COUNTER Online Metrics

The Journal Usage Factor project: results, recommendations and next steps

Peter Shepherd, COUNTER

Table of Contents

	Page
0. Executive summary and next Steps	2
1. Why usage-based measures?	6
2. Who will benefit from the Journal Usage Factor?	7
3. Journal Usage Factor Project: aims and objectives	8
4. Journal Usage Factor Stage 1: market research	8
5. Journal Usage Factor Stage 2: modelling and analysis	14
6. Journal Usage Factor: next steps	20
7. Journal Usage Factor Stage 3: organizational structure	21
8. Further information	21
9. References	21
10. Appendix	22
11. Acknowledgements	22

0. Executive summary and next steps

The widespread availability of reliable usage data for online journals has opened the door to usage-based measures of journal impact, value and status. Since 2002 COUNTER (1) has provided a standard for vendor-generated usage statistics for individual libraries and library consortia, while the MESUR project (2) has demonstrated the potential value of a wide range of usage-based metrics for assessing the impact of journals at a global level. A common, underlying theme in both projects is that usage-based alternatives to citation-based metrics are both desirable and increasingly practical.

While ISI's journal Impact Factors (IFs), based on citation data, have become generally accepted as a valid measure of the quality of scholarly journals, and are widely used by publishers, authors, funding agencies and librarians as measures of journal quality, there are misgivings about an over-reliance on Impact Factor alone in this respect. The availability of the majority of significant scholarly journals online, combined with the availability of credible COUNTER-compliant online usage statistics, raises the possibility of a parallel usage-based measure of journal performance becoming a viable additional metric. Such a metric may be termed 'Journal Usage Factor' (JUF),

The proposed JUF will provide information about the average use of the items in an online journal. Like Impact Factor, it is scale independent. In other words it should be able to be used to compare journals irrespective of their size. To gain widespread acceptance it should be robust and easy to understand.

At the outset of this project, the initial working hypothesis was that JUF would be derived using the calculation in Equation 1 below:

Equation 1: $$\text{JUF} = \frac{\text{Total usage over period x of items published during period y}}{\text{Total items published online during period y}}$$

While this approach would have the seductive advantage of familiarity, as it parallels the calculation used for Impact Factors, a detailed statistical analysis carried out by CIBER in the course of this project (see Appendix A) found it wanting and recommended that a statistically more meaningful approach would be simply to sort the number of downloads for each item used during period *y* and take the middle value (the median) as the JUF, an approach which will be investigated further in the next stage of the project. This recommendation, together with the others made by CIBER, is listed below. These recommendations focus on the statistical aspects of JUF, rather than on organizational or economic issues that have also been addressed in the course of the project and are described in this report, but will now have to be developed in more detail. A valid statistical model is the *sine qua non* of a meaningful Journal Usage Factor and it is important to establish confidence in this before taking the organizational and economic models further.

Key recommendations

Recommendation 1

This study shows that usage data are highly skewed; most items attract relatively low use and a few are used many times. As a result, the use of the arithmetic mean is not appropriate (see Appendix A, pages 8-10)

The Journal Usage Factor should be calculated using the median rather than the arithmetic mean

Recommendation 2

There is considerable variation in the relative use made of different document types and versions (see Appendix A, pages 10 and 11). This means that the usage factor will be affected substantially by the particular mix of items included in a given journal, all other things being equal.

A range of usage factors should ideally be published for each journal: a comprehensive factor (all items, all versions) plus supplementary factors for selected items (e.g. article and final versions).

Recommendation 3

Monthly patterns of use at the item level are quite volatile and usage factors therefore include a component of statistical noise (see Appendix A, page 12)

Journal Usage Factors should be published as integers with no decimal places

Recommendation 4

As a result of this statistical noise, the mean usage factor should be interpreted within intervals of plus or minus 22 per cent (see Appendix A, page 12)

Journal Usage Factors should be published with appropriate confidence levels around the average to guide their interpretation

Recommendation 5

This report shows that relatively short time windows capture a substantial proportion of the average lifetime interest in full journal content (see Appendix A, pages 15-19). Longer windows than 24-months are not recommended (see Appendix A, page 22) and this should be considered a maximum. There is possibly a case for considering a 12-month window (see Appendix A, page 21) but there are counter-arguments here: the impact of publishing ahead of print especially.

The Journal Usage Factor should be calculated initially on the basis of a maximum time window of 24 months. It might be helpful later on to consider a 12-month window as well (or possibly even a 6-month window) to provide further insights.

Recommendation 6

Usage in months 1-12 especially follows different patterns in different subject areas (see Appendix A, pages 15-19).

The Journal Usage Factor is not directly comparable across subject groups and should therefore be published and interpreted only within appropriate subject groupings.

Recommendation 7

Usage factors will tend to inflate across the board year-on-year as a result of many factors, including greater item discoverability through search engines and gateways. Changes to access arrangements (e.g. Google indexing) will have dramatic and lasting effects. The use of a two-year publication window would ameliorate some of these effects by providing a moving average as well as a greater number of data points for calculating the usage factor.

The Journal Usage Factor should be calculated using a publication window of two years

Recommendation 8

The usage factor delivers journal rankings that are comparable in terms of their year-on-year stability with those generated from citation metrics such as the ISI impact factor and SNIP (see Appendix A, pages 25-27)

There seems to be no reason why ranked lists of journals by usage factor should not gain acceptance

Recommendation 9

Usage factors below a certain threshold value (perhaps 100 but research is needed on a larger scale to explore this further) are likely to be inaccurate due to statistical noise (see Appendix A, pages 30-32). The size of the journal should also be taken into account.

Small journals and titles with less than 100 downloads per item are unsuitable candidates for Journal Usage Factors: these are likely to be inaccurate and easily gamed,

Recommendation 10

The usage factor does not appear to be statistically associated with measures of citation impact (see Appendix A, pages 35-36)

The Journal Usage Factor provides very different information from the citation Impact Factor and this fact should be emphasised in public communications.

