Patterns and costs of printed and online journal usage
Oliver Obst, Head, Medical Branch Library, ULB Muenster, Germany

Abstract
This study in an academic medical sciences library setting examines the correlation of usage of a matched set of print and online titles, the validity of e-journals usage statistics and the impact of online journals on print journal usage. The print and online usage was determined for 270 journals, both versions of which were available. Print usage was determined annually since 1997 using the reshelving and the error-copies method. Online usage statistics were delivered by five publishers and corrected for redundant multiple accesses. Print journal usage decreased by 22.3 and 30.2% respectively over each of 2 years after the introduction of online journals. Journals published both in print and online lost 30.4% of their print usage within 2 years. The total loss of usage of print-only titles in the same period was somewhat higher, at 45.8%. The average correlation between online and print usage is 0.60 and 0.67 respectively. For the examined titles, users accessed the online versions ten times as often as the print version. Two clearly distinguishable groupings emerged: while with Academic Press and Elsevier, e-journal usage exceeded print usage by a factor of 3 or 4, the e-journals of Blackwell, HighWire and Springer were used on average 14.6 times as frequently as the corresponding print journals. Each usage of a print article cost 2.79–50.82 €, each usage of an online article 0.31–15.10 €, depending on the publisher. On average, the usage of an online article was 5.4 times cheaper. Within 2–3 years the usage of online journals has outstripped that of print titles by a factor of ten, but the specific spectrum of usage remains much the same as when only print journals alone existed. Print titles not available online suffer a greater decline in usage compared with print/online journals. This confirms that what is read or purchased is determined primarily by ease of access and that there is a steady tendency to reduce the multiplicity of access modes to a manageable few. The availability of journals online seems to have created a new clientele, at least in the case of the German-language Springer journals. The connection between supplier and supply is much less clear with e-journals than it is with print titles. Therefore it is very important to stress and encourage the role of the library as the supplier of this sort of information in the university environment. Collection building issues are discussed in the light of the results.

Background
Five years since the introduction of online journals, this new medium has been widely accepted in the medical faculty, thanks to the ease
of access and its user friendliness. In 1998 only 26% of all purchase requests were for the online version of journals, whereas 1 year later the number had already risen to 72%, and by 2002 it was exceptional for faculty staff not to request e-journals. Their hesitancy is most likely to be associated with concerns about whether sufficient infrastructure is in place for students to have access to online journals, including the facility to print out articles. Were it not for such a consideration, they would most likely conclude that the online version is perfectly sufficient. We know also from other studies that e-journals now feature among the indispensable tools of most medical professionals.1-3

This paper is not about the purchase or administration of online journals, a topic that has been discussed elsewhere at length and in detail.4 The subject of this paper is the correlation of usage and cost of print and online titles at this intermediate stage, and the usage assessment of e-journals.

Introduction

For 2 years, beginning in 1997, this library acquired e-journals rather haphazardly, in much the same way as did many other libraries. As a branch of the University Library, the Medical Branch Library played a comparatively passive part in the initial contracts of the Northrhine-Westfalia (NRW) Consortium with Springer and Elsevier5 who, already in 1998, together supplied around 500 biomedical e-journals. A short time later, the (unintentional) shift of paradigms was accomplished: in 1999 for the first time the number of e-journals supplied by the Branch Library exceeded that of its print titles (Fig. 1). This trend was further enhanced by the arrangement of contracts that tended to favour package deals, as well as by the decrease in print stock on account of low budgets and high publication costs.

These factors led to a paradoxical situation: while in the past the greatest effort was made to evaluate and supply only the most important and cost-efficient titles of journals in print, the cancelled titles were now being re-offered in online form through the consortium contracts, alongside a large number of other, extraneous material that was of no special interest to library patrons.

From the very outset, the library tried to establish its own online access to medical journals, but this effort was for the most part tightly focused on specific existing or potential requirements. Subscriptions to two of the leading journals (Journal of Biological Chemistry and Proceedings of the National Academy of Sciences of the USA) were taken out in the online version as soon as they became available; hundreds of other online subscriptions followed.

