highly circulated and highly cited) overlap, and any deselection exercise based on this methodology would set out to identify these titles.

The data in the Excel spreadsheet were analyzed using Spearman's rank correlation coefficient[20] to determine the relationships between factors that might be used to inform judgments about individual titles being considered for deselection or retention. A correlation looks at two sets of numbers to test the degree to which an increase in one set can be used to "predict" an increase or decrease in the other. For example, if one set of numbers listed the distances traveled by cyclists and the other set listed the times taken, then a correlation value of 1 would mean that all cyclists rode at the same speed and the time required for a journey could be accurately predicted by its distance. A correlation of 0 means that there is no predictive relationship at all between the two sets of numbers, while a negative correlation exists when an increase in one number can be used to predict a decline in the other. Spearman's rank correlation looks at the relationship between two sets of ranked numbers according to their rankings rather than simply correlating the numbers themselves and is designed for situations where extreme "outliers" could skew the results.

Research Question	Spearman Correlations	591 Zoology	324 Politics
	TABLE 2 **Summary of Results**		
1	Circulation count/Months since check-in	–0.52	–0.42
2	Pre-2010 citations/2010 citations	0.88	0.91
3	Massey citations/Total citations	0.42	0.22
4	Total citations/Circulation count	0.32	0.31
4	Total citations/Months since check-in	–0.28	–0.29

Research Question 1: Are historic monograph circulation counts a robust predictor of future circulation counts?

The use of circulation counts as a primary source of data for deselection is premised on the notion that, although circulation of a given book may decline over time, this decline will be reasonably similar over all titles (at least within the same subject group) and is largely a factor of the age of the book. Most library systems do not provide a full borrowing history for each book; instead, they simply record the total number of borrowings (circulation count) and the date on which the book most recently moved through the system (last check-in). Because of this limitation, we cannot determine how often a book was borrowed two decades ago and compare this to how often it has been borrowed in the most recent decade to test whether the earlier figure had predicted the latter. However, the last check-in date can be used as a proxy for recent circulation on the basis that recently circulating books would have been checked in closer to the present than those that were not in use, in which case a negative correlation between circulation and the number of months since the last check-in would indicate that past circulation counts predicted future ones. This correlation did exist for both the samples—0.52 for the 591 sample and –0.42 for the 324 sample—but the partial nature

of the correlation means that historic circulation is an imperfect predictor of the future probability that a book will be borrowed. At the population level, it is safe to say that highly circulated books will continue to be borrowed more often than average, but at the individual title level this becomes a probability rather than a robust indicator. If a simple cutoff number of circulations had been imposed on this set of books ten years ago, a significant number of books would have been removed from the shelves that had in fact subsequently been circulated, while others that had been kept would have received no further circulation activity. This is only to be expected, of course, but it does cast doubt on the assertion that circulation figures on their own can be reliably extrapolated into the future.

Research Question 2: Are historic monograph citation counts a robust indicator of future citation counts?
It is well known that a publication will be cited more often in its first few years and that this figure will decline over time, although rates of decline (cited half-lives) will vary between disciplines and may also operate differently for monographs in comparison to journal articles.[21] However, if citation is an indication of "interest" in a book that was not cited frequently in the past, it might be expected that it is less likely to attract attention in the present or future than more highly cited titles. To test this hypothesis, pre-2010 citations for each title were calculated by subtracting the post-2009 citations from the totals and then the two sets of figures were correlated to determine if historic citations were reliable predictors of future citations. The correlation was 0.88 for the 591 sample and 0.91 for the 324 sample, indicating that past citedness could be extrapolated into the future for these disciplines.

Research Question 3: Do local-institution monograph citation counts mirror total citation counts?
Total citations by Massey-affiliated authors were obviously very low in comparison to all citations; here, the correlation was relatively much weaker: 0.42 for the 591 sample and 0.22 for the 324 sample. In terms of deselection, however, it would be hard to argue that the interests of the small number of researchers over a number of years who happened to have cited books from the collection accurately reflected the interests of researchers as a whole. However, in larger institutions with much greater numbers of published articles, local citation counts could potentially be significant.

