The journal usage factor: exploratory data analysis

STAGE 2 FINAL REPORT

27 May 2011

CIBER Research Limited

CIBER Research Limited

Contents

PART ONE
Executive summary

Background to this report

Understanding the usage factor

CIBER Research Limited

The proposed Journal Usage Factor (the JUF) provides information about the average use of the items in an online journal. Like ISI's citation impact factor, it is scale independent. In other words you should be able to use it to compare journals irrespective of their size. To gain widespread acceptance, it should be robust and easy to understand.

The JUF is given by the generic formula:

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

Alternatively, and as recommended in this study, we simply sort the number of downloads for each item used during period *y* and take the middle value (the median).

Use is counted monthly from the date of online publication, not on a rolling calendar year basis as is the case for the ISI impact factor.

In the first stage of this project, Laura and John Cox collated monthly COUNTER-compliant usage data for 326 titles from seven publishers. The publication years covered were 2005 to 2009 and a broad range of subjects were covered. They successfully demonstrated the feasibility of creating a coherent set of usage factors from diverse publishers, although several practical issues emerged, notably issues over the different schema used to record document type and version.

This report takes the project into a second stage: it takes a more detailed look at the critical properties of the usage factor.

CIBER has organised the data from stage one into a single item-level database and data mined that extensively using the Statistical Package for the Social Sciences (SPSS v.17).

We set out to get answers to some key questions:

- how should the usage factor be calculated and presented?
- what are the usage characteristics of different document types (e.g. original research articles, short communications, editorial material, etc.)
- what are the usage decay rates of different document types and versions?
- what is the most appropriate time window *(x)* for measuring use?
- what is the most appropriate publication period (*y*) for constructing the usage factor?
- how stable is the usage factor over time: can it be used to generate meaningful league tables of journal use?
- what is the relationship, if any, between the usage factor and measures of citation impact?
- to what extent could the usage factor be gamed, either by humans or machines, and are there digital signatures associated with such attempts to cheat the system?

This report begins to answer these questions.

CIBER would like to thank the UK Serials Group and COUNTER for funding this study. We are also very grateful to Laura Cox for her sterling preliminary work on the data and for her kind support.

Key recommendations 1 of 2

Critical properties of the usage factor

CIBER Research Limited

Recommendation 1

This report 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 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 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. articles in 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 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 page 12).

Journal usage factors should be published with appropriate confidence intervals 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 pages 15-19). Longer windows than 24-months are not recommended (see page 22) and this should be considered a maximum. There is possibly a case for considering a 12-month window (see 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 6-month window as well to provide further insights.

Recommendation 6

Usage in months 1-12 especially follows different patterns in different subject areas (see 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.

Key recommendations 2 of 2

Critical properties of the usage factor

CIBER Research Limited

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 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 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 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 impact 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 (pags 39-44) suggests that a cheap and expedient second line of defence would be to develop statistical forensics to identify suspicious behaviour, whether that 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 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

PART TWO
Patterns of use across items

This section looks at the frequency distribution of downloads at the item level.
Its purpose is to guide judgements concerning the best way to summarise `average' use.

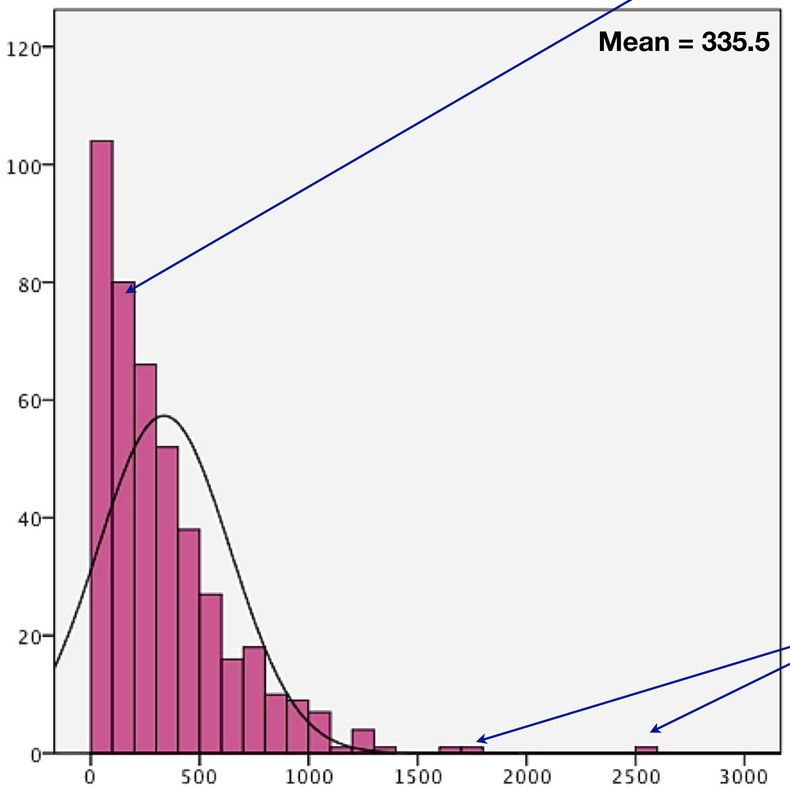
Patterns of use across items

A problem with averages: the Bill Gates problem

Medical and life sciences journals (n=48 titles) all items all versions

CIBER Research Limited

Many articles used a few times



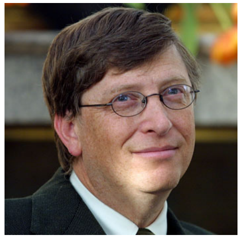
A few articles used many times

About this slide

This figure shows the frequency with which individual items are downloaded from 48 medical and life sciences journals The publication year is 2008 and this histogram shows use in month 2. The pattern is approximately lognormal: many items are used rarely, a few items are used many times.

A major downside in using the arithmetic mean to summarise this data is that the few heavily used items will exert a major effect: the mean will be a lot higher than other averages such as the mode or median. The International Mathematical Union has criticised ISI's citation impact factor for this reason.

An analogy. If Bill Gates gets onto a London bus, the average income of the passengers rises such that they are all millionaires several times over. Their income falls back to the UK average when he gets off. But median income remains the same throughout the journey.



