

4-17-2007

## Are They Any Use? Hazards of Price-Per-Use Comparisons in e-Journal Management

Jason S. Price

*Claremont University Consortium*

---

### Recommended Citation

Price, Jason S., "Are They Any Use? Hazards of Price-Per-Use Comparisons in e-Journal Management" (2007). *Library Staff Publications and Research*. Paper 7.  
[http://scholarship.claremont.edu/library\\_staff/7](http://scholarship.claremont.edu/library_staff/7)

This Conference Proceeding is brought to you for free and open access by the Library Publications at Scholarship @ Claremont. It has been accepted for inclusion in Library Staff Publications and Research by an authorized administrator of Scholarship @ Claremont. For more information, please contact [scholarship@cuc.claremont.edu](mailto:scholarship@cuc.claremont.edu).

## **Are they any use? Hazards of price-per-use comparisons in e-journal management.**

**Jason Price**

General hazards which could otherwise be known as observations without benefit of data, so as a first step I'm just going to tell you things that I believe or know to be true but I don't have data to support them. So, defining use narrowly, the vagaries of user behaviour, the different dissemination styles in teaching, particularly at a teaching institution such as Claremont Colleges where I come from, and the granular usage reports.

My perspective is narrow, it's e-journal-centric, it's package-centric and it's approached from an academic institution's perspective. Given those caveats, let's move forward. So, a narrow definition of user one general hazard. The Counter journal report one provides us with full text article requests, the total number of downloads from our resource. These are by far the most common numbers to use when measuring cost per use, but it really is only one dimension of use and in making a larger evaluation of resources it makes sense to look at other measures of use as well if not along side at least in addition. So additional measures which I'll just briefly mention, A-Z list, the quick user weblog files, also the data that the library itself has on uses of their resource. The time cited in recent papers from your institution so you can use for instance Web of Science data to find out who your faculty has been citing and what journals those are, compile that over a series of years and that provides another measure of local use. If you're interested in that I have a brief paper in ACS Livewire that you could go see. Impact factor is certainly a global measure of use although use through citation is still used and the number of papers published by local researchers by journal so this would probably necessitate a larger place.. aware of faculty publishing those are another measure of use of journals. And then faculty researcher surveys and print use of .. in many cases and also there's this emerging idea of using something very similar to Google's page rank on a large collection to try to catch use transmitted a little bit earlier than things like impact factor might.

General hazard number two, the vagaries of search use are hazards. Users may very well check for full text before judging the relevance from abstracts and probably the most extreme example of this is a case where they don't even really have the full titles to look at so using Google Scholar a lot of times the title of the paper will be cut off. How do you judge relevance from this? There's a little snippet which might help but they may very well just click in, see if they have access and then decide whether it's relevant and if they've downloaded the whole paper as part of that process which the links often do, then you've recorded a use for something that was useless to the user. So this a way that the numbers are not as meaningful as we might hope. A second deadly robot hazard, the Google accelerator. This is the pre-loading of links from a page that user brings up so if a user brings up a page of ten results that are all pdf's, if they have Google

Accelerator working it may download all of those links which the user may not even look at or go to but that could get counted as a use.

General hazard number three: dissemination style in teaching. So Professor A downloads one pdf and makes copies for students, either photocopies or they send them out by e-mail. This would count one use for many students, so that's not necessarily a reflection of how valuable that particular article is or how much that article is used in the institution. Professor B might send a link to the publisher of a pdf to 40 students then those 40 students would all use that link - that counts 40 uses for that one article which you might think is over-counting it. In addition, students might just go back to that link repeatedly to look at the paper rather than downloading it themselves so use is maybe not as precise as we think it is when we do these quantitative measures. In another example Professor C could post a pdf in the electronic reserve site, that's one download from the publisher but then many downloads that aren't measured in the publishers' statistics by students or whoever else is using it.