Research Question 4: Can monograph circulation counts be used as a proxy for citation counts?
The final research question related to the correlation between circulation and citation data. If there was a strong relationship between the two phenomena, they could be regarded as mutually validating and either could act as a reasonable proxy of the other. This turned out not to be the case. The correlation between citations and checkouts was 0.32 for the 591 sample and 0.31 for the 324 sample, while the correlation between citations and months since last check-in was –0.28 for the 591 sample and –0.29 for the 324 sample. As a group, then, well-cited books are borrowed more than others, but at individual title level the effect is too random for either figure to predict the other in a reliable way.

Discussion
To the advocate of circulation data–driven deselection, it might appear that the low level of correlation between citation and circulation counts means that lack of interest in citation bibliometrics as a data source for the purpose of deselection is fully justified; if

the citation count of a book cannot reliably predict the future likelihood of a borrower checking it out, then the fact that it has been of interest to authors of scholarly works might seem to be merely academic. However, the core assumption behind this line of argument is that it is the purpose of library books to circulate and that circulation data are an accurate reflection and primary determinant of the value of a book to a library; however, if we consider that the purpose of a library book is to provide information, we cannot measure their success in doing this solely on the basis of how often they are circulated. Obviously, library books are purchased in the expectation that library users will find them useful, but even at the time of purchase many titles in academic libraries are known to be of interest to only a small subsection of all library users and have been added to the collection for the purpose of providing valuable coverage of specific topics. To come back at some future date and measure them on the basis of circulation data is to subject them to a test that they were never intended to pass.

Any metric must relate to an identifiable meaning external to itself in order to be used as a valid basis for action. Speed can be used to determine the time a journey of a particular length will take because that meaning is coded into the notion of speed, whereas from the point of view of information provision the relationship between circulation and information value is tenuous and carries only a small amount of meaning. Adams and Noel[22] pointed out a number of serious problems with circulation counts as a meaningful data source, including varying loan rules for different patron types and the fact that renewals are counted equally with initial checkouts, while Danielson[23] emphasized that circulation figures are highly localized and contingent and cannot be compared between libraries. Most notably, Evans and Saponaro[24] stated that:

> There are problems in interpreting circulation data in terms of the value of a collection. Circulation data cannot reflect use generated within the library, such as reference collections and non-circulating journals. Even for circulating items, there is no way of knowing how the material was used; perhaps the book was used to prop open a window or press flowers. Nor can you determine the value derived by the person from a circulated item.

If a circulation count is to be taken as a significant measure of the value of a book, it would be expected that all books had an equal opportunity to circulate; but there are many other factors that can influence this. The influence of shelf position on circulation was first noted by Shaw;[25] and, while it might have a lesser effect in academic libraries if a significant proportion of borrowing is of items located through catalog searches, undoubtedly eye-level shelves are considerably easier to browse than those at floor level. A more recent study of browsing behaviour by McKay, Smith, and Chang[26] found that there was a distinct "neighbor effect" on borrowing and that "a nearby book being loaned increases the probability of a loan by at least 9-fold." Discoverability through keywords, subject headings, and titles is another variable, while classification itself has a critical effect on the likelihood that a book will be discovered by all members of its target audience. In the words of Atkinson,[27] library catalogs are "powerful and subtle censorship devices that exclude all but a few of the subjects considered in the document."

More fundamental in terms of the meaning of circulation data in relation to the information value of books is the question of timing. A circulation count is generated at the point at which the book is checked out, and, although there is a tendency to use the words *circulation* and *usage* interchangeably, they are not at all the same thing. To take a book off the shelves and borrow it might presume an expectation of use; but, because the barriers to borrowing are generally low, this expectation does not need to

be particularly high. A student may spend a few minutes at the shelves, fail to find a really useful title, and instead choose ten books that look vaguely promising, then renew them twice, and in the process generate thirty points of circulation data. Whether the books were in fact used, or even looked at, is simply not a piece of information that is conveyed or implied by the checkout. While it is true that books that circulate are as a class more useful than those that do not, when it comes to individual titles all that can be said is that they are more attractive to borrowers in some way, which could be as much a factor of title, appearance, classification, shelf position, or appearance on a reading list as of inherent usefulness. This last factor is particularly important, as book circulation to undergraduate students is heavily influenced by coursework and the recommendations of instructors. Short lists of recommended readings and large classes can result in libraries purchasing, and lending, multiple copies of individual titles that later cease to circulate as the instructor leaves or the emphasis of the course changes.