Patterns of use across items

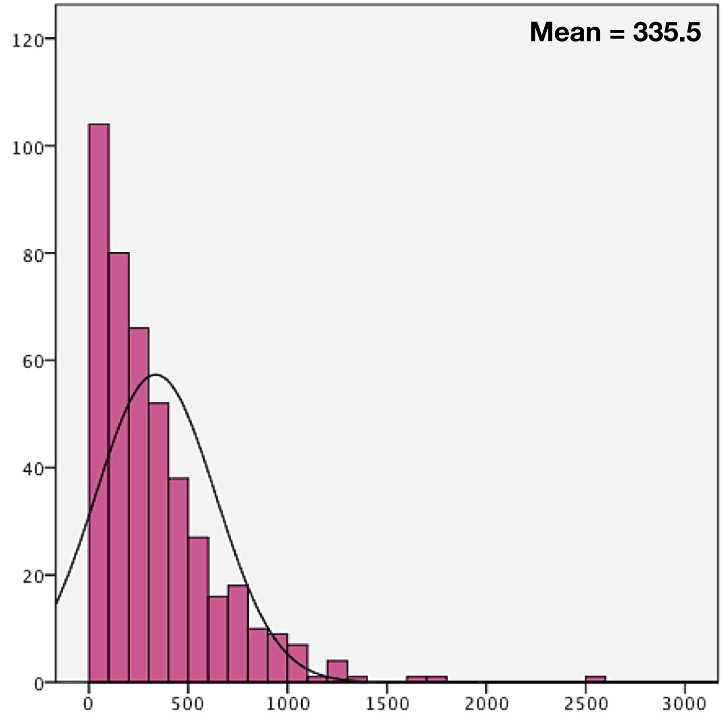
A problem with averages: the Bill Gates problem

Medical and life sciences journals (n=48 titles) all items all versions

CIBER Research Limited

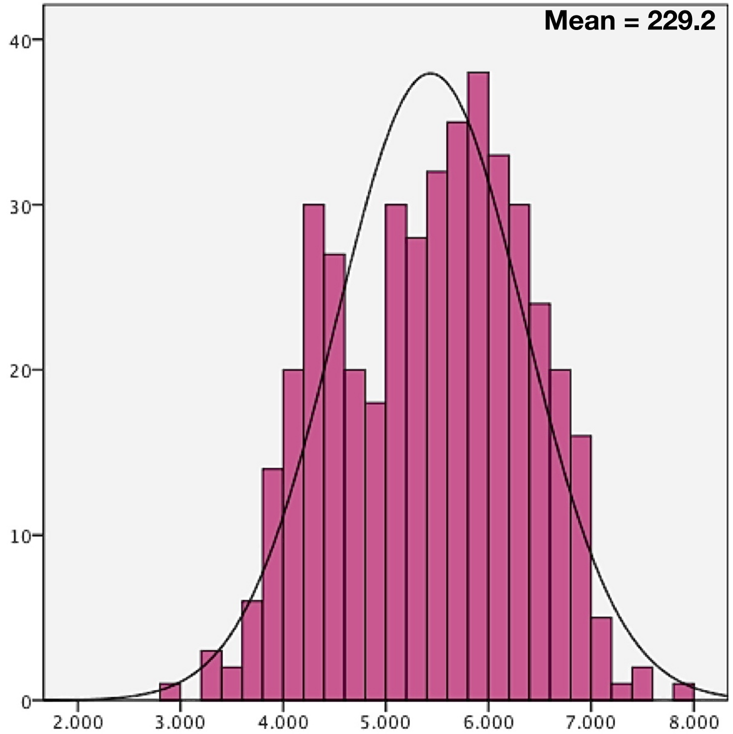
What can be done?

If we convert numbers of downloads to their natural logarithm, we get a more or less normal distribution. This is what CIBER has done in this report in order to run statistical tests on usage factors that make sense. In the `real world' the simple application of the median rather than the average would remove the distorting effect of a small number of very heavily downloaded items very effectively. In fact the CIBER estimates are very similar to the median.



natural log transform

from lognormal to normal



Patterns of use across items

2007 usage factors for medical and life sciences titles

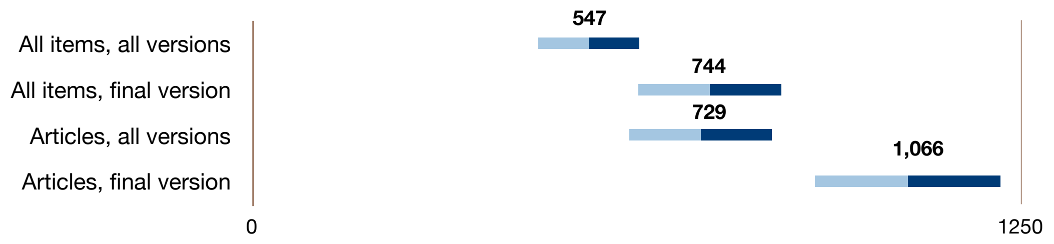
Medical and life science journals (n=48 titles)

CIBER Research Limited

We now present some real data. These journal usage factors are based on one publication year (2007) and use over the first 24 months.

These slides show the usage factor calculated as an arithmetic mean (top) and CIBER's estimation of the median (bottom). Both calculations reveal wide variation according to the types of items included or not included.

Publication year 2007, usage in months 1-24
Mean journal usage factors with 95% confidence intervals

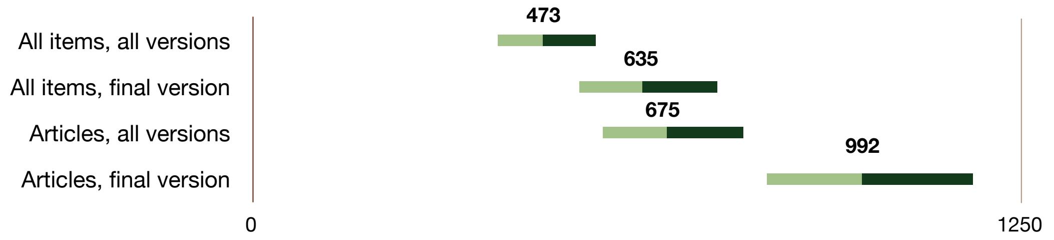


Articles, final versions attract significantly higher use than all the others
ANOVA F=13.1,p < 0.01

How to read these graphics

The charts show the average journal usage factors for different mixes of items and versions, with 95% confidence intervals.