Recommendation 11

Attempts to game the usage factor are highly likely. CIBER's view is that the real threat comes from software agents rather than human attack. The first line of defence has to be making sure that COUNTER protocols are robust against machine attack. The analysis in this report (see Appendix A, pages 39-44) suggests that a cheap and expedient second line of defence would be to develop statistical forensics to identify suspicious behaviour, whether it is human or machine in origin.

Further work is needed on usage factor gaming and on developing robust forensic techniques for its detection

Recommendation 12

Although the scope of this study was to consider the Journal Usage Factor only, future work could look at the other indicators that mimic other aspects of online use, such as a 'journal usage half-life' or a 'reading immediacy index'.

Further work is needed to broaden the scope of the project over time to include other usage-based metrics

Journal Usage Factor next steps: Stage 3

While the recommendation to use the median usage value for a journal, rather than the arithmetical average determined in Equation 1, somewhat reduces the data that has to be gathered to calculate the Journal Usage Factor , the data and metadata requirements remain stringent and will have to be specified in detail. Likewise, the processes, organizational and economic models required to implement JUF in a sustainable way will have to be developed.

Objectives of Stage 3

Stage 3 of the project builds on the key recommendations listed above. Work has already begun on Stage 3 and the following objectives have been set:

1. **Preparation of a draft Code of Practice for the Journal Usage Factor,** consistent with the COUNTER standards, which will cover: definitions of terms, data and metadata requirements, the article types to be counted, the article versions to be counted, and the way in which JUF is to be recorded and reported. This draft Code of Practice will be published in 2011.
2. **Further testing of the recommended methodology for calculating Journal Usage Factor:** in addition to the median value, other usage-based metrics will be tested. Publishers will be invited to participate in tests using the draft Code of Practice.
3. **Investigation of appropriate, resilient subject taxonomy for the classification of journals.** The currently available journal classification systems are inadequate and out of date. An alternative will be sought.
4. **Exploration of the options for an infrastructure to support the sustainable implementation of JUF.** If the JUF is to be credible, and usable by the key target groups (researchers, publishers, librarians, research funding agencies) an appropriate organizational structure, an independent audit and some form of central registry will be required. These and other infrastructure issues will be addressed in Stage 3.
5. **Investigation into the feasibility of applying the Usage Factor concept to other categories of publication**, such as online databases, books and reference works

It is envisaged that Stage 3 will be completed by the end of March 2012.

Organizational Structure of Stage 3

At the end of Stage 2 of the JUF project UKSG transferred responsibility for the project to COUNTER, which will now take it forward. To ensure ongoing equanimity in the supervision of the project, it will continue to have two Co-Chairs –a Publisher

and a Librarian – who will report to the COUNTER Executive Committee. The organizational structure of Stage 3 is:

Co-Chairs: Jayne Marks (SAGE Publishing, USA) and Hazel Woodward (Cranfield University, UK)

COUNTER Executive Committee Chair: David Sommer (David Sommer Consulting, UK)

JUF Stage 3 Project Director: Peter Shepherd (COUNTER)

International Advisory Board: in the course of being appointed

1. Why usage-based measures?

A growing body of reliable journal usage statistics

The burgeoning availability of reliable usage data for online journals has opened the door to usage-based measures of journal impact, value and status. Since 2002 COUNTER (1) has provided a standard for vendor-generated usage statistics for individual libraries and library consortia,

A complement to citation-based measures

While ISI's journal Impact Factors (IFs), based on citation data, have become generally accepted as a valid measure of the impact and status of scholarly journals, and are widely used by publishers, authors, funding agencies and librarians as measures of journal quality, there are misgivings about an over-reliance on Impact Factor alone in this respect (3). The availability of the majority of significant scholarly journals online, combined with the availability of increasingly credible COUNTER-compliant online usage statistics, raises the possibility of a parallel usage-based measure of journal performance becoming a viable additional metric - the Journal Usage Factor (JUF)

Journal Impact Factors: strengths and weaknesses

Strengths:

- well-established
- widely recognised, accepted and understood
- difficult to defraud
- endorsed by funding agencies and scientists
- simple and accessible quantitative measure
- independent
- global
- journals covered are measured on the same basis
- comparable data available over a period of decades
- broadly reflect the relative scientific quality of journals in a given field
- its faults are generally known

Weaknesses:

- bias towards US journals; non-English language journals poorly covered
- optimized for biomedical sciences, work less well in other fields
- can be manipulated by, e.g., self-citation
- over-used, mis-used and over-interpreted

- is an average for a journal; provides no insight into individual articles
- formula is flawed; two year time window too short for most fields
- can only be used for comparing journals within a field
- only a true reflection of the value of a journal in pure research fields
- impact of practitioner-oriented journal is understated
- under-rates high-quality niche journals
- does not cover all fields of scholarship
- over-emphasis on IF distorts the behaviour of authors and publishers
- time-lag before IFs are calculated and reported; new journals have no IF
- emphasis on IF masks the great richness of citation data

Journal Usage Factor: providing a new perspective

JUF is a complementary measure that will compensate for the weaknesses of Impact Factors in several important ways:

- JUFs will be available for a much larger number of journals
- coverage of all fields of scholarship that have online journals
- impact of practitioner-oriented journals is better reflected in usage
- usage is recorded and reported immediately upon publication of an article
- availability of JUF will reduce the current over-emphasis of IFs
- authors would welcome a usage-based measure for journals

Journal Usage Factor: a simple, transparent calculation

One strength of citation-based Impact Factors is that the calculation used to derive them is straightforward, transparent and is based on two sets of data that are readily available for journals: the number of items published in the journal and the number of citations made to the journal. It would be advantageous if the calculation of Journal Usage Factor could be similarly straightforward and transparent and one of the main objectives of the JUF project was to explore approaches that meet these criteria.