Turning to the potential of citations as a measure of value, similar reservations exist but they are less acute. Cronin[28] drew attention to the difficulty of assigning firm meaning to the act of citation in the absence of standardized practice, the impossibility of understanding the intentions of citing authors, and the presence of "noise" in the signals given off by citation counts, but the case for intentionality can be made on the basis of the barriers that exist to a work being cited and that citation being registered. Although there is no absolute proof that the citing author has engaged intellectually with the cited work, the citation carries with it a claim that he or she has and is at least an endorsement of the relevance of that work to the topic in question. The citation does not carry with it an automatic agreement with or endorsement of the cited work, but it does imply that it is worth taking seriously and that the reader can find further information on the topic in the cited work. The next barrier is the fact that the citing author needs to be a scholar of some ability and must write his or her own work and persuade an editor and reviewers that it is worthy of publication; ideally, the peer-review process itself will act as a check on the references to ensure that they are not profligate or gratuitous, although it is acknowledged that peer review can lead to "reference packing" in some cases. Finally, the journal or book in which the citing work appears must be one that is included in the relevant citation database, generally Web of Science or Scopus. What these barriers mean is that the act of citation is not open-ended and that, despite the presence of noise and the existence of anomalies, there is a recognized relationship between citation and significance that means that a well-cited book is probably one that has made a contribution to its field and is recognized by scholars as worthy of attention.

As well as being indicators of scholarly merit, citations serve another function, as pointers to the literature. Researchers do not discover documents purely by catalog or database searching or from browsing; citation chaining[29] or the following up of references is an important means of carrying out a comprehensive search of the literature and is explicitly facilitated in citation databases and through the use of Digital Object Identifiers attached to references. Although scholarly literature, particularly in the sciences, is often seen as having a short shelf life—Hurt remarked that "five years old is too old"[30]—in fact, all scholarship is based on past work; in the case of scholarly monographs, many of them attain the status of classics, a knowledge of which is critical to the understanding of current thinking. The present study found that MacArthur and Wilson's *Theory of island biogeography* was cited 2,186 times between 2010 and July 2015; even if many of these citations are tokenistic, they point to the indispensable nature of this title within the whole field of zoology. The linear metaphor for scholarly progress is simplistic, and use of "historic" monograph collections is not confined to historians of science—it is just as important for a full understanding of current debates and developments.

It is not the intention of this paper to advocate the use of citation data on its own as a determining factor in deselection decision-making, but rather to endorse Burdick's view that it is an essential safety net for librarians making "life and death" decisions about books of which they have no first-hand knowledge. The notion that decisions of this sort can be data-driven ignores the fact that what information does, in fact, drive is a model, and the strength of the model is based on its relationship to real-world meaning. Neither circulation nor citation data can stand as full proxies of the value of a title, but, in reflecting the status of a title within the scholarly community, citedness should be considered at least equally with circulation. Neither metric is "wrong," as they measure different phenomena; but the weak correlation between them means that an optimal decision requires both pieces of information. A circulation purist might argue that, if a book is rarely or never borrowed, its potential value is purely theoretical; but, for a research library, its coverage of the literature is a value in itself. What Kelly[31] has called the "materials-centered approach" presupposes that the needs of a single researcher are not to be weighed against the borrowing patterns of the group. Because circulation data are easy for librarians to procure, their use in deselection is straightforward and convenient; and, because libraries report on circulation statistics rather than on units of actual information, they appear to provide a good fit with both library missions and common-sense. However, real-life library use is multidimensional, and considered decisions need to be informed by an understanding of the book in its full context. By removing those titles that had not made a significant contribution to their disciplines and retaining those that had, a well-conducted deselection exercise would leave a core research collection of high value titles that had earned their place on library shelves and that continued to inform the research they had been purchased to support.