Publication year 2007, usage in months 1-24
CIBER median journal usage factors with 95% confidence intervals



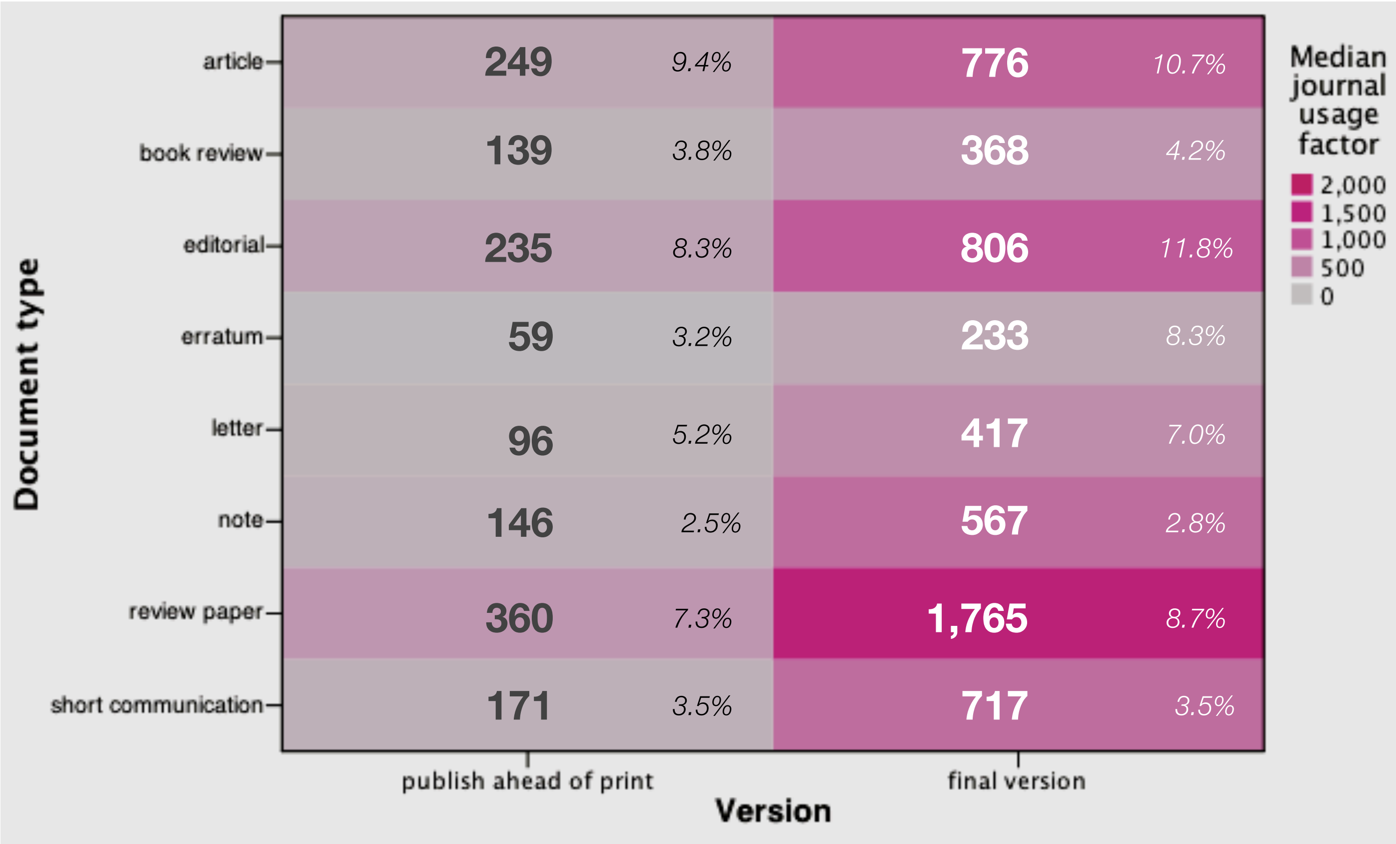
Articles, final versions attract significantly higher use than all the others
ANOVA F=11.0,p < 0.01

Patterns of use across items

2008 usage factors by document and version type

Medical and life science journals (n=48) all items, all versions

CIBER Research Limited



How to read this graphic

This heat map shows the 2008 24-month usage factor broken down by document type and version, with average downloads per category. The colour coding places the numbers in broad bands.

This is indicative only: publishers are not consistent in how they describe document types.

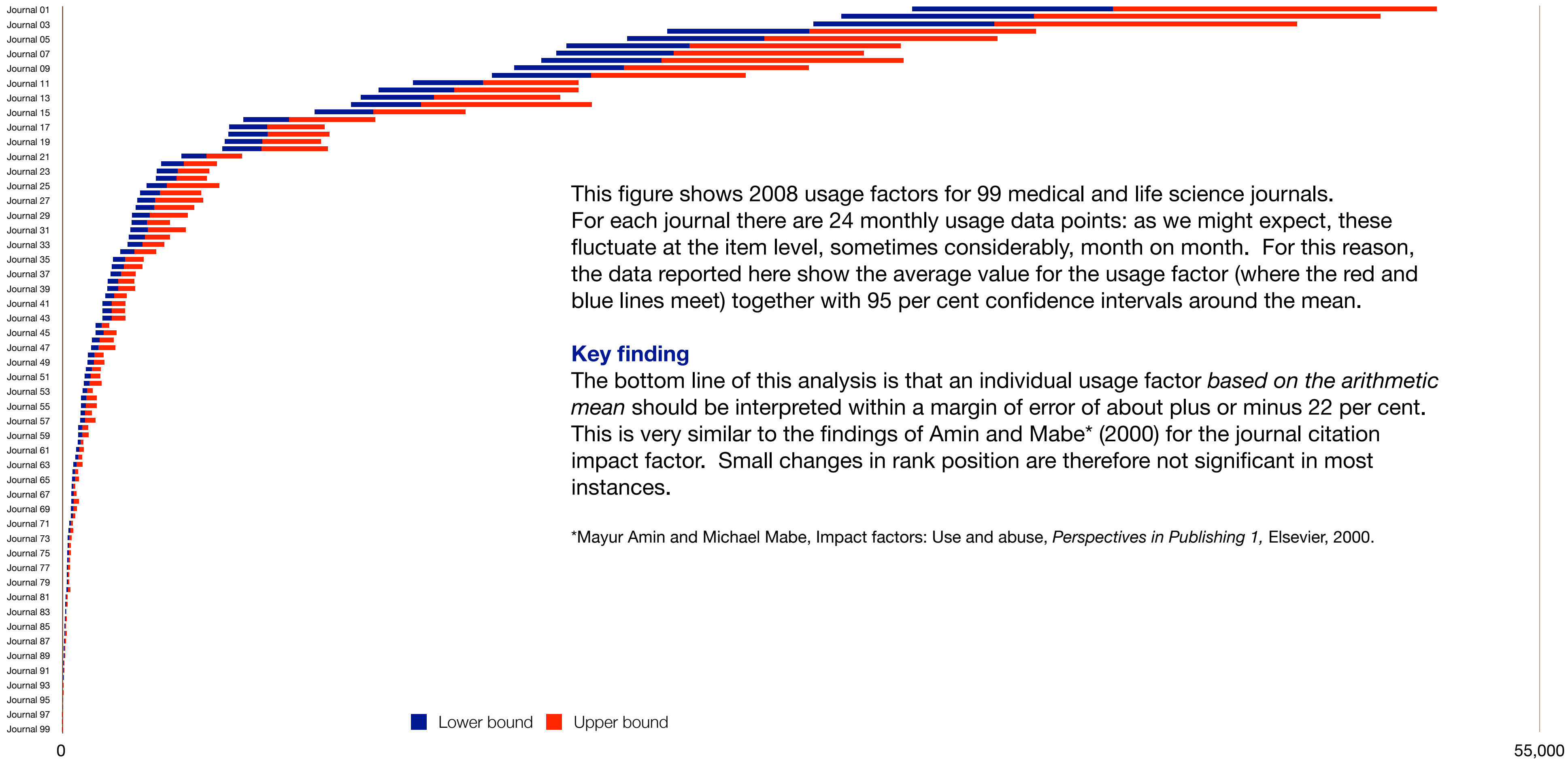
Key finding

Variation by document type and version is very considerable. It follows that the usage factor will vary according not just to the popularity of the journal but to the specific mix of editorial and original research content.