Journal Usage Factor: readily available data

There are now over 15,000 full-text online journals providing COUNTER compliant usage data, in many cases stretching back to 2002.

2. Who will benefit from the Journal Usage Factor?

There are four groups that would benefit most from the introduction of a Journal Usage Factor. They are:

a. **Authors,** especially those in practitioner-oriented fields, where citation-based measures understate the impact of journals, as well as those in areas outside the core STM fields of pure research, where coverage of journals by citation-based measures is weak.
b. **Publishers,** especially those with large numbers of journals outside of the core STM research areas, where there is no reliable, universal measure of journal impact, because citation-based measures are either inadequate or non-existent for these fields

c. **Librarians,** when deciding on new journal acquisitions, have no reliable, global measures of journal impact for fields outside the core STM research fields. They would use usage-based measures to help them prioritise journals to be added to their collections.
d. **Research Funding Agencies,** who are seeking a wider range of credible, consistent quantitative measures of the value and impact of the outputs of the research that they fund.

3. Journal Usage Factor project: aims and objectives

The overall aim of this project is to explore how online journal usage statistics might form the basis of a new measure of journal impact and quality, the Journal Usage Factor. The specific objectives of the project were: to examine the ways in which journal quality is currently assessed; to assess whether the JUF would be a statistically meaningful measure; whether it would be accepted by researchers, publishers, librarians and research institutions; whether it would be statistically credible and robust; whether there is an organizational and economic model for its implementation that would be acceptable to the major stakeholder groups.

The project was executed in two Stages, between 2007 and 2011. Stage 1 focused on market research into the overall feasibility and acceptability of the Journal Usage Factor in principle. Stage 2 focussed on modelling and analysis, in which real usage data from COUNTER-compliant publishers was used to test the formula for calculation of JUF, as well as the processes for doing so on a sustainable, ongoing basis.

4. Journal Usage Factor Stage 1: market research

Summary and key findings

The objective of Stage 1 of the project, carried out in 2007 and 2008, was to obtain an initial assessment of the feasibility of developing and implementing JUFs. This was done by conducting a survey in two Phases. The first was series of in-depth telephone interviews with a total of 29 authors/editors, librarians and publishers. The second was a web-based survey in which almost 1400 authors and 155 librarians participated. The feedback obtained helped determine not only whether JUF is a meaningful concept with the potential to provide additional insights into the value and quality of online journals, but also how it might be implemented. The results obtained also provided useful pointers for the topics to be explored further in subsequent stages of the project. The apparent eagerness of senior executives to take part in the interviews and the large number of responses to the web survey indicated a high level of interest in journal quality measures in general and in the JUF concept in particular.

Based on these results it appears that it would not only be feasible to develop a meaningful Journal Usage Factor, but that there is broad support for its implementation. The main conclusions that were drawn from this part of the survey were:

• the majority of publishers were supportive of the JUF concept, appeared to be willing, in principle, to participate in the calculation and publication of JUFs, and were prepared to see their journals ranked according to JUF
• there was a diversity of opinion on the way in which JUF should be calculated, in particular on how to define the following terms: 'total usage', 'specified usage period', and 'total number of articles published online'. The subsequent tests with real usage data in Stage 2, described below helped refine the definitions for these terms.
• there was not a significant difference between authors in different areas of academic research on the validity of journal Impact Factors as a measure of quality
• the great majority of authors in all fields of academic research would welcome a new, usage-based measure of the value of journals
• JUF, were it available, would be a highly ranked factor by librarians, not only in the evaluation of journals for potential purchase, but also in the evaluation of journals for retention or cancellation
• publishers were, on the whole, unwilling to provide their usage data to a third party for consolidation and for calculation of JUF. The majority appeared to be willing to calculate JUFs for their own journals and to have this process audited. This was generally perceived as a natural extension of the work already being done for COUNTER. While this may have implications for systems, these were not seen as being problematic.
• COUNTER was on the whole trusted by librarians and publishers and was seen as having a role in the development and maintenance of the JUF

Feedback from Librarians

The results, presented in Table 1, below showed that, without Journal Usage Factor, 'feedback from library users' was the most important consideration in the decision to purchase journals. Next in the list came price, followed by the reputation or status of the publisher and then Impact Factor. When the JUF was introduced to the mix, librarians ranked it second in order of importance. While one might not expect a JUF to supplant user feedback, it is significant that JUF was ranked ahead of price, IF and the reputation or status of the publisher.

Table 1: Librarians' views on the relative importance of key factors in the process of evaluating journals for potential purchase [in rank order]

Ranking without JUF	Ranking with JUF
1. Feedback from library users	1. Feedback from library users
2. Price	2. Journal Usage Factor
3. Reputation/status of publisher	3. Price
4. Impact Factor	4. Impact Factor
	5. Reputation/status of publisher

When it came to evaluating journals for retention or cancellation, librarians were now able to consider usage and cost per download statistics in addition to the factors listed previously. As the results in Table 2 indicate, feedback from library users remained the foremost consideration, but it is interesting to note that usage was ranked second in importance ahead of price and cost per download. IF and the reputation or status of the publisher appeared to be relatively unimportant.

When JUF was presented as an option, the re-ranked list in Table 2 shows that librarians perceived it to be important, ranking it third behind feedback from library users and usage. JUF is ranked ahead of price and cost per download. It is noteworthy that librarians thought a JUF could be more important than IF.