Reflection on Methodology and Future Research

As noted above, Scopus was chosen as the sampling medium for this study for essentially pragmatic reasons. A statistical analysis needs to be grounded in a clearly identifiable sample and the data gathered using straightforward and replicable methods; Scopus proved to be the most convenient instrument for these purposes. Looking at citations of monographs over a range of disciplines, Kousha, Thelwall, and Rezaie[32] found a significant degree of correlation between Google Scholar and Scopus (Spearman correlations 0.74 to 0.94), between Google Books and Scopus (0.61 to 0.83), and between Google Scholar and Google Books (0.74 to 0.83), which suggests that they are all measuring more or less the same phenomenon and that the likelihood that a book would be highly cited in one source but not in another is negligible. (This is in large part due to the skewed distribution of citations that those authors also commented on.) A 2015 paper by Kousha and Thelwall[33] titled "An Automatic Method for Extracting Citations from Google Books" describes the use of an API to harvest citation data and raises the exciting prospect that title-level citation metrics would become a standard tool for librarians working on deselection. Preliminary tests indicate that good results could be achieved using this method but that short titles would present more of a challenge to accuracy, and a higher rate of rejection of titles prior to implementation might be required to produce realistic results.

Acknowledgements

Notes

1. Terrance L. Cottrell, "Weeding Worries, Part 1: Books," *The Bottom Line* 26, no. 3 (2013): 98–102, doi:10.1108/BL-06-2013-0015; Rick Lugg and Ruth Fischer, "Future Tense: Doing What's Obvious: Library Space and the Fat Smoker," *Against the Grain* 21, no. 1 (2009): 75–76, available online at http://docs.lib.purdue.edu/atg/vol21/iss1/47 [accessed 23 November 2016

2. Rick Anderson, "Print on the Margins: Circulation Trends in Major Research Libraries," *Library Journal* (2011), available online at http://lj.libraryjournal.com/2011/06/academic-libraries/print-on-the-margins-circulation-trends-in-major-research-libraries/ [accessed 23 November 2016].

3. Patricia L. Thibodeau, "When the Library Is Located in Prime Real Estate: A Case Study on the Loss of Space from the Duke University Medical Center Library and Archives," *Journal of the Medical Library Association: JMLA* 98, no. 1 (2010): 25–28, doi:10.3163/1536-5050.98.1.010.

4. Karen Jensen, "Data-Driven Decisions for Library Liaisons: Exploring Strategies for Effectively Managing Diminishing Monograph Collections," *Collection Management* 37, no. 1 (2012): 9–22, doi:10.1080/01462679.2012.629854; John N. Ochola, "Use of Circulation Statistics and Interlibrary Loan Data in Collection Management," *Collection Management* 27, no. 1 (2003): 1–13, doi:10.1300/J105v27n01_01; Lisa M. Rose-Wiles, "Are Print Books Dead? An Investigation of Book Circulation at a Mid-Sized Academic Library," *Technical Services Quarterly* 30, no. 2 (2013): 129–52, doi:10.1080/07317131.2013.759496.

5. J. Corey Miller, Keith H. Coble, and Jayson L. Lusk, "Evaluating Top Faculty Researchers and the Incentives That Motivate Them," *Scientometrics* 97, no. 3 (2013): 519–33, doi:10.1007/s11192-013-0987-7.

6. Vincent Larivière, "On the Shoulders of Students? The Contribution of PhD Students to the Advancement of Knowledge," *Scientometrics* 90, no. 2 (2011): 463–81, doi:10.1007/s11192-011-0495-6.

7. Richard W. Trueswell, "A Quantitative Measure of User Circulation Requirements and Its Possible Effect on Stack Thinning and Multiple Copy Determination," *American Documentation* 16, no. 1 (1965): 20–25, doi:10.1002/asi.5090160106.

