

Using Cited Half-life to Adjust Download Statistics

J. Parker Ladwig and Andrew J. Sommese

“Supplying accurate CPU [cost-per-serial use] information to faculty and appropriate marketing of the alternate modes of delivery ... become the key to achieving an optimal cost-efficient serials collection in an academic library.”¹

A model is presented for adjusting use statistics using a journal's ISI Journal Citation Reports cited half-life. The goal is to improve the method used to evaluate the raw electronic download figure. The proposed model will still undercount total use, but the undercounting will be proportional across disciplines and less severe. By using this model, librarians can avoid making cancellation decisions that may cost their libraries more money in the long run.



In the spring of 2004, the University Libraries of Notre Dame began another round of journal cancellations. One overall goal was to be as cost-efficient as possible. First, the subscription cost of a journal was divided by the number of full-text downloads for one year to calculate a cost per download. Then, this figure was compared to the average commercial document delivery (docdel) cost. Those journals that cost more per download than the estimated docdel cost became candidates for cancellation.

When the methodology was explained to the Mathematics Department's library committee, questions were raised not only about the cancellations, but also about the methodology employed. One obvious question was,

“How could only one year's worth of download statistics be a fair measure of use?” Upon reflection, the second author of this article uncovered an even more serious flaw: the downloads had not been adjusted for the journals' half-lives. Because the electronic runs are short (e.g., six years for many Springer Verlag journals as of 2003), raw download numbers would be reasonable for short half-life journals but would significantly undercount the downloads for long half-life journals.

To demonstrate the significance of the ISI Journal Citation Reports cited half-life, this article will discuss the importance of journal use statistics, explain cited half-life and its importance, and then present a model for adjusting download statistics (including the model's assumptions and problems).

J. Parker Ladwig is the Mathematics Librarian in University Libraries at the University of Notre Dame; e-mail: ladwig.1@nd.edu. Andrew J. Sommese is the Duncan Professor of Mathematics in the Department of Mathematics at the University of Notre Dame; e-mail: sommese@nd.edu. The authors would like to express their thanks to the referees of this article for their helpful suggestions.

Importance of Use Statistics

A main reason for the library to collect use statistics is so that it can maximize the return on its investment (ROI). For example, when someone buys a car, he or she expects to maximize his or her investment. If the \$20,000 car is expected to last ten years, this is equivalent to expecting to get at least \$2,000 worth of use from it each year (ignoring the effects of inflation). In the same way, if a library pays \$500 for a subscription to one volume of a journal (assuming that only one volume is published for the year), the library expects that it will get at least \$500 worth of use over the volume's lifetime.

Journal use statistics are collected in order to perform this sort of calculation. Because it is difficult to separate the use of one volume of a journal over its lifetime (assuming one volume is published per year), a library examines the use of the entire run, calculates the cost per use (CPU), and asks if the CPU is greater than the expected ROI. One measure of expected ROI is the cost of alternate modes of access, namely, interlibrary loan (ILL) and docdel. Because docdel is clearly more expensive than ILL, only the cost of docdel is compared to the journal's CPU. If the CPU (i.e., the cost of using the entire run of the journal over a year) is more than the cost of one article via docdel, there is a strong argument for canceling the subscription and investing any net savings in docdel or ILL.

This sort of analysis does not work in making decisions to subscribe to a journal. Even if more is spent on docdel per year for a particular journal than the annual subscription, there is not sufficient information to make a subscription decision (i.e., the publication years of the requested articles are generally unknown). Even if most of the articles are from the last few years and their docdel cost is significantly higher than a subscription, the cost for the first year or two will include both the subscription cost and the cost for docdel (or ILL) for articles from the years the library does not own. Thus, the model presented

in this paper is useful for cancellation decisions but would need to be modified for decisions about new subscriptions.

Definition of Cited Half-life

ISI defines and explains cited half-life as follows:

The cited half-life is the number of publication years from the current year which account for 50% of current citations received. This figure helps you evaluate the age of the majority of cited articles published in a journal. Each journal's cited half-life is shown in the Journal Rankings Window. Only those journals cited 100 or more times have a cited half-life.

The chronological distribution of the cumulative percent of citations received per publication year is shown in the Cited Half-Life Calculation dialog box.

A higher or lower cited half-life does not imply any particular value for a journal. For instance, a primary research journal might have a longer cited half-life than a journal that provides rapid communication of current information. Cited Half-Life figures may be useful to assist in collection management and archiving decisions. Dramatic changes in Cited Half-Lives [sic] over time may indicate a change in a journal's format. Studying the half-life data of the journals in a comparative study may indicate differences in format and publication history.²

To illustrate cited half-life, consider the ISI citation figures for *Nature Cell Biology*, *Communications in Partial Differential Equations*, and *Mathematische Annalen* listed in table 1. Data from 2003 are the most recent available.

In 2003, *Nature Cell Biology* had an ISI cited half-life of 2.7 years. Specifically, of

TABLE 1
Percent of Citations to Each Journal
 “Breakdown of the citations to the journal by the cumulative percent of 2003
 cites to articles published in the following years” (JCR)

	2003	2002	2001	2000	1999	1998	1997	1996	1995	1994	1993- all
Nat. Cell. Biol.	5.1	28.8	57.6	84.4	99.9	99.9	99.9	99.9	99.9	99.9	100.0
Comm. PDE	0.5	3.6	10.7	18.0	24.3	29.5	35.9	42.2	47.5	53.1	100.0
Math. Ann.	0.5	2.5	6.4	9.3	11.8	15.1	17.5	19.6	21.7	24.2	100.0