Patterns of use across items

How informative are journal rankings based on usage factors?

Medical and life science journals (n=99 titles) all items all versions



This figure shows 2008 usage factors for 99 medical and life science journals. For each journal there are 24 monthly usage data points: as we might expect, these fluctuate at the item level, sometimes considerably, month on month. For this reason, the data reported here show the average value for the usage factor (where the red and blue lines meet) together with 95 per cent confidence intervals around the mean.

Key finding

The bottom line of this analysis is that an individual usage factor *based on the arithmetic mean* should be interpreted within a margin of error of about plus or minus 22 per cent. This is very similar to the findings of Amin and Mabe* (2000) for the journal citation impact factor. Small changes in rank position are therefore not significant in most instances.

*Mayur Amin and Michael Mabe, Impact factors: Use and abuse, *Perspectives in Publishing 1,* Elsevier, 2000.

Patterns of use across items

Preliminary conclusions

CIBER Research Limited

The journal usage factor should not be based on an arithmetic average of the raw data as this makes it too sensitive to a few very highly downloaded items. Several approaches are possible to a `better' estimate of the average and by far the simplest is to take the median. The median is simply the "middle number" in a sorted list of numbers and it is a very good way to think about the `average case' when the data are skewed.

There is considerable variation in use according to document type and version. This does not matter in one sense: if the intention of the exercise is to simply report on usage for the journal as a whole package, then it is completely valid to include all items and all versions. This would certainly have practical advantages for data processing.

However, and the same stricture* applies to the classic impact factor, a journal usage factor could easily be manipulated by simply changing the balance of item types in the final issue. The data on page10 and, especially page 11 suggests that this is a very serious issue for usage data.

It is possible to rank journals according to their usage factor but care should be taken to include additional information in any published data. Any published journal usage factors should include appropriate error bars around the average, as on page 12.

*See, for example, Mayur Amin and Michael Mabe, Impact factors: Use and abuse, *Perspectives in Publishing 1,* Elsevier, 2000.

PART THREE
Patterns of use across time

This section looks at the obsolescence rates of all items published in 2006.
Its purpose is to guide a judgment as to the optimal time window for the journal usage factor.

Patterns of use across time

Monthly use of all items published in 2006

Engineering journals (n=21 titles) all items, all versions

CIBER Research Limited

About this slide

In order to have an informed discussion about the optimal length of the time window to record downloads for the usage factor, we need to understand how items are used over time, especially the point at which interest in journal content begins to wane.

In this and the following slides, we take all document types (in all versions) published in 2006 and look at their monthly pattern of use over the subsequent three years.

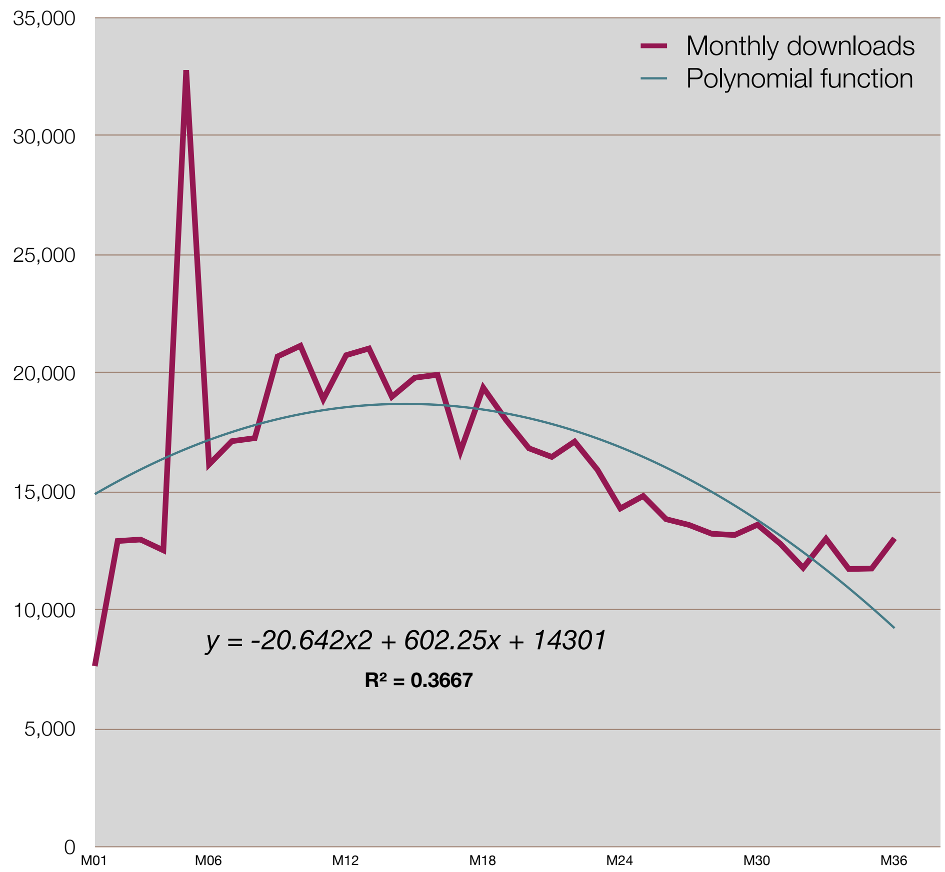
Ideally, we need a longer time series to be sure, but this is all we have.

The trend line, which admittedly in this case does not give an excellent fit to the data, suggests that aggregate usage of 2006 engineering items will trickle to near zero (i.e. become `asymptotic') at around **45 months** after publication.

The life span of original research articles and review papers is likely to be longer, as the `all items' approach used here will contain much relatively ephemeral material such as editorial material and rapid communications.

How to read this graphic

This chart shows the number of downloads each month with a trend line.