8. Scott Britton and John Renaud, "SPEC Kit 337: Print Retention Decision Making" (2013), available online at http://publications.arl.org/Print-Retention-Decision-Making-SPEC-Kit-337/ [accessed 23 November 2016]; G. Edward Evans and Margaret Zarnosky Saponaro, *Collection Management Basics* (Santa Barbara, Calif.: Libraries Unlimited, 2012); 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; Maggie Liu, "Collection Analysis on Social Work: A Cost-Effective Approach," *Collection Building* 34, no. 2 (2015): 59–64, doi:10.1108/CB-06-2014-0031; Rick Lugg, "Data-Driven Deselection for Monographs: A Rules-Based Approach to Weeding, Storage, and Shared Print Decisions," *Insights: The UKSG Journal* 25, no. 2 (2012): 198–204, doi:10.1629/2048-7754.25.2.198; Cynthia Ehret Snyder, "Data-Driven Deselection: Multiple Point Data Using a Decision Support Tool in an Academic Library," *Collection Management* 39, no. 1 (2014): 17–31, doi:10.1080/01462679.2013.866607; Simona Tabacaru and Carmelita Pickett, "Damned If You Do, Damned If You Don't: Texas A&M University Libraries' Collection Assessment for Off-Site Storage," *Collection Building* 32, no. 3 (2013): 111–15, doi:10.1108/CB-02-2013-0006; Doug Way, "Developing and Implementing a Disapproval Plan," *College & Research Libraries News* 74, no. 6 (2013): 284–87; Patricia Bravender and Valeria Long, "Weeding an Outdated Collection in an Automated Retrieval System," *Collection Management* 36, no. 4 (2011): 237–45, doi:10.1080/01462679.2011.605290; Stanley J. Slote, *Weeding Library Collections: Library Weeding Methods*, 4th ed. (Englewood, Colo.: Libraries Unlimited, 1997); Suzanne Ward, "The Other Side of the Coin : De-Selecting Material from a Research Library's Storage Facility," in *Proceedings of the Charleston Library Conference* (2010), doi:10.5703/1288284314483; Karen C. Kohn, "Usage-Based Collection Evaluation with a Curricular Focus," *College & Research Libraries* 74, no. 1 (2013): 85–97.

9. Stephen Bulick, William N. Sabor, and Roger Flynn, "Circulation and in-House Use of Books," in *Use of Library Materials: The University of Pittsburgh Study*, eds. Allen Kent et al. (New York: Marcel Dekker, 1979), 49.

10. Debbi Dinkins, "Circulation as Assessment: Collection Development Policies Evaluated in Terms of Circulation at a Small Academic Library," *College & Research Libraries* 64, no.1 (2003): 46–53.

11. Michael W. Handis, "Practical Advice for Weeding in Small Academic Libraries," *Collection Building* 26, no. 3 (2007): 84–87, doi:10.1108/01604950710761643; Rick Lugg and Ruth Fischer, "Future Tense—Weeding: The Time Is Now," *Against the Grain* 20, no. 4 (2008): 87–88, available online at http://docs.lib.purdue.edu/atg/vol20/iss4/35 [accessed 23 November 2016]; Jim Martin, Hitoshi Kamada, and Mary Feeney, "A Systematic Plan for Managing Physical Collections at the University of Arizona Libraries," *Collection Management* 38, no. 3 (2013): 226–42, doi:10.1080/01462679.2013.797376.

12. Peggy Johnson, *Fundamentals of Collection Development and Management* (Chicago: American Library Association, 2013), 322–23.

13. Britton and Renaud, "SPEC Kit 337: Print Retention Decision Making."

14. Lea Currie and Amalia Monroe-Gulick, "What Do Our Faculty Use? An Interdisciplinary Citation Analysis Study," *Journal of Academic Librarianship* 39, no. 6 (2013): 471–80, doi:10.1016/j.acalib.2013.08.016; Irene Ke and Jackie Bronicki, "Using Scopus to Study Researchers' Citing Behavior for Local Collection Decisions: A Focus on Psychology," *Journal of Library Administration* 55, no. 3 (2015): 165–78, doi:10.1080/01930826.2015.1034035; Charlene Kellsey and Jennifer Knievel, "Overlap between Humanities Faculty Citation and Library Monograph Collections, 2004–2009," *College & Research Libraries* 73, no. 6 (2012): 569–83, doi:10.5860/crl-280; Karen C. Kohn and Larissa Gordon, "Citation Analysis as a Tool for Collection Development and Instruction," *Collection Management* 39, no. 4 (2014): 275–96, doi:10.1080/01462679.2014.935904; Cory Tucker, "Analyzing Faculty Citations for Effective Collection Management Decisions," *Library Collections, Acquisition and Technical Services* 37, no. 1–2 (2013): 19–33, doi:10.1016/j.lcats.2013.06.001; Concepción S. Wilson and Carol Tenopir, "Local Citation Analysis, Publishing and Reading Patterns: Using Multiple Methods to Evaluate Faculty Use of an Academic Library's Research Collection," *Journal of the American Society for Information Science and Technology* 59, no. 9 (2008): 1393–1408, doi:10.1002/asi.20812.