Three years after the first releases, two factors emerged, which were responsible for the Branch Library’s independent participation in consortium contracts, and its concluding contracts separately from the main University Library, first with publishers like Academic Press and Harcourt, later also with Blackwell, Cell Press, Elsevier and Nature Publishing Group. These factors were:

1 A non-cancellation clause: the University Library could not guarantee the continuation of all the institute’s journal subscriptions. This consequently contributed to the discontinuation of the contract with Elsevier. The Branch Library, by contrast, only needed to guarantee the few decentralized subscriptions to publications that the Medical Faculty still maintained, as most journals already had been centralized.

2 Discipline-specific supply: publishers with a range of primarily medical titles, such as Blackwell Science and Harcourt Health Services were
not particularly attractive to the consortium of university libraries.

Whereas the Branch Library annually evaluated the usage of its print journals, for a long time nothing was known about the level of use of the presented e-journals. Libraries must have knowledge of access information, even with this type of supply, in order to pursue a sound purchasing policy. In this paper, we are mainly concerned with the following questions:

1. Are there any discernible patterns in the usage of the two forms of presentation? Do titles in print form, which were widely read before the introduction of online versions, continue to show high usage rates?

2. How has the usage of print journals been altered by the supply of online versions? Is the usage of print journals decreasing in direct proportion to online usage increase?

3. What are the cost efficiencies of print and online journal?

The results of this and of other studies may provide recommendations for the efficient management of e-journals.

Methods

Print and online usage was determined for journals, both versions of which were available. Very few publishers provided access statistics of online journals; this study made use of the following statistics, all of which related to package deals apart from those provided by HighWire Press (see Table 1).

All publishers, apart from HighWire Press, delivered these statistics without explaining how the data was obtained and without defining the type of access, as has already been criticized by the Association of Research Libraries. HighWire provided the most detailed statistics: access to homepage, current issue table of contents (TOC), all TOCs, searches, abstracts, full text HTML and PDFs were separately itemized. Access was also divided according to ‘total usage’ and ‘unique events’. (See Berg on the problem of the usage concept in e-journals.)

To determine the usage of print titles, random sample tests were conducted in two phases: (i) prior to the introduction of online titles (1997–1998), and (ii) after the establishment of e-journal availability (1999–2001). In the first phase, the usage of print journals was ascertained by the reshelving method, whereas in the second phase the error-copies method was utilised. The error copies consisted of faulty copies of articles, which the users had thrown away. These were collected from the paper baskets alongside users’ checked-off reading lists. Print and online usage were each projected over 1 year. The correlation of print versus online usage was determined according to Bravais-Pearson.

Results

Total usage of print titles was determined by the total copy volume from which monograph copies (25%) were subtracted. The resulting number was divided by 10 to give the numbers of articles copied (adapted from Bauer). As shown in Fig. 2, the total usage of print titles has clearly been decreasing since 1999. There is a decrease of 22.3% in 2000 and 30.2% in 2001 as compared
with the respective previous year (45.8% decrease over 2 years). As demonstrated by Table 2, the decrease in usage of those titles that offered combined print and online access is clearly lower than that (30.4% for slightly more than 2 years).

By contrast, the usage of e-journals first doubled and then nearly tripled within the last couple of years, as illustrated by Fig. 3, which reproduces usage data relating to the Journal of biological chemistry and Blood (HighWire Press).

Table 3 illustrates the correlation as well as the ratio of absolute usage of both versions. In the first and second phase, online and print usages were clearly in concurrence with a correlation coefficient of 0.67 in Phase 1 and 0.60 in Phase 2. Online usage exceeds parallel print usage by a factor of 7.96, and print usage prior to e-journal publication by a factor of 4.59.

The following is an example in explanation of Table 3: during the year 2000 the Medical Branch Library received a total of 57 journals from Elsevier in print and online form. In the year 2000, 11 455 articles of these titles were downloaded
full text, whereas print versions were used 3489 times during the same period. The correlation coefficients of 0.72 for both 1997–1998 and 1999–2001 indicate that the usage of online titles followed a very similar pattern to that of print versions, regardless of whether it was prior or subsequent to the publication of e-journals.