A fourth general hazard is usage report granularly so when we look at statistics from Counter reports we get the number of uses per title, we don't know which years the articles in that title came from when we look at the numbers and so at this point we can't separate the front file from the backfile use, whether that backfile is one we purchased or whether it's freely available. Either way, we're not really measuring use of what we paid for, which I think in most cases for librarians making decisions, is what they want to know. So that's a place where we need to take a step forward or we may get somewhat hazardous results.

So specific hazards of cost per use: determining cost is not as easy as you might think. Comparing the cost per use to inter-library loan cost, comparing across publishers in cost per use ignoring the by-title data and lack of benchmarks which is probably my personal favourite. So I'll show you data that addresses each of these. So briefly though, first, Counter stands for Counting Online Use of Network Electronic Resources, it is a standard and code of practice that enables comparison and usage statistics from different vendors. Probably most of you are familiar with that but just briefly, my snapshot view of what it is and what it does, it standardises terminology and definitions, the layout and the format of the journal and database reports are standardised, the processing of user input so if a user triple clicks on a link because it's too slow for them Counter has set up a way that there is a standard as to how long it has to take between clicks for it to be counted as an extra use, or an extra view. And the frequency and availability of reports are standardised. And finally testing and regular audits to begin quite soon. Only then can a publisher claim to be Counter-compliant and be listed on the Counter website. So, specific hazard one I said was determining cost. It seems pretty easy, and it is, to calculate overall costs. If you've never done this and you're considering it all you have to do is make sure you have one Counter report from the end of a year. If you have that you can do it. You need one year's cost and one year's article requests, for the overall package and they come packaged

that way in the spreadsheet. So if we spent \$58,600 on our publisher e-access fee for a package and we had about 36,000 articles views then you would think, yes by dividing that our cost per use for this package was \$1.64. However, there are hazards there because this is what we pay in terms of the e-access fee, I think worldwide now most packages of electronic journals, do, particularly with consortia deals, require committing to maintain a minimum span, you commit to keep those titles forever, I call them lock-in deals. So in order to get this kind of access for this amount we also have to maintain a \$420,000 worth of subscriptions from this publisher and that money goes to the agent separately in small chunks. So in calculating our real cost per use for this package you really need to combine both sets of payments and then you come out with a number like \$13.40 per download. So just to drive that point home if we divide all the titles we get from this publisher into our subscribed titles and our unsubscribed titles, the subscribed titles we're paying complete list price plus any access fee so if we just look at the use of those 192 titles and their total cost we come up with a really relatively high cost per use, more than \$30 per download. Here it's much lower for the unsubscribed titles because we pay only a small portion of their list price in order to get access but we wouldn't get this cost per use without this cost per use so an appropriate number is the one that combines all that we pay for all the electronic access that we have.

Second specific hazard: comparing this number with inter-library loans. So for that same package our cost per viewer use was \$13.40 - what does this tell us? Is it high, is it low? It's hard to say but I often hear both content providers and librarians say, well if inter-library loan costs us about \$20 per article and our cost per view for this collection is \$13.40, well then that's a good deal. Well, I think this is a red herring because it's clearly the case that the views that we're measuring in an online environment are not similar to the number of requests for those same articles that would come through an inter-library loan. Some of them are for judging relevance, some of them they would only want if available immediately, I really think at this point knowing what the cost per use or cost per view is doesn't tell us much because we don't know how to relate inter-library loan numbers to full-text downloads. I think, although the study would be difficult, it would be really interesting to look at how those two relate.

So this is the third and most involved example but I think it's also quite important: Specific hazard three is cross-package comparisons. So this package 1 is the same one I've been talking to you about although this is data from over two or three years so the costs are higher and the cost per view is slightly different. You take your total use, you add your subscribed cost, your unsubscribed cost and you get an overall cost per view. Well, the way this works it suggests that package one coming in at about \$13 is better value than package 3 which comes in at \$22 and ultimately the Counter standard was built to allow us to compare usage statistics across vendors so this is the kind of thing you might want to do but I think it's particularly dangerous and I have some reason for that which I'll try to show you now. So, when we look at the full text requests of views we're combining the pdf