Table 2: Librarians' views on the relative importance of key factors in the process of evaluating new journals for retention or cancellation [in rank order]

Ranking without JUF	Ranking with JUF
1. Feedback from library users	1. Feedback from library users
2. Usage	2. Usage
3. Price	3. Journal Usage Factor
4. Cost per download	4. Price
5. Impact Factor	5. Cost per download
6. Reputation/status of publisher	6. Impact Factor
	7. Reputation/status of publisher

Feedback from authors

The main aim of this part of the survey was twofold:

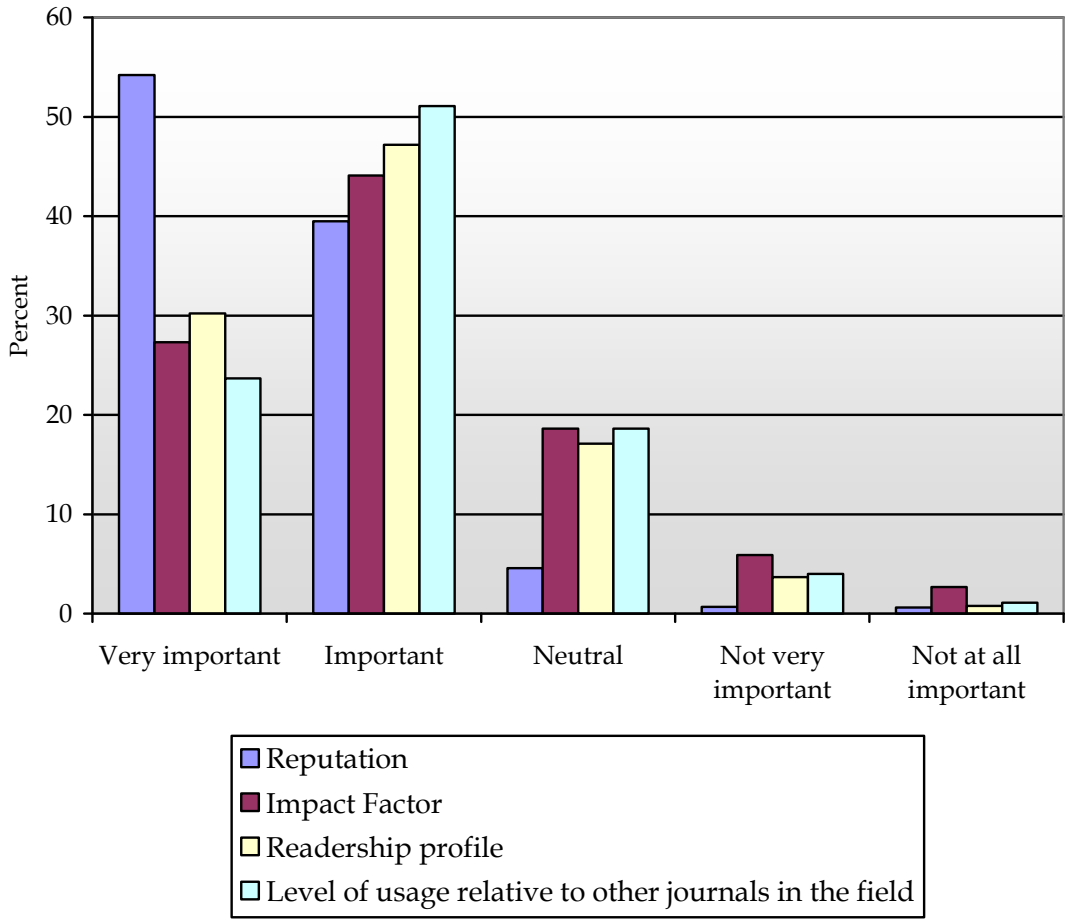
- To discover what academic authors thought about the measures that are currently used to assess the value of scholarly journals (notably IFs)
- To gauge the potential for usage-based measures

A total of 1394 academic authors participated in the study. Authors have a number of factors to consider when deciding which journal to submit their work to for publication. The survey set out to understand where Impact Factor fits alongside other factors that are known to be important to authors.

The results presented in Figure 1 below show that a journal's **reputation** is *the* most important factor in authors' decision-making process. Authors want their work to be read by their peers so it is not surprising that a journal's **readership profile** ranks

second overall in terms of relative importance. Clearly a journal's **Impact Factor** plays an important role in the majority of authors' deliberations about where to publish, but it appears to be a supporting rather than a lead role. The results indicate that authors discern a clear distinction between a journal's reputation and its Impact Factor. Finally, a journal's **level of usage** relative to other journals in the field is shown to be a significant factor. This recognition by academic authors of the importance of a journal's level of usage provides encouragement for the development of a usage-based quantitative measure.

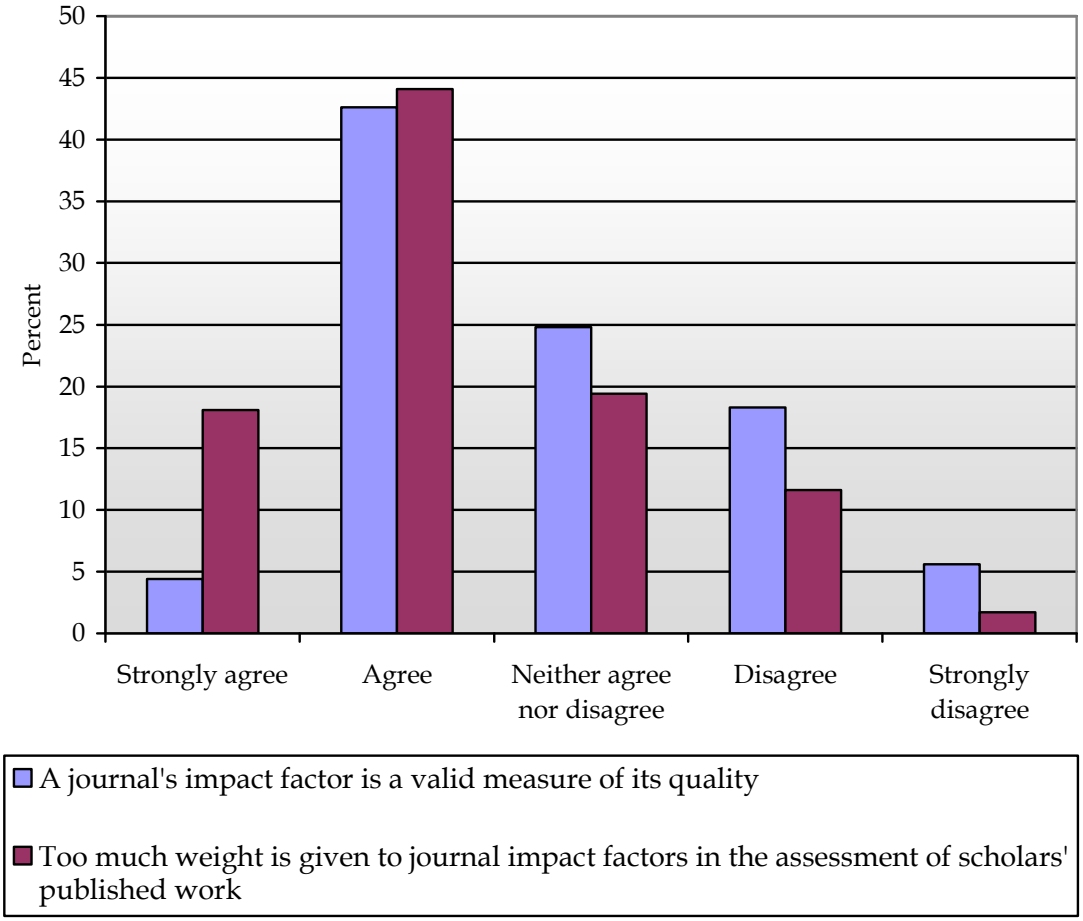
Figure 1: Academic authors' views on the relative importance of four key aspects of journal publishing when considering where to submit their work for publication