Interestingly, titles that were included in package deals were only accessed an average of 200 times, while HighWire titles, for which subscriptions had been taken out separately on demand, were accessed 2361 times, i.e. almost 12 times as frequently.

Springer titles had a slightly higher correlation coefficient in Phase 2, namely 0.62, as compared with 0.57 in Phase 1. PDF documents were accessed 28 258 times in 1 year, while the analogous print documents were accessed 1460 times. Prior to the availability of online titles, print titles were used much more often, namely 3054 times. Meanwhile, online Springer titles have come to be used 19 times more often than their corresponding print counterparts, while at the outset of online journal availability the frequency of usage was only 9 times greater. Figure 4 presents a graph of the online and print usages during Phase 1. The concentration of German-language medical speciality publications in the upper triangle of the diagram (indicating above average online usage) is striking.

A comparison of the usage of print HighWire titles with the usage of online accessible ones provides a good example of how precisely the print version usage prior to e-journal availability has been transposed into the electronic information environment: when ranked by amount of usage in print versus online, the order of the seven highest ranking titles is identical (Table 4).

Discussion

Concerning the validity of usage statistics

Firstly, we need to consider to what extent the usage statistics under consideration are valid and to what extent they can be put to useful account in this present study. We must bear in mind that, in order for the figures of access arrived at by means of the three methods (reshelving, error-copies method and online statistics) to permit valid comparison, we need to restrict them to the usage of full text.

The reshelving method applied during the period 1997–1998 is generally recommended for
its comparative reliability. However, it must be remembered that the actual usage is underestimated by 33% (according to the author’s calculations) to 60% (according to Bauer, Note 4. The figure of 60% results from 10 pages copied per article). This method also records the usage of tables of contents, abstracts, news columns etc., which are not at all necessary for the comparison of full text and which would enter into the equation with reversed signs. Instead of having to estimate both these divergences, we will simply assume that they cancel each other out.

The error-copies method employed during the period 1999–2001 assesses the faulty copies of articles from journals and the checked-off reading lists of users. The accordance with the reshelving method at 0.82 seems sufficiently precise, as two periods evaluated with the reshelving method show an accordance rate of 0.84 (the author’s calculations), which is only slightly higher.

The error-copies method has three undeniable advantages: firstly, it is a concealed method, which means that it cannot be manipulated; secondly, it is not very time-consuming; and thirdly, it in no way restricts usage. In contrast to the reshelving method, the error-copies method registers only the usage of full text, as the few copies of cover leaves or tables of contents are not counted. The error-copies method does not register in-house use, but the reading of complete articles has diminished in the course of an observed steep decline in the readership of journals. (According to the most recent estimates of library users, 70% fewer doctors and scientists frequent the library compared with two years ago.) In the final analysis, the three assessment methods can be reduced to full-text access only, thereby being directly comparable.

The enquiry into print usage data is very time-consuming; the compromise between validity/comprehensiveness and staff capacity/pragmatism leads mostly to random sample tests, which require corresponding interpolation. At first glance, this situation appears to be entirely different in the case of e-journals. Apparently, the automatic registration of access information in the publisher’s or the consortium’s main server’s log files seems to provide an impeccable instrument for usage assessment. This data, however, is replete with its own specific pitfalls:

- Frequently used titles could be called up out of the cache of proxy servers and would then not be registered in the log files. (This, however, seems to be a very unlikely situation, considering how very rare multiple use of individual articles is.)
- Journal supplies could be accessed by search robots or spiders, thus greatly enlarging the figures of access.
- There may exist considerable variations in definitions of access, thereby falsifying or encumbering comparison.
- Users may call up one and the same text repeatedly in the course of several seconds or minutes. This phenomenon of redundant multiple accesses at short intervals can grossly swell user statistics, as demonstrated by Berg.

The correspondence between online and print usage is obvious.