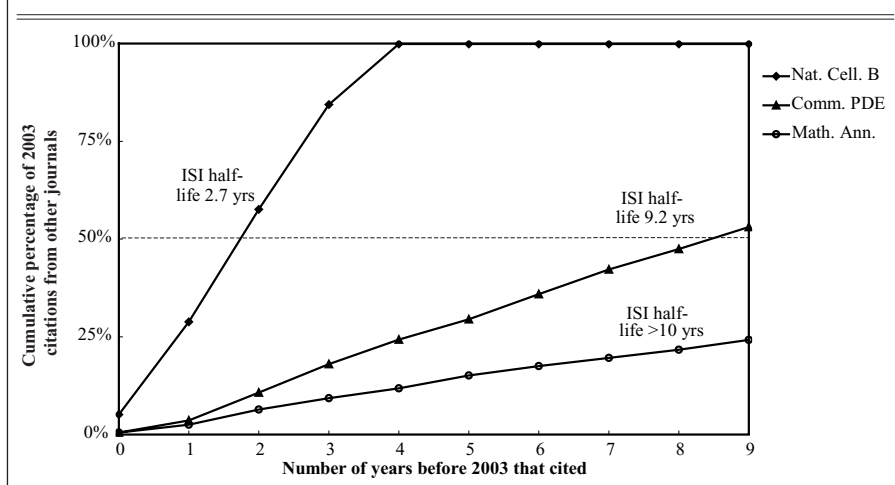
the total number of citations referring to *Nat. Cell. Biol.* from all the journals tracked by ISI, 5.08 percent were to articles published in 2003; 28.84 percent to articles published in 2002 or 2003; 57.62 percent to articles published in 2001, 2002, or 2003; and 84.39 percent to articles published in 2000, 2001, 2002, or 2003. Thus, with just the four most recent years of *Nat. Cell. Biol.* readily available, a researcher would be able retrieve more than 80 percent of the current citations to it.

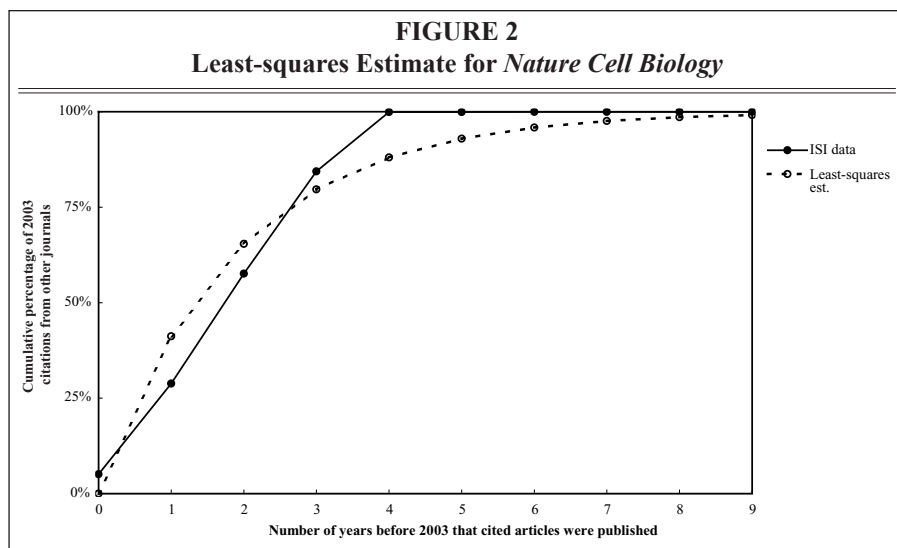
Communications in Partial Differential Equations had a cited half-life of 9.5 years. Of the total number of citations referring to *Comm. PDE*, 0.52 percent were to articles published in 2003; 3.60 percent to articles published in 2002 or 2003; 10.72 percent to articles published in 2001, 2002,

or 2003; and 17.99 percent to articles published in 2000, 2001, 2002, or 2003. Thus, with just the four most recent years of *Comm. PDE* available, a researcher would be able to retrieve less than 20 percent of the citations to it.

For *Mathematische Annalen*, a leading pure mathematics journal, the situation is even more dramatic. The cited half-life is listed as > 10 (if the ISI cited half-life is calculated as described below, it would be approximately 23 years). Of the total number of citations referring to *Math. Ann.*, 0.51 percent were to articles published in 2003; 2.53 percent to articles published in 2002 or 2003; 6.39 percent to articles published in 2001, 2002, or 2003; and 9.31 percent to articles published in 2000, 2001, 2002, or 2003. Thus, with just

FIGURE 1
ISI Half-lives for Three Journals





the four most recent years of *Math. Ann.* available, a researcher would be able to retrieve less than 10 percent of the citations to it.

Please note, however, one oddity in ISI's method of determining half-life. ISI considers the publication year (e.g., 2003) as "year one" of half-life rather than "year zero." Thus, when the half-life is 3.0 years, ISI does not mean that half of the articles were cited after 2000 (= 2003 - 3) but, rather, after 2001 (2003 was the first year, 2002 the second, 2001 the third). For our calculations, the beginning of the publication year must start at zero; therefore, the ISI half-life is simply adjusted by subtracting one year. (See figure 1.)

Cited half-life as an Exponential Decay Curve

As the term *half-life* suggests, the fraction of citations from a given year satisfies to a first approximation a curve for exponential decay. One can illustrate this by plotting the ISI data against a least-squares fitted exponential decay curve (discussed below). (See figure 2 for an example.)