Nearly half of the academic authors surveyed believe a journal's IF to be a valid measure of its quality. The data presented in Figure 2 indicate that this endorsement was not overwhelming: whereas 47% of authors either strongly agreed or agreed that Impact Factor is a valid measure of quality, 24% either strongly disagreed or disagreed, and 25% took a neutral stance.

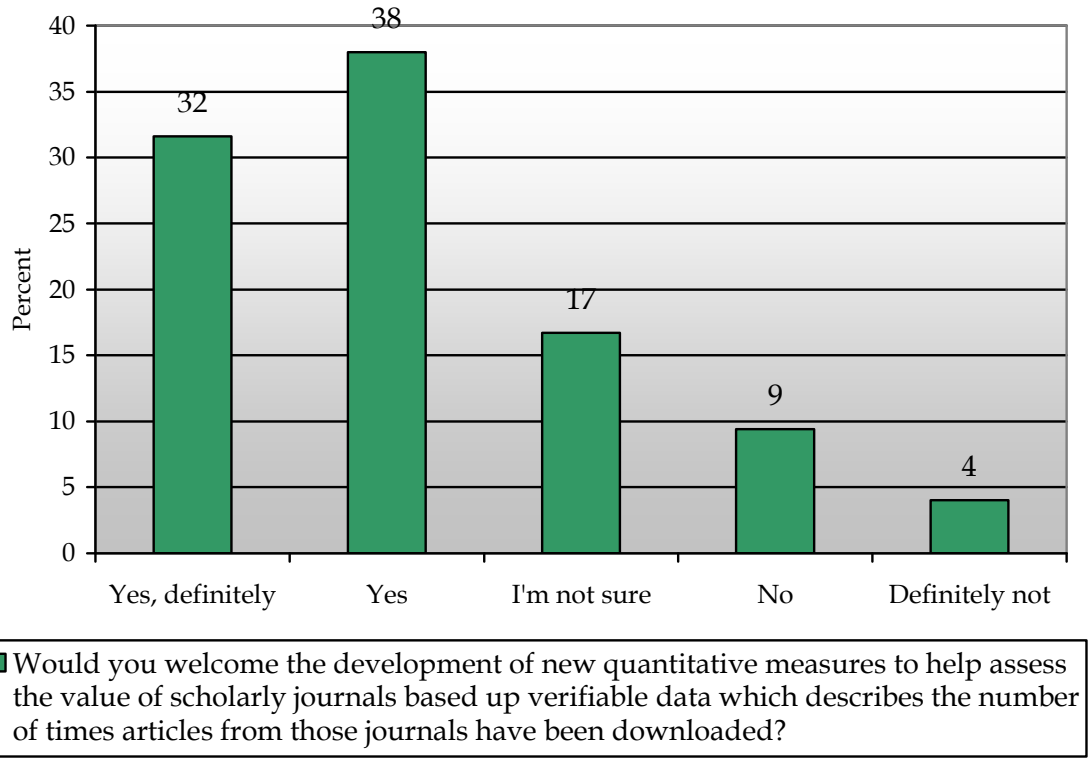
Overall there was a higher level of agreement with the following statement: too much weight is given to journal IFs in the assessment of scholars' published work. 62.5% of academic authors either strongly agreed or agreed that this is the case, compared to just 13% that either disagreed strongly or disagreed. 19% had no particular opinion either way.

Figure 2: Academic authors' views on the value and use of journal Impact Factors



Authors were then asked the following question: *Would you welcome the development of new quantitative measures to help assess the value of scholarly journals based upon verifiable data which describes the number of times articles from those journals have been downloaded?* The pattern of responses, presented in Figure 3, is clearly positive. 70% of academic authors replied 'yes, definitely' or 'yes' in response to the question.

Figure 3: Proportions of academic authors who would welcome a new measure for the assessment of the value of scholarly journals based on article downloads



Principal conclusions of Stage 1

Impact Factor: IF, for all its faults, is entrenched, accepted and widely used. There is a strong desire on the part of authors, librarians and most publishers to develop a credible alternative to IF that will provide a more universal, quantitative, comparable measure of journal value. It is generally acknowledged that no such alternative currently exists, but that usage data could be the basis for such a measure in the future. 70% of authors surveyed would welcome a new, usage-based measure of the value of scholarly journals.