and the html formats. This on the X axis is various publishers and on the Y axis is mean pdf to html ratio. It sounds complicated but I think I can make it simple. So from Highwire for every html download there is, there's also a pdf download. This is data from Phil Davis at Cornell, we worked together on this paper. But for Wiley, for every one html download, there are 20 pdf downloads. Is this a problem, what does it mean? That was one of the things we tried to figure out. Well, I separately observed something that I think may explain it, at least in some cases. So I went ahead and looked at the html/pdf ratio in all of our three packages that we're looking at right now and it turns out that package 1 had a pretty low ratio, about one to one, whereas these other two had very few html per pdfs. So the question is, how many of the pdfs recorded here in this bar are duplicates of html use recorded in this bar - are we looking at the use of the same article, and that was the observation that I made of Phil's data. So if you use a link resolver, our institution, to get a ScienceDirect article, what we get at Claremont and this is not the same as all institutions, there is variations even within the publisher, what we get is the full text. So here's the introduction, background, the figures are all here. So I automatically get the html when I request that article. To get to the pdf, I have to click on that link. So what have I done if I wanted the pdf in the first place? I've recorded two uses, an html use and a pdf use. Most other sites are not set up this way and in fact other Science Direct sites at other institutions aren't. But ours is, so no wonder the ratio is higher and that means, if I'm comparing the total number of downloads from Elsevier and other providers I'm comparing apples to oranges. So instead if we go back and compare our numbers from package one and package three when we looked at the total downloads, it looked like three was a lot better than one, one being Science Direct. If we, however, look at the cost per pdf download we get extremely different numbers, in this case package one seems to be the same value as package three in terms of cost per use. So this can have significant impact on how you compare and evaluate publishers. And Counter has already developed a response to this through Elsevier and Ebsco working together with Counter to produce a unique article filter that measures the unique article requests so it de-duplicates if you've got html then pdf, it doesn't count that as two it counts that as one. What they found was that they had about a 20% to 30% decrease in the total use numbers, so there was 20% or 30% duplication. And then there was some variation across different publishers but those numbers are publisher wide. What I want to do, and maybe this'll be me next week, is look at the data for specific institutions. For instance I could take the institutions that have the kind of Science Direct set-up that we do and I don't even know how it got set up that way, and others that have it set up differently and so it might be that for us the duplication is much greater.

But now it's time for a reality check. Should we expect the cost per use to be equivalent amongst packages. I think there're a lot of reasons why we shouldn't. If different publishers provide different quality maybe we're willing to pay different amounts. Secondly, scope. So we are considering purchase of the China academic journals from EastU. We have only five scholars who read Chinese and this database is entirely Chinese language

journals. They have hundreds and hundreds of journals, massive amount of content, huge undertaking, expensive thing to do. We shouldn't expect to pay the same cost per view for that than we would for a larger package that has wider appeal. Business model could affect it for profit versus cost recovery, you might expect to pay different amounts to different publishers, and exposure in Google Scholar, again, is an issue. I can't ignore this because of how much I know people come through it, if one publisher is heavily exposed in Google Scholar and we send all of our introductory students in there to find a couple of articles they need to study, if they're looking through and they look at 50 before they find the right one and there's a lot of one publisher's content in there and not a lot of another publisher's, again we skew the number of downloads, the number of uses and therefore the cost per use data.

Specific hazard number 4: Ignoring by title data, so, so far everything I've shown you was based on two numbers out of a Counter report per year plus the cost of that resource, that's not hard to do. But the by-title data is important too and I hear resignation from most librarians I talk to, 'well, we're locked into these titles we can't really manage that, why do I care what the cost per use is of my individual titles?' Well, we hear of some really high ones, this is really a cost per use for these particular titles of \$500 per download, \$800 per download, \$1200 per download. They have been mitigated by the fact that these are the subscribed titles that we get, some other lease titles for much lower amounts, but still these titles are high-priced titles that we are not using, why should we keep them, when most journal deals allow for swapping of titles. So we will have eventually faculty requests or new titles that we want to add, these are the ones we should get rid of and replace with other journals from the same publisher to keep that spend up but to select different titles. One other use, and I'll just go over this briefly, of the by-title data is the cut off the long tail, so we saw examples of this yesterday but here, really, it's only a very small percentage of the titles, titles ordered by use for each institution here. What this is, is all the titles that had almost no use, or zero use, so it's a very few number of titles that had the majority of the use in this package in our consortia. Well, what we did was we said, I'm going to look at each institution's use order and cut off the end of the tail, take all of those titles which no-one's really using much and create a smaller shared package which then costs us less and presumably we would use more of. Here's evidence that we do use more of it so the original data is the blue data, the proportion of titles containing 80% of the use roughly 20% and in some cases a lot less than that. After we made this new list based on people's individual title use data, we broadened the use so closer to 35% of the titles represent 80% of the use so more of the titles in that package are important and in fact there are many fewer with zero uses.