Many physical models serve as analogies. For example, in the field of physics,

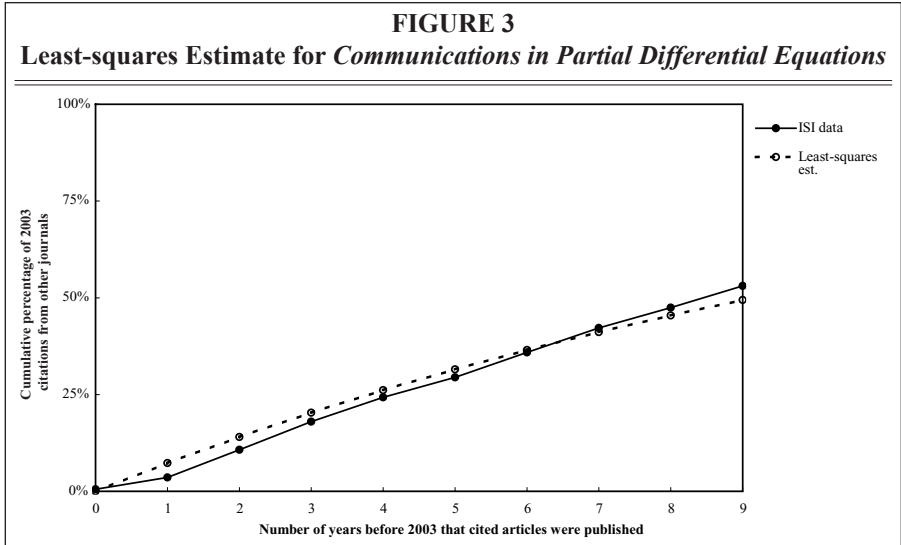
the half-life of a radioactive substance is the time required for half

of it to decay away.... A plot of the remaining nuclei as a function of time shows a steady decrease as the curve tends to, but never actually reaches, zero. The kind of behavior is called exponential decay.... The fraction of the original material remaining after n generations is $(1/2)^n$, instead of 2^n for exponential growth.³

The fractional amount decaying each year does not change as the source disappears. Say, for example, that the half-life of a chunk of radium is 1,620 years. After 1,620 years, one half of the chunk is left. After 3,240 years, the original chunk does not have 0 percent radium left but, rather, $100\% \cdot 1/2 \cdot 1/2 = 100\% \cdot 1/4 = 25\%$. After T years the fraction of the total atoms of radium left is

$$\left(\frac{1}{2}\right)^{\left(\frac{T}{1620}\right)}$$

The empirical data provided by ISI may be used to find the exponential decay curve with the best least-squares fit to the data. As illustrated in figures 2, 3, and 4 (with the half-life corrected as indicated above), the least-squares fitted decay curves fit the data surprisingly well, especially for long half-life journals.

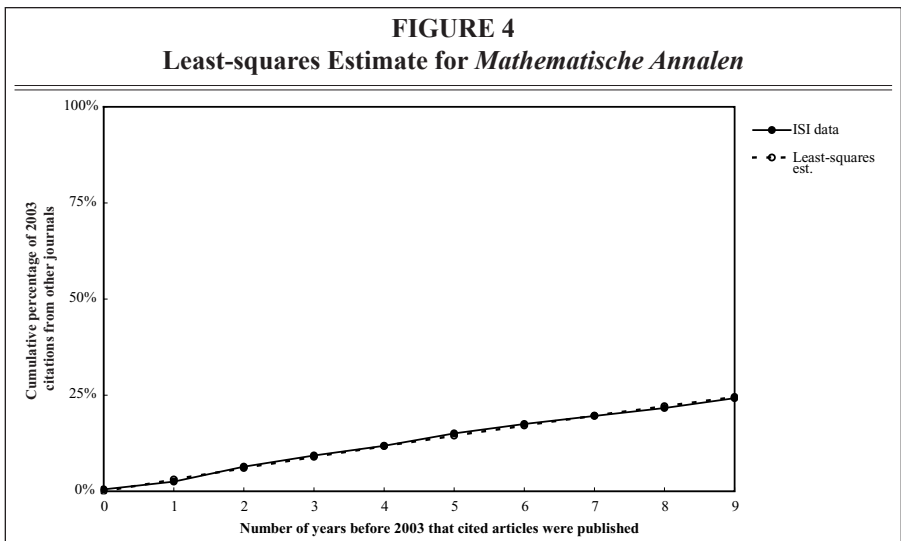


Importance of Cited Half-life

Among its many advantages, the half-life is particularly important for providing more accurate use data for all types of journals. For example, assume that only the print version of a journal is available, that the library has a complete run, and that the library is gathering reasonably accurate annual use statistics. If two of the example journals were used in print one hundred times over the course of a year, it would be expected that fifty of those uses for *Nat. Cell. Biol.* were to articles

published in the past 1.7 years ($= 2.7 - 1$), and fifty of those uses for *Comm. PDE* were to articles published in the past 8.5 ($= 9.5 - 1$) years.

Imagine now a second scenario where the electronic version is available for the past seven years, and prior to that, only the print was available (back to volume one, issue one). Further, assume that the year's use data include *only electronic downloads*. For *Nat. Cell. Biol.*, because its corrected half-life is 1.7 years and the electronic version is available for seven



years, virtually 100 percent of the total use (print and electronic) is reflected in the download statistics (assuming that one electronic download is equal to one print use). For *Comm. PDE*, because its corrected half-life is 8.5 years, only 36 percent of the total use is reflected in the downloads; and for *Math. Ann.*, because of its very long half-life, only 18 percent is reflected in the downloads.

Not only is the total use undercounted in the second scenario (only electronic downloads were captured), but the undercounting is not comparable. For one journal, the electronic use captures nearly 100 percent of the total use and for another, less than 20 percent.

The striking difference in these figures also points to a difference across fields that, if not adjusted for, could lead to cancellations ultimately costing the library more than was initially saved. For example, of the 327 mathematics and applied mathematics journals listed in ISI, 39 percent have an ISI cited half-life of at least ten years, 70 percent have an ISI cited half-life of at least six years, and 11 percent have no half-life given (e.g., new journals). Compare this with biochemistry and molecular biology. Of the 261 journals listed in ISI, 7 percent have an ISI cited half-life of at least ten years, only 38 percent have an ISI cited half-life of at least six years, and only 1 percent have no half-life given.