<table>
<thead>
<tr>
<th>Journal</th>
<th>Online usage</th>
<th>Print usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Journal of Biological Chemistry</td>
<td>10 622</td>
<td>1387</td>
</tr>
<tr>
<td>Proceedings of the National Academy of Sciences</td>
<td>5 248</td>
<td>775</td>
</tr>
<tr>
<td>Blood</td>
<td>3 467</td>
<td>946</td>
</tr>
<tr>
<td>American Journal of Physiology</td>
<td>1 847</td>
<td>943</td>
</tr>
<tr>
<td>Journal of Cell Biology</td>
<td>1 580</td>
<td>543</td>
</tr>
<tr>
<td>EMBO Journal</td>
<td>1 543</td>
<td>347</td>
</tr>
<tr>
<td>Journal of Experimental Medicine</td>
<td>1 068</td>
<td>91</td>
</tr>
<tr>
<td>Annual Review of Biochemistry</td>
<td>312</td>
<td>34</td>
</tr>
<tr>
<td>Journal of Applied Physiology</td>
<td>278</td>
<td>49</td>
</tr>
<tr>
<td>Journal of Lipid Research</td>
<td>206</td>
<td>71</td>
</tr>
<tr>
<td>Drug Metabolism and Disposition</td>
<td>118</td>
<td>71</td>
</tr>
</tbody>
</table>
According to the ACCELERATE study, the phenomenon of redundant multiple accesses constitutes 28.2% of the total number of accesses. A nearly identical figure of 28.6% was determined from the statistics presented by Blackwell Science (the only publisher, by the way, who provides an IP address as well as times of access), the publishers Blackwell and Elsevier give a figure of 22%, and Academic Press cites one of ‘upwards of 20%’ (personal communication).

Another option for determining the size of this error is supplied by the statistics of HighWire Press, which offers access rates without ‘duplicate usage’. Duplicate usage counts not only individual events of redundant multiple accesses, but all cases of multiple accesses in a university over a longer period of time (2 months). For the Journal of Biological Chemistry, for instance, the resulting ‘unique events’ on average are about 38% below the monthly rate of ‘total usage’. This percentage corresponds easily with the above figures, if 28% is considered as referring to redundant multiple accesses and 10% to intentional accesses by the same IP address or by a different one. (Redundant multiple access by definition refers to individual and short interval access. Intentional multiple access is defined as access to one and the same document by different individuals in multiple sessions.) Observed during the course of 1 year, the unique events fall to below 48.5% of the total usage. The percentage of intentional multiple access events would therefore be around 23% p.a.; the figure of 19% given in the ACCELERATE study compares very well with this.

In view of all these sources of potential error, the counting method utilised with print journals appears to be more reliable than previously assumed. Moreover, its weak points are well known and can therefore be considered during calculations. In contrast to this, the statistics of online access can hardly be evaluated at all, as it is relatively unclear what has been added up. This type of assessment suggests a validity that really is non-existent. A number of assumptions and corrections are required in order to make even halfway-confident statements about the real figures of e-journal access. Even the guidelines of the ICOLC have thus far not been able to eliminate such inconsistencies. Several teams and standardization projects are presently at work on the urgently required specification of access statistics and their standardized assessment.

Discussion of results

Journals published in both print and online versions lost 30.4% of their print usage within approximately 2 years. The total loss of usage of all print-only titles in the same period was somewhat higher, at 45.8%. Rindfleisch identified a reduction of 16–43% p.a., thus confirming the order of magnitude of usage loss. These findings point to the fact that print titles that are not available online suffer the same or a very similar decline in usage as those that are available as combined print and e-journals. This result, which at first glance seems paradoxical, appears to confirm that what is read or purchased is determined primarily by ease of access and that there is a steady tendency to reduce the multiplicity of access modes to a manageable few or ‘users may be compromising quality for convenience’.