All authors and librarians interviewed thought that Usage Factor would be helpful in assessing the value, status and relevance of a journal. These results were confirmed by the much larger sample of authors and librarians in the web survey. The majority of the publishers also thought it would be useful, but their support would depend on their confidence in the basis for the JUF calculation .

Ranking journals by JUF: While the great majority of authors were in favour of ranking journals by JUF, there was less unanimity among the publishers. Indeed the publisher responses, both positive and negative, tended to be qualified. The majority were positive, but need to be convinced that the JUF calculation would be robust and fair. The minority who were negative appeared to accept that such rankings are going to happen in any event and they would rather it is done by an organization that they

trust. Librarians indicated that, if JUF were available, it would become the second most important factor (after 'feedback from library users') in decisions in the purchase of new journals, while it would be the third most important factor (after 'feedback from library users' and 'usage') in retention/cancellation decisions.

Organizations that could compile and comment on JUF data: there is no existing organization which commands the confidence of both librarians and publishers **and** has the capability to compile/comment on JUF data. Librarians, on the whole, do not have sufficient confidence in publisher-only organizations and publishers, on the whole, do not have sufficient confidence in librarian-only organizations, to fill this role. Indeed, it may require a partnership between organizations. The type of organization required will depend on the role to be filled. If, for example, publishers were to be responsible for the consolidation and calculation (audited) of JUFs, a much smaller central organization would be required than if it were to be responsible for the consolidation of usage data, calculation of UFs and publication of JUFs.

The majority of publishers appear to be willing, in principle, to calculate and publish JUFs for their journals, according to an agreed international standard and appreciate that there would be benefits to them in doing so. Some publishers are more reluctant than others, but would participate if UF were defined and implemented in a way that is acceptable to the market.

In summary, there is significant support, even among established publishers whose journals perform well in IF rankings, for the development and implementation of JUFs. Having said that, this survey has brought into focus a number of structural questions that will have to be dealt with if JUFs are to be credible

5. Journal Usage Factor Stage 2: modelling and analysis

a. Initial modelling and analysis

This part of the project was carried out by Frontline GMS and John Cox Associates.

Based on the results of Stage 1, UKSG, RIN (UK Research Information Network), ALPSP (Association of Learned and Professional Society Publishers). the International STM Publishers Association and a group of publishers decided to fund a further, Stage 2 study. The overall aim of Stage 2 was to assess the viability of JUF as a reliable, implementable, cost-effective tool for assessing the relative status and value of journals by testing each of the individual elements in Equation 1 below using real publisher usage data from a range of vendors.

Equation 1: $$\text{JUF} = \frac{\text{Total usage over period x of items published during period y}}{\text{Total items published online during period y}}$$

Test usage data for 150,000 articles from 326 journals, covering five broad subject areas (Engineering, Humanities. Medicine and Life Sciences, Physical Sciences, and Social Sciences) was obtained from seven publishers (ACS

Publications, Emerald, IOP Publishing, Nature Publishing Group, OUP, Sage and Springer), covering the publication years 2006 -2009.

Publishers were asked to differentiate between different versions of articles, i.e. the Version of Record (VoR), Accepted Version and Proof (4), this proved to be difficult for some publishers, who do not make the distinction when replacing an earlier version of an article with the Version of Record. Publishers were also asked to classify items within the journal as 'Article' or 'Non-article' content and to exclude the standing matter (such as cover pages, contents, indexes, acknowledgements etc.). This also proved to be difficult for some publishers due to the complex way in which they label items; some publishers had in excess of 300 item types. All were able to exclude the standing matter, but for some the contractor had to accept 'all content' rather than classified data.

The precise selection of journals for each subject area was agreed with each participating publisher. The intention was to form a balanced range of around 40-50 titles for each of the five broad subjects. In reality, however, the number of journals in each broad subject was as follows: engineering, 38; humanities, 35; medicine and life sciences, 102; physical sciences; 32; social sciences, 119. This was due to the participating publisher's lists and disciplines selected. Many of the publishers publish on behalf of learned societies, and some publishers chose to exclude society journals to avoid a lengthy process of asking permission for each journal to be included.

The usage data collected from the participating publishers provided coverage from 2006 throughout 2009; full data for 2006 was only available from one publisher, and for 2007 was only available from some publishers. This did, however, allow JUF calculations for a range of publication periods.

Evaluation of the JUF variables

The effects on JUF of four variables were analysed in the course of the study. These variables were: content type (all content vs articles only); article version (accepted article, proof, version of record); publication period; usage period.

Content type

The JUFs for all content and for articles only were compared, and evaluated in the context of their practical implementation. Little significant difference in JUFs was observed between all content and articles only in the Humanities, Physical Sciences and Business & Management. In the Social Sciences, the JUFs were lower in the articles-only category, indicating that readers made considerable use of non-article content. In Medicine & Life Sciences, and in the sub-set of Clinical Medicine, JUFs were higher in the articles-only calculation, indicating that readers are much more concerned to use articles than other editorial content. No firm conclusions could be drawn in Engineering, as the JUFs fluctuated from period to period. It is clear that non-article content is relevant and is used across the disciplines, though much less so in Medicine & Life Sciences.

Item type control is difficult to manage. Using all content (i.e. all editorial content including articles, editorials, book reviews etc, but not standing matter

such as editorial board lists, subscription and permissions details etc) reduces the likelihood of item misdescription by eliminating the need for detailed categorisation, and reduces the impact on publishers. Editorial matter is published for a purpose, and its usage forms part of the usage of the journal as a whole. Even with the adoption of all content, publishers will have to adhere strictly to the specification and avoid extraneous items such as standing matter from creeping in.

For consistency across all disciplines, the balance of advantage appears to lie with a JUF based on all content, as providing a better, more robust metric than one based solely on articles.

Further research and testing on a wider range of journals, across more disciplines, will be necessary in order to confirm these conclusions.