Thus, to compare fairly the total use of one journal with another across disciplines, electronic download statistics should be adjusted by incorporating cited half-lives.

Model for Adjusting the Download Statistics

Let DL denote the number of electronic downloads in a given year. This is the "use statistics" that publishers provide.

Let ER denote the electronic run (i.e., the number of years of the journal from T years in the past to the present with electronic, but not paper available).

Let FC denote the full-count of use in a given year. The goal is the steady-state

full-count at some date in the future. To understand this, recall how this adjusted figure will be used. It will be divided into the yearly subscription cost to obtain the CPU. Because costs are involved in canceling a subscription and then resubscribing at a later date, the model is calculating the electronic downloads per year that one would expect in the future assuming the ER increases one more year each year.

To derive AF , the overall adjustment factor to be applied to DL , convert DL to FC as follows:

$$\boxed{FC = DL \cdot AF}$$

Letting HL denote the adjusted ISI cited half-life of the journal (i.e., = ISI HL - 1);

$$AF = \frac{1}{\left[1 - \left(\frac{1}{2}\right)^{\left(\frac{ER}{HL}\right)}\right]} \quad \text{(equation 1)}$$

The Mathematics

The mathematics is straightforward. As noted in the radium example above, the fraction of the total atoms of radium remaining after T years is:

$$\left(\frac{1}{2}\right)^{\left(\frac{T}{1620}\right)}$$

Thus, the total percentage spent after T years is:

$$\left[1 - \left(\frac{1}{2}\right)^{\left(\frac{T}{1620}\right)}\right] \cdot 100\%$$

By modeling citations to a given volume of a journal as atoms decaying from a chunk of radium for a single volume of a journal, the fraction of total of expected citations unaccounted for j years after publication is found to be:

$$\left(\frac{1}{2}\right)^{\left(\frac{j}{HL}\right)}$$

Thus, for a single volume of a journal, the total citations accounted for in the j th year after publication is:

$$\left[\left(\frac{1}{2} \right)^{\binom{j-1}{HL}} - \left(\frac{1}{2} \right)^{\binom{j}{HL}} \right] C,$$

where C is the total number of citations from a given volume of a journal. For simplicity, one volume is assumed to equal a year's worth of a journal's articles.

Thus, assuming the same number of total citations for each volume of the journal is the same number C , the total number of citations TC in the current year from all volumes up to T years ago is the sum of the above quantities for j from 1 to T , that is:

$$TC = \left[1 - \left(\frac{1}{2} \right)^{\binom{T}{HL}} \right] C.$$

Note that as T increases TC approaches C . Thus the citations in a given year from the most recent T volumes of the journal account for a fraction equal to:

$$\left[1 - \left(\frac{1}{2} \right)^{\binom{T}{HL}} \right]$$

of the number of total citations of the journal per year that will be approached in the future.

Assume that DL is proportional to TC (i.e., $DL = k \cdot TC$ for some constant k

and T equal to ER . Thus, by the above, $FC = k \cdot C$. Notice that this FC is the future steady-state count.

From this, we conclude that

$$FC = k \cdot C = k \cdot \frac{TC}{\left[1 - \left(\frac{1}{2} \right)^{\binom{ER}{HL}} \right]} = \frac{DL}{\left[1 - \left(\frac{1}{2} \right)^{\binom{ER}{HL}} \right]}.$$

Thus, 1

$$AF = \frac{1}{\left[1 - \left(\frac{1}{2} \right)^{\binom{ER}{HL}} \right]}.$$

Note that if ER was equal to the full run of the journal, this would still give a small increase to DL .

Model Assumptions

The model is based on a number of assumptions summarized in table 2. For each assumption, difficulties are first presented, then a justification.

1. Citations Decay Exponentially

Assumption: The fraction of this year's citations of the journal from volume one until T years ago will be:

$$\left[1 - \left(\frac{1}{2} \right)^{\binom{T}{HL}} \right].$$

TABLE 2 The Model's Assumptions	
Assumption	Difficulty
1. Citations decay exponentially.	1. May be linear, but doesn't fit data
2. Half-life is continuous.	2. ISI assumes discrete, but doesn't fit publishing practices
3. Half-life is relevant for journals.	3. Not true for all articles, but more true for journals
4. Half-life is similar across all volumes.	4. Probably does change over time, but difficult to correct
5. Local half-life and ISI half-life are proportional.	5. Proportionality fits data, but need more research
6. Citation and use are proportional.	6. Proportionality fits data, but need more research

Difficulties: First, half-life itself may not be the best way to describe journal obsolescence. This is discussed by Endre Száva-Kováts, whose article, "Unfounded Attribution of the 'Half-Life' Index-Number of Literature Obsolescence to Burton and Kebler," is required reading for anyone fond of half-life data. "[I]n their 1960 article Burton and Kebler first made critical and later ambiguous statements, and finally attribute only 'some validity' to the idea of literature half-life."⁴

Second, in 1961, R. E. Burton and B. A. Green Jr. suggested that statistical "median-age" be used instead of half-life, implying a linear rather than an exponential relationship.⁵ Despite the term *median-age*, Burton and Green employ an exponential decay curve to graph the citation pattern. Further, the use of a linear median-age rather than an exponential half-life does not fit the data presented by ISI. (See figures 2, 3, and 4.) Even if median-age were used instead of half-life, however, it would still adjust the download figures in such a way as to give a better estimate of use than a simple cost per download calculation.

Third, the model is not needed for short half-life journals because nearly all the use is likely to occur in the run of electronic access available. Thus, the model is intentionally designed to be most helpful for medium to long half-life journals.

Justification: Despite these difficulties, the model's assumption of exponential decay is a good first approximation as illustrated by the remarkably good fit of the model data to the ISI data in figures 2, 3, and 4.