The proportion of 1 : 2 and 1 : 1 respectively of online to print usage for the year 1998 found by Schümmer and the author has meanwhile clearly shifted in favour of online usage. As shown in Table 3, the proportion of the online/print usage of the journals of the five publishers that were examined was on average 7.96. However, it must be considered that only the usage of actually offered volumes was counted. This means that the e-journals are automatically at a disadvantage compared with the print journals, as quite naturally there exist more volumes of print journals, as these have a longer history than e-journals. This should be particularly obvious in the case of Academic Press or Blackwell, as only the issues of about 1.5 and 2.5 years respectively were available for analysis. As the usage of the two most recent volumes constitutes only about 40% of the —hypothetical— total usage, this could provide an explanation for the conspicuously low figures of access. This is further supported by the fact that the online/print correlation clearly corresponds to the size of the online inventory. The concurrence increased with the numbers of volumes supplied online (see Table 5).

As for print titles, this error is hardly of any significance, as the most intensely used latest 20
volumes of the titles are available almost without exception, thanks to a demand-oriented purchasing policy. Including approximately 10% of unrecorded in-house usage, the print usage may have been underestimated by a maximum of 15%. By allowing for the subtraction of multiple access events to e-journals (−28%), correction of the print usage (+15%), and standardization to the same number of volumes in supply (see Obst for calculations according to the diagram)\textsuperscript{20}, the data in Table 3 can now be modified as in Table 5.

The result is a ratio of about ten to one between online usage and print usage. This factor is not the same for all publishers, and two clearly distinguishable groupings begin to emerge. While with Academic Press and Elsevier, e-journal usage exceeded print usage only by a factor of 3 or 4, the e-journals of Blackwell, HighWire and Springer were used on average 14.6 times as frequently as the corresponding print journals. Similar ratios were found by Rindfleisch (22 ± 13 or 16 ± 5 respectively with or without \textit{Science})\textsuperscript{17} and Morse (15.3).\textsuperscript{21}

Most striking was the close correlation between the decrease in print usage (Table 2, last column) and the online/print ratio (Table 5, last column). The greater the amount of online usage the greater the print decrease. With a correlation coefficient of 0.85, the decrease in print usage was almost predictable from the online/print ratio.

As the assessment of print usage was identical for all journals, the pronounced differences in the print/online ratio referred to above are likely to be due to specific publisher features. The following reasons could generate such levels of different usage:

\begin{itemize}
  \item The method of assessment for access statistics. For example, a detailed evaluation of log files in the form of session analysis according to Berg\textsuperscript{8} reduces the usage up to 50% (this, however, has already been taken into account in the present study). Apart from that, the access events may have been increased by search robots such as spiders or off-line browsers.
  \item Differing attractiveness of different modes of supply. A variety of factors could cause unpopularity among users: difficulty of access, missing comprehensiveness, bad full-text quality or lack of updates. Conversely, greater attractiveness could be achieved by an upgraded online version (multimedia supplements, updates, search options).
  \item If certain titles of a specific publisher generate an above average count in online access, this can be explained either by disproportionate usage on the part of its present clientele or by an expansion to new groups of users. The first case could apply to Springer’s journals for further professional or vocational training, which in their print versions might have been rarely used as they are mostly privately accessed through memberships. But it is also possible that the disproportionately frequent use of online versions compared with that of print titles could have occurred because they are more easily accessed. An expansion to include new user groups seems plausible, especially in the case of German-language medical Springer journals, which are of interest not only to the medical faculty, but also to health-conscious individuals of non-medical background, or which contain articles found on publishers’ main servers in interdisciplinary full-text searches. This view is supported by the observation, that even locations which have no medical faculty achieve high access rates for these titles.
\end{itemize}

In spite of this, it is doubtful whether these phenomena can explain a fivefold difference in the online/print ratio. Divergent methods of data assessment or log file evaluation are the most probable causes. However, it cannot be overlooked that these data are upheld by the very suppliers who provide the journals.