Article version

In view of the inconsistencies among publishers in their approaches to differentiating between versions of articles, as well as the desirability of capturing usage as soon as an article appears online, it was decided that, for the purposes of this project, the balance of advantage lies with including all versions in the JUF metric. This approach not only enables complete usage to be captured in the JUF, but also minimises problems of data accuracy. This issue can be revisited once publishers adopt a consistent policy on article version control, based, for example, on the recommendations of the NISO/ALPSP Technical Working Group on Journal Article Versions (4).

Publication period

In considering whether to recommend a one-year or two-year publication period, three factors were taken into account:

- The Impact Factor is based in a publication period of one year, with citations measured in the following two years. Whether the JUF should follow the same structure as the IF is a matter of preference.

- The data demonstrated that there were occasionally unexplained peaks or troughs in usage. A longer publication period would have a 'smoothing' effect on the JUF, to reduce the impact of such usage events;

- A two-year publication period would reduce the effect on the JUF of early publication where it is offered by the publisher, and also 'smooth' the effect of the tapering usage at the end of the usage period.

It was decided that a two-year publication period provides consistency and a smoothing effect that will provide a more reliable metric than one based on one year only. It is recommended that a two year publication period be adopted.

Usage period

Four usage periods were considered: 1-12 months, 1-24 months, 13-24 months, and 13-36 months after the month of publication. It was agreed that

relying on usage in the periods 13-24 months and 13-36 months would not be desirable, for the following reasons:

- Delaying the capture of usage data until 12 months after publication excludes usage that takes place immediately the article becomes available. It was apparent from the data that usage in the first few months is substantial, and reflects the importance of timely access to researchers, particularly in STM disciplines. To ignore this usage would be to base the JUF on incomplete usage and seriously distort the result;

- By definition, the publication period would be some years old – e.g. 2007 publication and usage in 2008-09 would result in a JUF being available well into 2010, or a 2007-08 publication period and usage in 2009-10 would produce a JUF in 2011. The resulting JUF would be historical, rather than current, and devalue the metric.

In evaluating the advantages and disadvantages of usage periods of one and two years immediately after publication (i.e. usage in 1-12 and 1-24 months after the month of publication), it is considered that a one-year period suffers from the tapering effect of usage of articles published later in the year. In order to provide a more reliable base of usage data, a two year period is considered to be preferable, and the recommendation is to adopt a usage period 1-24 months after the month of online publication.

b. Detailed statistical analysis

The final phase of the project was a more detailed statistical analysis of the test usage data and an examination of the ctirical properties of the Journal Usage Factor, carried out by CIBER, which was designed to validate the conclusions and recommendations of the initial statistical analysis, specifically in the following areas: content type, article version, publication period and usage period.

Since the inception of this project, the working assumption has been that JUF should be calculated by the generic formula in Equation 1 below:

Equation 1 $$\text{JUF} = \frac{\text{Total usage over period x of items published during period y}}{\text{Total items published online during period y}}$$

An alternative approach, and one which is recommended as a result of this study, is to simply sort the number of downloads for each item used during period *y* and take the middle value (the median) as the JUF(se is counted in units of 12 months from the date of online publication of each article, not on a calendar year basis as is the case for the ISI Impact Factor).

This phase of the project was designed to answer the following key questions:

- how should the Journal Usage Factor be calculated and presented?
- what are the usage characteristics of different article types (e.g. original research articles, short communications, editorial material, etc.)
- what are the usage decay rates of different article types and versions? (See also the large-scale analysis of online journal usage over time in The STM Report: an overview of scientific and scholarly journal publishing (5))
- what is the most appropriate time window *(x)* for measuring use?
- what is the most appropriate publication period (*y*) for constructing the Journal

Usage Factor?

- how stable is the Journal Usage Factor over time: can it be used to generate meaningful league tables of journal use?
- what is the relationship, if any, between the Journal Usage Factor and measures of citation impact?
- to what extent could the Journal Usage Factor be gamed, either by humans or machines, and are there digital signatures associated with such attempts to cheat the system?

The full report of the detailed CIBER analysis may be found in Appendix A of this document. Summarised below are its **key recommendations**:

Recommendation 1

This study shows that usage data are highly skewed; most items attract relatively low use and a few are used many times. As a result, the use of the arithmetic mean is not appropriate (see Appendix A, pages 8-10)

The Journal Usage Factor should be calculated using the median rather than the arithmetic mean

Recommendation 2

There is considerable variation in the relative use made of different document types and versions (see Appendix A, pages 10 and 11). This means that the usage factor will be affected substantially by the particular mix of items included in a given journal, all other things being equal.

A range of usage factors should ideally be published for each journal: a comprehensive factor (all items, all versions) plus supplementary factors for selected items (e.g. article and final versions).

Recommendation 3

Monthly patterns of use at the item level are quite volatile and usage factors therefore include a component of statistical noise (see Appendix A, page 12)

Journal Usage Factors should be published as integers with no decimal places

Recommendation 4

As a result of this statistical noise, the mean usage factor should be interpreted within intervals of plus or minus 22 per cent (see Appendix A, page 12)

Journal Usage Factors should be published with appropriate confidence levels around the average to guide their interpretation

Recommendation 5

This report shows that relatively short time windows capture a substantial proportion of the average lifetime interest in full journal content (see Appendix A, pages 15-19). Longer windows than 24-months are not recommended (see Appendix A, page 22) and this should be considered a maximum. There is possibly a case for considering a

12-month window (see Appendix A, page 21) but there are counter-arguments here: the impact of publishing ahead of print especially.

The Journal Usage Factor should be calculated initially on the basis of a maximum time window of 24 months. It might be helpful later on to consider a 12-month window as well (or possibly even a 6-month window) to provide further insights.

Recommendation 6

Usage in months 1-12 especially follows different patterns in different subject areas (see Appendix A, pages 15-19).

The Journal Usage Factor is not directly comparable across subject groups and should therefore be published and interpreted only within appropriate subject groupings.