2. Half-life Is Continuous

Assumption: The exponential decay model is a continuous time model; treating time as discrete leads to error.

Difficulty: One particularly worrisome area is implicit in the ISI measurement of half-life. The ISI half-life figure assumes that a year's issues of a given journal may be treated as if they appeared at the beginning of the year, when, in fact, they

are spread out over the year. ISI's assumption does not cause harm for long half-life journals, but it does for those with a short half-life.

Justification: Because there is little effect on long half-life journals and, in fact, more accuracy for short half-life journals, the model employs a more straightforward calculation based on continuity.

3. Half-life Applies to Journals

Assumption: The exponential decay model applies for journals, but not for articles.

Difficulties: Helmut M. Artus argued forcefully that "the generally accepted assumption of the steady obsolescence of scientific literature is refuted," and one would agree that half-life is undoubtedly not true in a useful way for all articles.⁶ As an example of this, consider Leonard Roth's article, "On the Projective Classification of Surfaces."⁷

Algebraic geometry was at the center of mathematics at the end of the nineteenth century. For a variety of reasons, objects of such complication were being studied that controversies arose over what had been proved and not proved. The subject slumbered until the middle of the twentieth century, when general tools (e.g., sheaf cohomology and algebra) had advanced to the point that many of the difficult complications could be handled by the new machinery. Many of the invariants of the classical period that had not been rigorously defined had natural interpretations as invariants of the new machinery. This led to a renaissance of the subject in the middle of the twentieth century. Sommese (one of the authors of this paper) discovered the important article by Roth and quoted it prominently in his article, "Hyperplane Sections of Projective Surfaces I—The Adjunction Mapping."⁸ Performing a citation search on Roth's article shows that the first citation in ISI is by Sommese. Thereafter, a sequence of twenty citations (excluding Sommese's) continues through 2004.

Justification: Some articles continue to be cited despite their age; others are cited

once or twice and then forgotten. This distinction is an important one to remember, but for the present model, what is being used is the citation pattern for a *collection of articles* (i.e., a journal), not the citations of any individual article.

4. *Half-life Is Similar across All Volumes*

Assumption: Different volumes of the same journal have the same half-life.

Difficulties: As the explanation of the cited half-life from the JCR points out, "Dramatic changes in Cited Half-Lives [sic] over time may indicate a change in a journal's format."

This could be caused by a change in the number of articles published in a given year (e.g., more pages with same density of print or the same number of pages, but denser format); a change in editorial policies; or a change in a field of study.

Justification: An investigation into half-lives changing over time was not pursued. It would be worthwhile to explore the effect of such a change further.

5. *Local Half-life and ISI Half-life Are Proportional*

Assumption: Local half-life is proportional to the corrected ISI half-life figures.

Difficulty: Even if ISI's half-life figures are a good proxy for the citation patterns of the general scholarly community, they might differ markedly from the citation pattern of a particular university or research institute. In "Library Journal Use and Citation Half-Life in Medical Science," Ming-Yueh Tsay studied cited half-life and local use half-life. "[I]n general, journals with shorter citation half-lives also have shorter use half-lives." But, "[t]here is ... a [statistically] significant difference between the mean citation half-life and the mean use-half life for journals of each category [studied]."⁹

Justification: Further research is needed on this assumption.

6. *Citation and Use Are Proportional*

Assumption: The journal being cited and the journal being downloaded should

be in about the same proportions (with some time lag).

Difficulty: The other, more significant problem is that in-house use half-lives may differ significantly from citation half-lives. Tsay demonstrates that for the journals held by the medical library she studied, use half-life was less than cited half-life (e.g., for 266 clinical medicine titles, the use half-life was 3.02 years, but the citation half-life was 6.06 years).¹⁰ The implication is that in this case, citation half-lives may *overestimate* local use.

In "Biology Journal Use at an Academic Library: A Comparison of Use Studies," Diane Schmidt and Elizabeth B. Davis argued that "this technique [of studying citations] does not address the influence of background reading or information gathered for personal, as opposed to professional, use. Another problem ... is that, in general, they measure only the use of journals by faculty or occasionally graduate students."¹¹ The implication here is that citation half-lives *underestimate* local use.

Justification: Because the estimate is applied across the board, one journal's cost per use will be more comparable to another's than an unadjusted calculation. However, further investigation is needed on this assumption.

Practical Problems

Practical problems stemming from the model's assumptions and methods for handling them are summarized in table 3. Because Notre Dame's cancellation project was postponed, the effects of these practical decisions are demonstrated with a sample set of journals listed in table 4.

1. **Do the download statistics need to be adjusted at all?**

The first thing a user of download statistics will have to decide is whether to adjust them. The authors are convinced that download statistics should be adjusted, except in the case of a small set of journals with generally short half-lives. Even if the corrected download statistics do not figure

TABLE 3 Practical Difficulties	
Choice	Our Decision
1. Do the download statistics really need to be adjusted?	1. We had many long half-life journals, so the adjusted download figures were calculated.
2. How should you generate one year's download statistics?	2. We had data problems, so only one complete year's worth of data was used.
3. How should you calculate the electronic run?	3. We had runs available for each journal, so they were used.
4. What do you do if the print and electronic runs overlap?	4. We didn't have good print statistics, so the print was ignored.
5. What do you do if part of the run is available for little or not cost?	5. We didn't consider this at the time, so the current issues were focused on.
6. What do you do with a half-life > 10 years?	6. We used a corrected half-life of 9.0 and were ready to perform least-squares analysis for the borderline cases.
7. What if the half-life is unavailable from ISI?	7. We were prepared to use a corrected half-life of 9.0 or to calculate the half-life needed to make a given cost-per-use cutoff.
8. How do you convert print use to downloads?	8. We needed to make some estimate, so five downloads: one print use was the ratio used.

largely in a library's cancellation decisions, the adjusted figures can inform decisions about borderline cases. Corrected half-lives for the sample set of journals is listed in table 5. Of the seventeen journals listed, seven have half-lives greater than eight years.