Costs of online and print usage can be compared only in a limited way, as in any one particular

\begin{table}
\centering
\caption{Proportionate usage and correlation-corrected values.}
\begin{tabular}{lrrr}
\hline
          & Online & Correlation & Ratio  \\
stock &       & online/print & online/print \\
\hline
AP**      & 1.5 year & 0.43       & 4.0  \\
Blackwell & 2.5 year & 0.46       & 9.8  \\
Elsevier  & 5.5 year & 0.72       & 3.2  \\
HighWire**& 5.0 year & 0.76       & 13   \\
Springer  & 4.3 year & 0.62       & 21   \\
Average   &           & 10.2       &      \\
\hline
\end{tabular}
\footnote{the mean number of volumes available at the time of the online usage assessment; **after deduction of multiple-access events according to statements by the publisher.}
\end{table}
year there may be expenses either for the print or for the electronic version, but not for both. Therefore we have not considered the actual costs but rather the prices quoted by the publishers for the purchase of the print and electronic versions respectively (Table 6).

This table clearly shows the difference between print and online purchase. Print usage on average costs 5.4 times as much as online usage. But here, too, we detect pronounced differences among the various publishers. Top among the print journals is Springer with 50.82 € per usage (while HighWire titles cost only a fraction of this). The reason for this seems to be that they are German-language publications. In their annual evaluation rounds faculty staff tend to cling to such titles as seem familiar to them, i.e. German-language publications. Of the e-journals, the Academic Press journals at 15.10 € per usage even surpass some print-only usage costs. This is presumably because, in comparison with other publishers, Academic Press supplies the lowest number of volumes online. HighWire is also the most cost-efficient publisher when it comes to e-only publications: at 0.31 €, the difference from competitors by more than one order of magnitude is even greater than the case of the print only publications.

### Conclusion

The correlation found of 0.60 and 0.67 respectively clearly supports the hypothesis that use of e-journals is very similar to that of print journals, i.e. frequently accessed print titles are also frequently accessed in their online version. The greater the number of volumes available online, the greater the concurrence becomes. Schümmer confirms the results at hand as he came to the conclusion that several print titles in strong demand are also frequently accessed online.13

The results presented here have shown that within 2–3 years the usage of online journals has outstripped that of print titles by a factor of 10 (and in special cases even by a factor of 100), the specific spectrum of usage remaining much the same as when only print journals existed. Subsequent to the results of this and other studies, the following recommendations can be made for the management of e-journals:

1 Package deals seem to have no future. There is much to suggest that the local stock of journals (at least after evaluation and cancellation procedures have been completed) is the result of a conscious purchasing policy that is dictated by demand requirements. As demonstrated by the ACCELERATE study,14 the supply of additional titles had hardly any effect in improving the stock of literature. Online usage per title increased by an order of magnitude as soon as these titles were not sold through package deals but rather individually purchased as a result of careful selection. Consequently, subscription to single, selected titles as well as a pay-per-view system should be considered.
For a sound purchasing policy, the e-journal statistics available at this point are of no more use than the statistics of print usage. The need to catch up in this respect is great and urgent indeed. Publishers’ contracts ought to contain precise instructions regarding the evaluation of log files. As an alternative, an independent body could be entrusted with this evaluation.

The purchase of back files should always be considered, as their usage lies at above average frequency at mostly very low additional cost. This is particularly true of the social sciences, but it applies to the medical field as well.

Are e-journals an economical alternative? The Medical Library pays approximately 10 € for the use of a print journal article, and only one fifth of this for an e-journal article. Including space, system, and staff costs, the Drexel library study came up with 20 € and 2 € respectively.

Electronic journals alter not only the quantity of usage but also its quality. On several levels they introduce changes in behaviour patterns among library clients, physicians and scientists of the medical faculty. The connection between supplier and supply is much less clear with e-journals than it is with print titles. Most users do not realize that the use of e-journals is only made possible through the financial mediation of the library. However, there is no return to the simplicity and security of the printed journal. Customers and libraries are already becoming irreversibly accustomed to the benefits of the e-journal. As journal management remains one of the greatest challenges for libraries, strategic planning as well as professional marketing has to be used to encourage the role of the library as the supplier of this valuable type of information.

References