Patterns of use across time

Monthly use of all items published in 2006

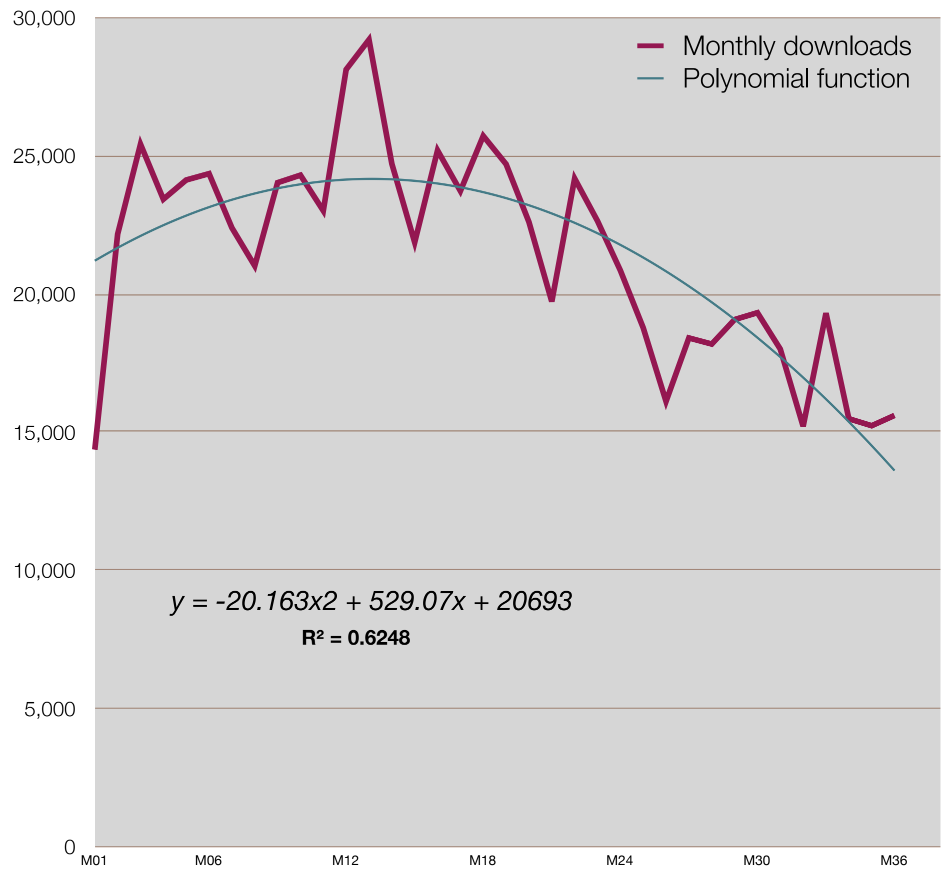
Humanities journals (n=24 titles) all items, all versions

CIBER Research Limited

About this slide

Humanities items follow a generally similar pattern to engineering but with a shorter and more delayed peak.

The trend line, which offers a reasonable fit to the data, suggests that aggregate usage of 2006 humanities items will trickle to near zero (i.e. become `asymptotic’) at around **48 months** after publication.



Patterns of use across time

Monthly use of all items published in 2006

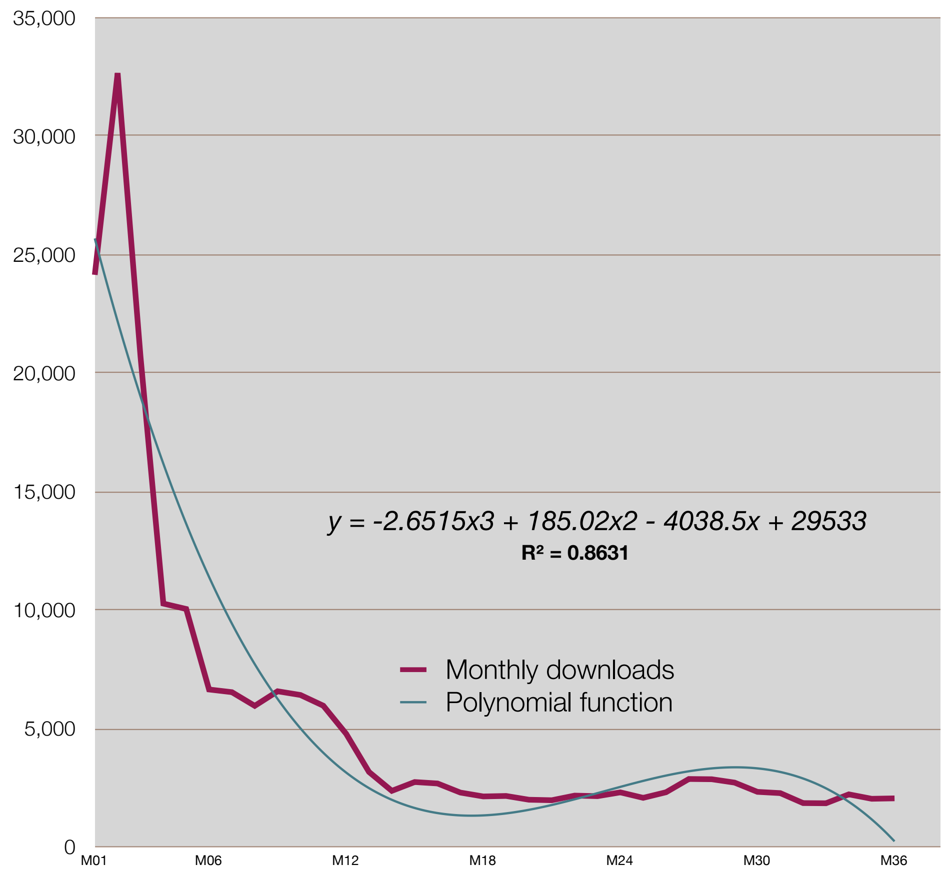
Physical sciences journals (n=3 titles) all items, all versions

CIBER Research Limited

About this slide

The monthly pattern of use for physical sciences items is very different from the other broad subjects in this study. There is a very sharp initial peak followed by continuing and steady interest in items in the period months 14-36 [caution: we only have three journals].