2. What time period should be used to compute *DL*, the download statistic for a year? Should the *DL* be from the most recent year or should an average of several years be used?

If there are *K* years of download data, and the total count of all downloads is *TDL* for those *K* years, *TDL/K* could be use for *DL* to smooth out fluctuations. Caution should be exercised, however, because:

- Usage might increase from year to year as the comfort level and dependence on electronic journals continues to increase. *TDL/K* could be replaced with *TDL/K* times the overall ratio of increase over the *K* years (i.e., if the total downloads for a given university increased overall by 3%, the *TDL/K* for each journal could be adjusted by 3%).

- If *K* is not an integer, seasonal variations in downloads will skew the figures (i.e., if *K* only covered 18 months, *K* would be 1.5). Because the variation could be caused by whether classes are in session or not, for example, one could attempt to make an adjustment for seasonal variation. However, it is probably best for *K* to be an integer.

- The electronic run will not be constant over the *K* years (i.e., this year there might be five years of electronic access, but last year only four).

- Lastly, publishers generally acknowledge that there are problems with the statistics when they first began to collect them.

A library should begin with one year's worth of statistics and build from there over time. In table 4, download statistics are listed for 2003 and 2004; in table 5, only download statistics from 2004 are used.

3. What method should be used to calculate the electronic run?

Another minor difficulty is calculating the number of years of electronic availability,

ER. Even for the same publisher, the *ER* for individual journals can vary from a few years to more than ten (especially for new journals). Also, the calculations assume no overlap between electronic and print versions. (See the discussion below.) The library should use the number of years for the entire electronic run, realizing, of course, that even the adjusted download figures from the model will undercount, especially for disciplines more comfortable with paper journals than electronic journals. These subjects also appear to be ones with the preponderance of long half-life journals (e.g., journals in the humanities). However, because the model has been applied across the board, the adjusted download figures are more comparable than the raw download figures by themselves. (See table 5 for *ER* figures for each journal. Notice that *Nature Cell Biology* only began in 1999 and is adjusted accordingly.)

4. What should be done if the print and electronic runs overlap?

At Notre Dame, many print subscriptions have been cancelled in favor of the electronic version only. With few exceptions, however, there was a period of time when the library received both the print and electronic versions of a journal.

If a library is gathering reasonably accurate print and electronic use statistics, the statistics could simply be added together for the overlapping time period. Of course, there is the problem of the definition of "use." Download statistics count the use of one article as one use; print statistics generally count one current issue or one bound issue as one use. Obviously, these are not comparable.

For the sample set of journals, the assumption was that there was no overlap between electronic and print and that the print use statistics could be ignored (they are not generally collected well or with

TABLE 4
Sample Set of Journals

Rank	Title	2004 Cost	2003 DL	2004 DL	2004 CPU
1	Nature Cell Biology	\$899	n/a	348	\$2.58
2	SIAM Journal on Numerical Analysis	\$508	61	69	\$7.36
3	Evolution and Human Behavior	\$808	89	109	\$7.41
4	Library & Information Science Research	\$315	17	13	\$24.23
5	Accounting, Organizations & Society	\$1,633	65	50	\$32.66
6	Acta Psychologica	\$936	43	28	\$33.43
7	Probabilistic Eng Mechanics	\$832	771	24	\$34.67
8	International Journal of Industrial Organization	\$1,169	40	32	\$36.53
9	Physics Reports	\$5,599	198	153	\$36.59
10	Immunology Letters	\$2,734	89	69	\$39.62
11	Journal of Molecular Structure: Theochem	\$7,633	216	192	\$39.76
12	Earth Science Reviews	\$1,334	13	33	\$40.42
13	Mathematische Annalen	\$2,760	58	60	\$46.00
14	Communications in Partial Differential Equations	\$1,995	n/a	38	\$52.50
15	Poetics	\$433	19	3	\$144.33
16	Technological Forecasting & Social Change	\$839	5	4	\$209.75
17	Journal of Logic and Algebraic Programming	\$923	1	1	\$923.00

regularity). The adjustments were made using only the complete electronic run.

5. What should be done if part of the run is available for little or no cost?

Because the ultimate goal is to maximize use while minimizing cost (i.e., maximizing the

ROI), there are implications for journals available for little or no cost after some number of years (e.g., through JSTOR or via various open-access arrangements).

As a first example, consider *Mathematische Annalen*. The years 1996 to the present of this journal are available elec-

TABLE 5
First Adjustment to the Sample Set of Journals

Rank	Adj. Rank	Title	2004 Cost	2004 DL	2004 ER	2003 HL	AF	FC	Adj. CPU
1	1	Nature Cell Biology	\$899	348	6	1.7	1.1	381.0	\$2.36
2	2	SIAM Journal on Numerical Analysis	\$508	69	8	9*	2.2	150.0	\$3.39
3	3	Evolution and Human Behavior	\$808	109	8	3.0	1.2	129.4	\$6.25
14	4	Communications in Partial Differential Equations	\$1,995	38	4	8.5	3.6	136.5	\$14.61
9	5	Physics Reports	\$5,599	153	7	7.7	2.1	327.3	\$17.11
5	6	Accounting, Organizations & Socty	\$1,633	50	10	9*	1.9	93.1	\$17.54
4	7	Library & Information Science Research	\$315	13	10	5.3	1.4	17.8	\$17.68
6	8	Acta Psychologica	\$936	28	10	9*	1.9	52.1	\$17.95
12	9	Earth Science Reviews	\$1,334	33	10	9*	1.9	61.4	\$21.71
13	10	Mathematische Annalen	\$2,760	60	9	9*	2.0	120.0	\$23.00
7	11	Probabilistic Eng Mechanics	\$832	24	10	6.1	1.5	35.3	\$23.54
8	12	International Journal of Industrial Organization	\$1,169	32	10	6.6	1.5	49.2	\$23.75
10	13	Immunology Letters	\$2,734	69	10	4.4	1.3	87.0	\$31.42
11	14	Journal of Molecular Structure: Theochem	\$7,633	192	10	4.3	1.2	239.8	\$31.82
16	15	Technological Forecasting & Social Change	\$839	4	10	8.6	1.8	7.2	\$116.07
15	16	Poetics	\$433	3	10	n/a			\$144.33
17	17	Journal of Logic And Algebraic Programming	\$923	1	10	n/a			\$923.00