Recommendation 7

Usage factors will tend to inflate across the board year-on-year as a result of many factors, including greater item discoverability through search engines and gateways. Changes to access arrangements (e.g. Google indexing) will have dramatic and lasting effects. The use of a two-year publication window would ameliorate some of these effects by providing a moving average as well as a greater number of data points for calculating the usage factor.

The Journal Usage Factor should be calculated using a publication window of two years

Recommendation 8

The usage factor delivers journal rankings that are comparable in terms of their year-on-year stability with those generated from citation metrics such as the ISI impact factor and SNIP (see Appendix A, pages 25-27)

There seems to be no reason why ranked lists of journals by usage factor should not gain acceptance

Recommendation 9

Usage factors below a certain threshold value (perhaps 100 but research is needed on a larger scale to explore this further) are likely to be inaccurate due to statistical noise (see Appendix A, pages 30-32). The size of the journal should also be taken into account.

Small journals and titles with less than 100 downloads per item are unsuitable candidates for Journal Usage Factors: these are likely to be inaccurate and easily gamed,

Recommendation 10

The usage factor does not appear to be statistically associated with measures of citation impact (see Appendix A, pages 35-36)

The Journal Usage Factor provides very different information from the citation Impact Factor and this fact should be emphasised in public communications.

Recommendation 11

Attempts to game the usage factor are highly likely. CIBER's view is that the real threat comes from software agents rather than human attack. The first line of defence has to be making sure that COUNTER protocols are robust against machine attack. The analysis in this report (see Appendix A, pages 39-44) suggests that a cheap and expedient second line of defence would be to develop statistical forensics to identify suspicious behaviour, whether it is human or machine in origin.

Further work is needed on usage factor gaming and on developing robust forensic techniques for its detection

Recommendation 12

Although the scope of this study was to consider the Journal Usage Factor only, future work could look at the other indicators that mimic other aspects of online use, such as a 'journal usage half-life' or a 'reading immediacy index'.

Further work is needed to broaden the scope of the project over time to include other usage-based metrics

6. Journal Usage Factor: next steps

While the recommendation to use the median usage value for a journal, rather than the arithmetical average determined by Equation 1, somewhat reduces the data that has to be gathered to calculate the Journal Usage Factor, the data and metadata requirements remain stringent and will have to be specified in detail. Likewise, the processes, organizational and economic models required to implement JUF in a sustainable way will have to be developed.

Objectives for Stage 3

Stage 3 of the project builds on the key recommendations of Stage 2. Work has already begun and the following objectives have been set:

1. **Preparation of a draft Code of Practice for the Journal Usage Factor,** consistent with the COUNTER standards, which will cover: definitions of terms, data and metadata requirements, the article types to be counted, the article versions to be counted, and the way in which JUF is to be recorded and reported. This draft Code of Practice will be published in September 2011.
2. **Further testing of the recommended methodology for calculating Journal Usage Factor:** publishers will be invited to participate in tests using the draft Code of Practice.

3. **Investigation of an appropriate, resilient subject taxonomy for the classification of journals.** The currently available journal classification systems are inadequate and out of date. An alternative will be sought.
4. **Exploration of the options for an infrastructure to support the sustainable implementation of JUF.** If the JUF is to be credible, and usable by the key target groups (researchers, publishers, librarians, research funding agencies) an appropriate organizational structure, an independent audit and some form of central registry will be required. These and other infrastructure issues will be addressed in Stage 3.
5. **Investigate the feasibility of applying the Usage Factor concept to other categories of publication**

It is envisaged that Stage 3 will be completed by the end of March 2012.

7. Organizational Structure of Stage 3

At the end of Stage 2 of the JUF project UKSG transferred responsibility for the project to COUNTER, which will take it forward. To ensure a continued equanimity in the supervision of the project, it will continue to have two Co-Chairs –a Publisher and a Librarian – who will report to the COUNTER Executive Committee. The organizational structure of Stage 3 is:

Co-Chairs: Jayne Marks (Sage) and Hazel Woodward (Cranfield University)

COUNTER Executive Committee Chair: David Sommer (David Sommer Consulting)

JUF Stage 3 Project Director: Peter Shepherd (COUNTER)

International Advisory Board: to be appointed

8. Further information

Further information on the JUF project may be found on the UK Serials Group website at: http://www.uksg.org/usagefactors and on the COUNTER website at: http://www.projectcounter.org/news.html

9. References

1. Project COUNTER**:** http://www.projectcounter.org/index.html
2. MESUR: http://www.mesur.org/MESUR.html
3. Amin, M and Mabe, M, Impact Factors: Use and Abuse, Perspectives in Publishing, Elsevier, 2000 http://www.elsevier.com/framework_editors/pdfs/Perspectives1.pdf
4. NISO/ALPSP Technical Working Group on Journal Article Versions: http://www.niso.org/publications/rp/RP-8-2008.pdf
5. Mark Ware and Michael Mabe, The STM Report: an overview of scientific and scholarly journal publishing (2009), p59 www.**stm**-assoc.org/2009_10_13_MWC_**STM_Report**.pdf

10. Appendix

Appendix A: The Journal Usage Factor: exploratory data analysis (CIBER Research Limited)

11. Acknowledgements

a. Thanks go to the following sponsors who provided the financial support for this project:-

- GOLD

UK Serials Group Research Information Network

- SILVER
  - ALPSP
  - American Chemical Society
  - STM
  - Nature Publishing Group
  - Springer

b. Thanks go to the following publishers who provided usage data for the analysis carried out in Stage 2:

  - American Chemical Society
  - Emerald
  - IOP
  - Nature Publishing
  - OUP
  - Sage
  - Springer

c. Thanks go to John Cox Associates and to Laura Cox of Frontline GMS, who collected, consolidated and analysed the usage data provided by publishers, and to Ian Rowlands of CIBER, who carried out a further detailed statistical analysis of the data