There is not enough data with only three titles to justify calculating an end point for physical sciences articles, but the well-fitting trend line suggests it may have been reached at or just after **36 months.**



Patterns of use across time

Monthly use of all items published in 2006

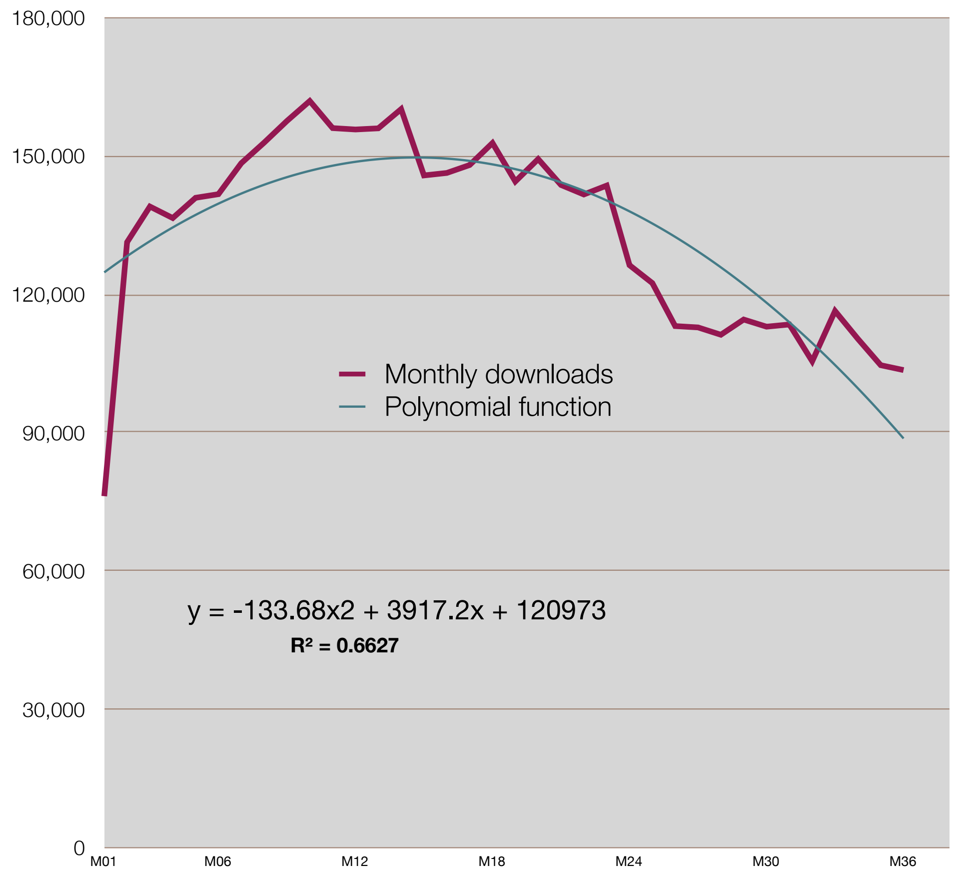
Social sciences journals (n=115 titles) all items, all versions

CIBER Research Limited

About this slide

The pattern in the social sciences is broadly similar to that for humanities items.

The trend line, which offers a reasonable fit to the data, suggests that aggregate usage of 2006 social sciences items will trickle to near zero (i.e. become `asymptotic’) at around **47 months** after publication.



Patterns of use across time

Monthly use of all items published in 2006

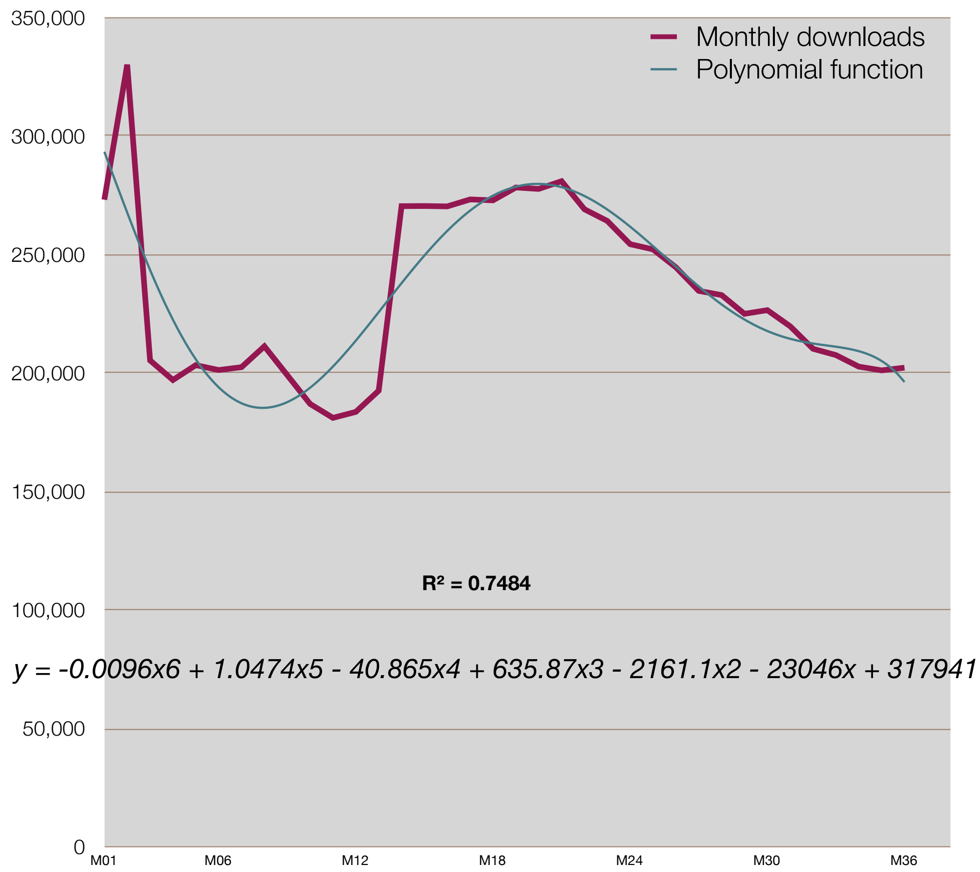
Medical and life sciences journals (n=47 titles) all items, all versions

CIBER Research Limited

About this slide

Monthly usage in the medical and life sciences shows an interesting double peak: a very immediate one in the first few months and another from month 12, which may well be due to a delayed open access and / or possible citation effects.

The trend line, which offers a good fit to the data, suggests that aggregate usage of 2006 medical and life science items will trickle to near zero (i.e. become `asymptotic') at around **40 months** after publication.