* ISI Half-life > 10

tronically by subscription from Springer Verlag. However, EMANI (Electronic Mathematical Archiving Network Initiative) makes all the issues from 1996 and earlier available for free. This is laudable, but because the older issues are free to anyone (even nonsubscribers), the download statistics from Springer for the subscribed issues should not be adjusted. As a result, the cost per download is higher and it is more likely that the subscription should be replaced by some other means of access (like commercial document delivery).

As a second example, consider journals that are available before a certain moving wall at a cost less than a subscription. The *SIAM Journal on Numerical Analysis* (*SINUM*) was available before 1996 through JSTOR (*JSTOR:SINUM*). Because the decisions for both these subscriptions are separate and because "all the years for each period" are available electronically, the raw download statistics should be used with no half-life corrections. A 2004 subscription to *SINUM* is \$508, and the cost of *JSTOR:SINUM* is about \$25. During 2004, the raw downloads were sixty-nine and sixty-six, respectively. Here are the results of the analysis:

- The cost per download was \$7.36 for *SINUM* and \$0.38 for *JSTOR:SINUM*.
- If *JSTOR:SINUM* was not available, the figures for *SINUM* would be adjusted, resulting in a cost of \$3.39 per download as indicated in table 5.

6. What should be done with an ISI half-life of "> 10" years?

For a journal with an uncorrected ISI half-life of at least ten years, ISI simply gives > 10 as the half-life. When ISI's data are corrected, the journal has a half-life > 9 years. Can this be calculated more accurately?

For these cases, a least-squares fitted exponential curve can be employed to compute a more accurate estimate of half-life. The procedure would be as follows:

Let C_i denote the cumulative fraction of cites from the journals in the current year

(for us, 2003) plus the i years before that. Find the value of HL such that

$$\sum_{i=0}^9 \left(C_i - \left[1 - \left(\frac{1}{2} \right)^{\left(\frac{i}{HL} \right)} \right] \right)^2$$

is minimized (e.g., using the solver function in ExcelTM).

The citation data were adjusted so that 0 percent was the prediction for cites from year zero (i.e., cites in 2003 journals to the 2003 volume of the journal in question). For long half-life journals this is close to true because the fraction C_0 is typically less than 0.5 percent.

More accurate computations of the > 10 half-lives using least-squares minimization could give valuable guidance in contested cases. However, for journals with half-lives greater than four or five years but less than ten, least-squares minimization tends to give a half-life slightly less than the half-life given by ISI. For journals with half-lives less than four years, least-squares minimization should not be used. (For the sample set's results employing least-squares minimization, see table 6.)

7. What if the half-life is unavailable from ISI?

JCR does not provide cited half-lives for journals cited less than one hundred times nor does it provide half-lives for every journal published (for example, new journals). As noted above, of the 327 mathematics and applied mathematics journals listed in ISI, 11 percent (36 journals) have no half-life given. Of the 261 biochemistry and molecular biology journals listed in ISI, 1 percent (3 journals) have no half-life given. The analysis uncovered the fact that nearly 50 percent of the journals across all subjects did not have half-life data available.

If ISI has data on a particular journal (even if it has been cited less than one hundred times), it might be able to estimate the half-life as was done for those with > 10 years (above). Another approach is to

TABLE 6
Final Adjustment to the Sample Set of Journals

Rank	Adj. Rank	LS Adj. Rank	Title	2004 Cost	2004 DL	2004 ER	2003 LS HL	AF	FC	Adj. CPU
1	1	1	Nature Cell Biology	\$899	348	6	1.7	1.1	381.0	\$2.36
2	2	2	SIAM Journal on Numerical Analysis	\$508	69	8	11.6	2.6	181.5	\$2.80
3	3	3	Evolution and Human Behavior	\$808	109	8	3.0	1.2	129.4	\$6.25
5	6	4	Accounting, Organizations & Society	\$1,633	50	10	15.3	2.7	136.9	\$11.93
14	4	5	Communications in Partial Differential Equations	\$1,995	38	4	8.5	3.6	136.5	\$14.61
6	8	6	Acta Psychologica	\$936	28	10	10.8	2.1	59.2	\$15.82
13	10	7	Mathematische Annalen	\$2,760	60	9	13.7	2.7	164.0	\$16.83
9	5	8	Physics Reports	\$5,599	153	7	7.7	2.1	327.3	\$17.11
4	7	9	Library & Information Science Research	\$315	13	10	5.3	1.4	17.8	\$17.68
12	9	10	Earth Science Reviews	\$1,334	33	10	10.7	2.1	69.1	\$19.31
7	11	11	Probabilistic Eng Mechanics	\$832	24	10	6.1	1.5	35.3	\$23.54
8	12	12	International Journal of Industrial Organization	\$1,169	32	10	6.6	1.5	49.2	\$23.75
10	13	13	Immunology Letters	\$2,734	69	10	4.4	1.3	87.0	\$31.42
11	14	14	Journal of Molecular Structure: Theochem	\$7,633	192	10	4.3	1.2	239.8	\$31.82
16	15	15	Technological Forecasting & Social Change	\$839	4	10	8.6	1.8	7.2	\$116.07
15	16	16	Poetics	\$433	3	10	n/a			\$144.33
16	17	17	Journal of Logic and Algebraic Programming	\$923	1	10	n/a			\$923.00

simply use a figure of 9.0 years across all subjects without a half-life (ISI half-life of = 10.0 - 1). If journals having a half-life of > 10 were left with a corrected half-life of 9.0, this estimate gives journals without half-lives the maximum benefit.