15. Kristin Hoffmann and Lise Doucette, "A Review of Citation Analysis Methodologies for Collection Management," *College & Research Libraries* 73, no. 4 (2012): 321–35, doi:10.5860/crl-254.

16. K. Brock Enger, "Using Citation Analysis to Develop Core Book Collections in Academic Libraries," *Library and Information Science Research* 31, no. 2 (2009): 107–12, doi:10.1016/j.lisr.2008.12.003.

17. Amrita J. Burdick, "Science Citation Index Data as a Safety Net for Basic Science Books Considered for Weeding," *Library Resources & Technical Services* 33, no. 4 (1989): 367–73.

18. Susannah Beatty-Tucker and Susanne Steiginga, "Scopus Content Update: 75,000 Book Titles and Counting | Elsevier Scopus Blog" (2015), available online at http://blog.scopus.com/posts/scopus-content-update-75000-book-titles-and-counting [accessed 23 November 2016].

19. Kayvan Kousha, Mike Thelwall, and Somayeh Rezaie, "Assessing the Citation Impact of Books: The Role of Google Books, Google Scholar, and Scopus," *Journal of the American Society for Information Science and Technology* 62, no. 11 (2011): 2147–64, doi:10.1002/asi.21608.

20. W. Pirie, "Spearman Rank Correlation Coefficient," in *Encyclopedia of Statistical Sciences* (Hoboken, N.J.: John Wiley & Sons, 2004), doi:10.1002/0471667196.ess2499.pub2.

21. Rong Tang, "Citation Characteristics and Intellectual Acceptance of Scholarly Monographs," *College & Research Libraries* 69, no. 4 (2008): 356–69, doi:10.5860/crl.69.4.356.

22. Brian Adams and Bob Noel, "Circulation Statistics in the Evaluation of Collection Development," *Collection Building* 27, no. 2 (2008): 71–73, doi:10.1108/01604950810870227.

23. Robert Danielson, "A Dual Approach to Assessing Collection Development and Acquisitions for Academic Libraries," *Library Collections, Acquisition & Technical Services* 36, no. 3/4 (2012): 84–96, doi:10.1016/j.lcats.2012.09.002.

24. Evans and Saponaro, *Collection Management Basics*, 144.

25. Ralph R. Shaw, "The Influence of Sloping Shelves on Book Circulation," *Library Quarterly: Information, Community, Policy* 8, no. 4 (1938): 480–90.

26. Dana McKay, Wally Smith, and Shanton Chang, "Lend Me Some Sugar: Borrowing Rates of Neighbouring Books as Evidence for Browsing," in *IEEE/ACM Joint Conference on Digital Libraries* (IEEE, 2014), 145–54, doi:10.1109/JCDL.2014.6970161.

27. Ross Atkinson, "The Citation as Intertext: Toward a Theory of the Selection Process," *Library Resources & Technical Services* 28, no. 2 (1983): 109–19.

28. Blaise Cronin, "The Need for a Theory of Citing," *Journal of Documentation* 37, no. 1 (1981): 16–24, doi:10.1108/eb026703.

29. Carole George et al., "Scholarly Use of Information: Graduate Students' Information Seeking Behaviour," *Information Research* 11, no. 4 (2006), available online at www.informationr.net/ir/11-4/paper272.html [accessed 23 November 2016].

30. C.D. Hurt, *Information Sources in Science and Technology*, 3rd ed. (Englewood, Colo.: Libraries Unlimited, 1998).

31. Matthew Kelly, "The Materials-Centred Approach to Public Library Collection Development: A Defense," *Library Philosophy and Practice (e-Journal)*, no. 1232 (2015), available online at http://digitalcommons.unl.edu/libphilprac/1232 [accessed 23 November 2016].

32. Kousha, Thelwall, and Rezaie, "Assessing the Citation Impact of Books."

33. Kayvan Kousha and Mike Thelwall, "An Automatic Method for Extracting Citations from Google Books," *Journal of the Association for Information Science and Technology* 66, no. 2 (2015): 309–20, doi:10.1002/asi.23170.