Patterns of use across time

Cumulative monthly use of all items published in 2006

Medical and life sciences journals (n=48 titles) all items, all versions

CIBER Research Limited

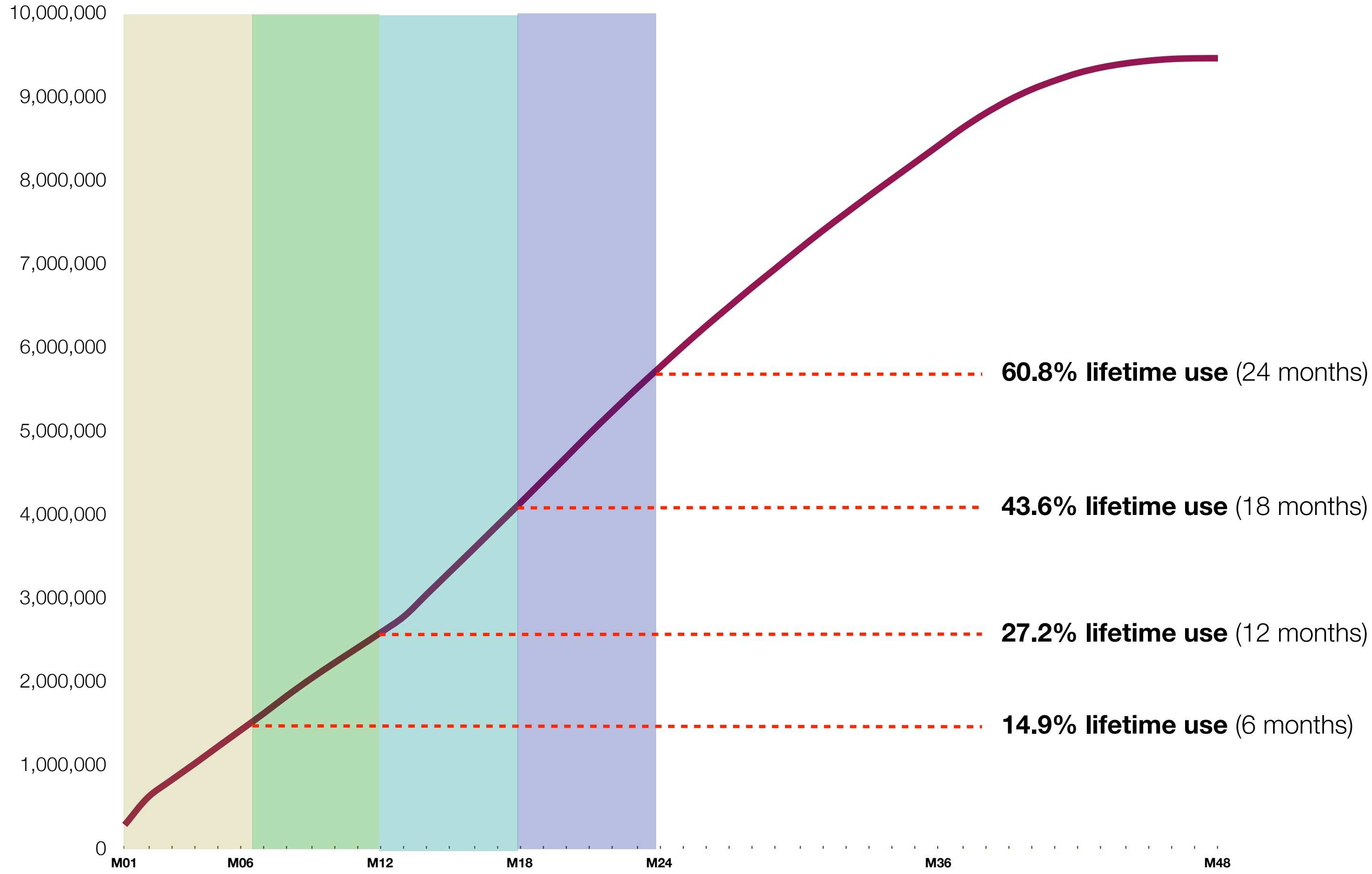
Implications of different time windows

This slide, for the medical and life sciences estimates how much cumulative life time use has been achieved after 6, 12, 18 and 24 months. This is important in the context of deciding which time window is most appropriate for the calculation of the usage factor.

Several conclusions can be drawn from this and the previous slides.

The first is that a 48-month time window would capture maximum information in the period between initial publication and the beginnings of a steep decline in user interest. Of course, this is far too long for any practical purpose and a more realistic choice would be to go for 6, 12, 18 or 24 months as these are still very informative.

An advantage of a really short window (6 months) would be that any effects due to delayed open access or citation effects could be mitigated against. More research is needed here.



Patterns of use across time

Cumulative use of all items published in 2006 by usage time window

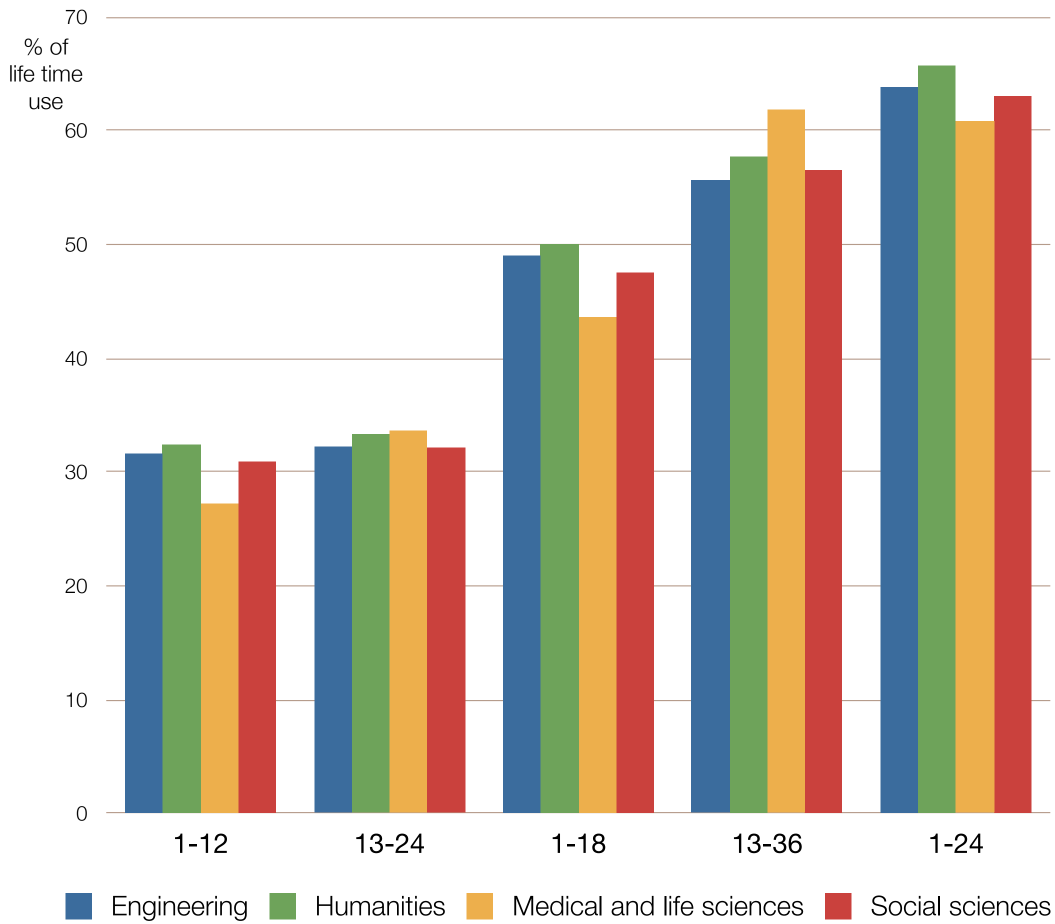
Comparison by broad subject area

CIBER Research Limited

About this slide

The table and figure summarise the data in the previous slides. Physical sciences are excluded because of the difficulty of finding an asymptote. The patterns of use across time are similar across the four remaining and highly contrasted disciplines.