For borderline cases, the model could be employed in reverse using the goalseek function in Excel™. If there were a particular CPU cutoff, say, \$40 per download, one could calculate what the half-life would need to be for a particular journal to make that cutoff. In table 5, *Poetics* would need a half-life of approximately twenty-one years to make the \$40 CPU cutoff. This might be reasonable. The *Journal of Logic and Algebraic Programming*, however, would need a half-life of 156 years. This is clearly unreasonable and the journal should be a candidate for cancellation.

8. How is use from electronic downloads converted to expected document delivery requests?

One of the major problems with analyzing use statistics is estimating the conversion factors between downloads, print uses, and commercial document delivery requests. Would a journal used five times electronically be used only once in paper? Would a print volume used once result in one commercial document delivery request? Does it matter if the docdel is mediated or unmediated? Estimating this conversion factor is critical when determining whether a subscription (print or electronic) is more cost-effective than docdel. Unfortunately, there appears to have been no previous published research in this area.

During the project, the plan was to use a conversion factor of five downloads to one mediated docdel request. Conversion factors for print use were not estimated. The download to docdel figure was based on docdel requests before the library had access to journals electronically and on a sense that patrons would request fewer articles if they had to ask for them (rather than clicking on a hyperlink).

Final Remarks

For the sample set, the average cost of docdel for the University of Notre Dame, roughly \$35 per article, was used. An estimated internal processing cost of \$5 was added, and the cutoff was set at \$40 per article. Journals that cost more than \$40 per adjusted, converted use would be candidates for cancellation. In tables 4, 5, and 6, journals are ranked from lowest to highest CPU. In tables 5 and 6, previous rankings for the journals are included. Even the results in table 5 are better than those in table 4.

No model is perfect, but the half-life model fits the citation data surprisingly well. The goal was to improve the method used to evaluate the raw download figures. Indeed, the proposed model will still undercount the same *areas* that the raw download figures undercount, but the undercounting will be proportional across disciplines and less severe.

It is important when evaluating whether approximations are “acceptable” to keep the goal in mind. The goal is a reduction in undercounting, and undercounting is much more severe for long half-life journals with short electronic runs available.

For example, look at a journal with a very long half-life, *Mathematische Annalen*. According to JCR, this respected mathematics journal has a half-life of > 10 years with an electronic run of nine years. The ISI database does not give the exact half-life; however, only 24.2 percent of citations are from articles published between 1994 and 2003. Using least-squares minimization, the “ISI half-life” is 23.2 years. *Math. Ann.* moves from a \$46 CPU in table 4 to a \$23 CPU in table 5 and to a \$17 CPU in table 6.

As time passes, the electronic runs of journals will increase, and there will be sufficient year-to-year raw download figures to make reasonable extrapolations. Thus, the model for adjusting downloads that this article proposes will become less urgent.

Notes

1. Marisa Scigliano, "Serial Use in a Small Academic Library: Determining Cost-effectiveness," *Serials Review* 26 (2000): 43–52.
2. JCR Glossary. Available online at <http://jcr4.isiknowledge.com/www/help/hjcrpls2.htm>. [Accessed 27 February 2005].
3. Lawrence A. Coleman, "Exponential Growth and Decay," *Macmillan Encyclopedia of Physics*, vol. 2 (New York: Simon and Schuster Macmillan, 1996), 533.
4. Endre Száva-Kováts, "Unfounded Attribution of the 'Half-Life' Index-Number of Literature Obsolescence to Burton and Kebler: A Literature Science Study," *Journal of the American Society for Information Science* 53 (Nov. 2002): 1098–1105.
5. R. E. Burton and B. A. Green Jr., "Technical Reports in Physics Literature," *Physics Today* (Oct. 1961): 35–37. "While the phrase 'literature half-life' has been applied to this figure, it more properly should be referred to as the median age."
6. Helmut M. Artus, "'Halbwertszeit wissenschaftlicher Literatur' – Naturgesetz oder Forschungsartefakt?" *Nachrichten für Dokumentation* 34 (Apr. 1983): 79–86. He also claims: "Ein bloß quantifizierendes Vorgehen reicht nicht aus, um das gleichermaßen kognitive wie soziale Phänomen 'Literaturnutzung' in den Griff zu bekommen." "A purely quantitative procedure is insufficient to come to terms with the cognitive and likewise social phenomenon of 'literature use.'" (Translated by Robert L. Kusmer, Associate Librarian, University Libraries of Notre Dame.)
7. Leonard Roth, "On the Projective Classification of Surfaces," *Proceedings of the London Mathematical Society, Second Series* 42 (1937): 142–70.
8. Andrew J. Sommese, "Hyperplane Sections of Projective Surfaces I—The Adjunction Mapping," *Duke Mathematical Journal* 46 (1979): 377–401.
9. Ming-Yueh Tsay, "Library Journal Use and Citation Half-life in Medical Science," *Journal of the American Society for Information Science* 49 (Dec. 1998): 1283–92.
10. *Ibid.*, 1286.
11. Diane Schmidt and Elizabeth B. Davis, "Biology Journal Use at an Academic Library: A Comparison of Use Studies," *Serials Review* 20 (summer 1994): 45–63.