So finally, and this is my favourite and I think it's the most important, you have a cost per use, you've calculated, you've been careful, you know that your cost per use is about \$20 per article for a given package. What does that mean? Well, if you ask another institution they might find theirs is only \$10, well that's a bad deal I'm paying twice as much as they are for articles

from this same package and we have the same interface setup and whatever else. But you ask somebody else you may find you're paying less than half. What we really need to do to benchmark cost per use is not so much compare across publishers as compare across institutions within our buying group, whatever that buying group maybe. So here's that one we were looking at before, it doesn't stand out so much anymore, it's about in the middle. These guys though are getting a raw deal though, it looks like so we have options, we can then on the next renewal ask the publishers to consider changing the pricing options for those particular institutions. Alternatively, if that's too complicated we can also re-distribute the cost within the consortium on our end so that cost is more reasonable. And again there are still other explanations other than value that might explain these numbers so you'd want to look at that and see if there are institution-specific explanations. But this is where I think cost per use data gets value is from looking at the costs of other organisations in your buying group and you can actually get that data from within your consortium whereas outside obviously there are problems with sharing both the use data and the cash.

And then something for the publishers in the room is you can kind of flip that and say, okay here are a couple of institutions that have extremely low cost per use and the publisher has a few add-on packages that they want to sell, this just happened in our consortium, they offered us these new titles, librarians didn't know what they were or why would want them necessarily but I can go or encourage the publisher to go to those institutions and say 'look, your people are using these a lot, your value of this package is huge, we recommend that you expand it because your users have obviously found this source and are using things from it.'

So, then in summary, some recommendations. Ensure that you have the right costs, that's a little bit easier said than done but it's worthwhile because otherwise it's just a number. Be wary of cross-publisher comparisons, consider both overall and pdf use or, I'm hoping that the filter that Elsevier and Ebsco have developed can be implemented broadly in the next Counter release. Ultimately across publishers should the price be the same. For single package evaluation, look at patterns on a title level particularly when that is a value to you when you are considering cancellation or swapping. Benchmark versus consortium or peers - that's really critical, so we need to support effort and the thing is you as individual librarians at your library aren't going to be able to each separately ... *(change of tape, but missing)* it's a fantastic standard and it's continuing to improve, it's really encouraging to someone who wants to use usage data meaningfully to know that they're working on it and doing lots of things. Couple of challenges, indication of subscription types, subscribe versus lease titles, how we pay for things and being able to know which title falls where, that's probably outside the offices of Counter but it's something librarians have to come up with in some cases if they're doing evaluations. Separation of backfile data, that's something that I noticed there'd been some discussion in Counter about the possibility of reporting use data, data on article use by year so for a particular journal you had year by year use so that you could separate out the backfile data if you knew that 1997 to the

present is your current subscription you could just take the things out before that and not include those numbers in your analyses. The unique article filter is great, I'd love to encourage that to happen, they've always provided good by-title data and a recent thing that makes the consortia argument that I have set out more feasible is that the Counter release two already in place requires single password consortium access to aggregate and by-institution statistics so what I'm saying is, now I have one password, I can look at the overall use in the consortium and I can also gain access to each institution's statistics so that it is relatively easy to gather this data and do this kind of analysis as the consortia needs to make decisions about renewals and re-negotiations.  
Thank you for your attention.