Usage months	**Engineering**	**Humanities**	**Medical and life sciences**	**Social sciences**
1-12	31.6%	32.4%	27.2%	30.9%
1-18	49.0%	50.0%	43.6%	47.5%
1-24	63.8%	65.7%	60.8%	63.0%
13-24	32.2%	33.3%	33.6%	32.1%
13-36	55.6%	57.7%	61.8%	56.5%



Patterns of use across time

Usage decay by document type

2006 publications in all subjects (n=210 titles) all versions

CIBER Research Limited

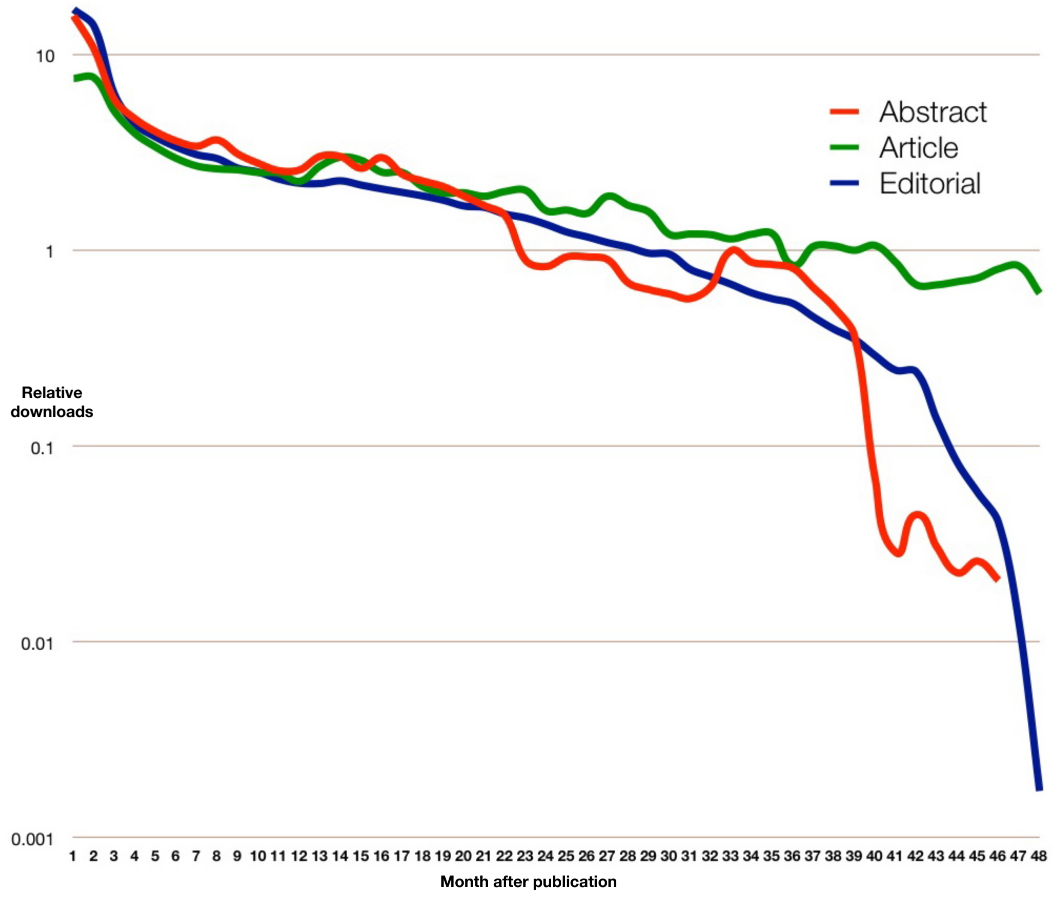
The previous slides have considered time effects on use for all document types together, thus modelling consumer interest in the journal as a complete package of research, editorial and other content. We might expect, however, that interest in different document types will wane faster for some than others.

Key finding

Over the first 22 months there is very little difference in the decay curves for abstracts, original research articles and editorial content. After this point, user activity changes. Articles maintain their interest over the next two years, albeit fading away slowly. Interest in abstracts and editorial matter fades away dramatically after year three.

How to read this graphic

This chart shows how interest in different document types wanes over time. The vertical axis is a measure of download activity, presented here on a log scale, just to make the patterns more clearly visible.



Patterns of use across time

Preliminary conclusions

We ideally need a longer time series to be sure but it appears that a good working estimate for the useful lifetime of most journal content (all items in all versions) is about four years. The longevity of original research articles and review papers is likely to be longer than this, and possibly more highly differentiated between subjects but if all items are used, then this seems a reasonable position to take.

All the subjects we looked at show a peak roughly between months 6 and 12 and a broadly similar (and steady) pattern of cumulative item use in years 1-3.

Tentatively, a time window based on months 1-24 would seem to be the most appropriate: capturing information both about the peak and the subsequent steady state growth. If the estimations of lifetime use are accurate, roughly four years for all items, then it would appear that a 1-24 month window will capture a substantial proportion of lifetime use, probably of the order of 60 per cent as a global figure. From the data presented here, a 1-24 month time window would not suffer greatly from the confounding effect of different document types fading in interest at different rates.

An alternative strategy would be to go for a very short time window, 6 months, on the basis that the confounding effects of embargo periods, whether relating to publisher delayed open access or the deposit of Stage II manuscripts in open repositories, would be minimised. Even with a 6-month window around a sixth of lifetime use would be captured.

PART FOUR
Stability of the usage factor
This section asks how stable the usage factor is over time.

Stability of the usage factor

Context: year-on-year changes in rankings for ISI impact factors

Medical and life science journals (n=36 titles) ISI impact factors

CIBER Research Limited

How stable are journal rankings based on the classic ISI impact factor? As a context for comparing the properties of the usage factor, we begin this section by looking at changes in rankings among 36 medical titles over three years based on the classic citation measure: the ISI impact factor.

Key findings

Journal rankings in the medical and life sciences based on the ISI impact factor are pretty consistent over the period 2006-2008. Journal order correlates highly and significantly from year to year.

Rank order correlations

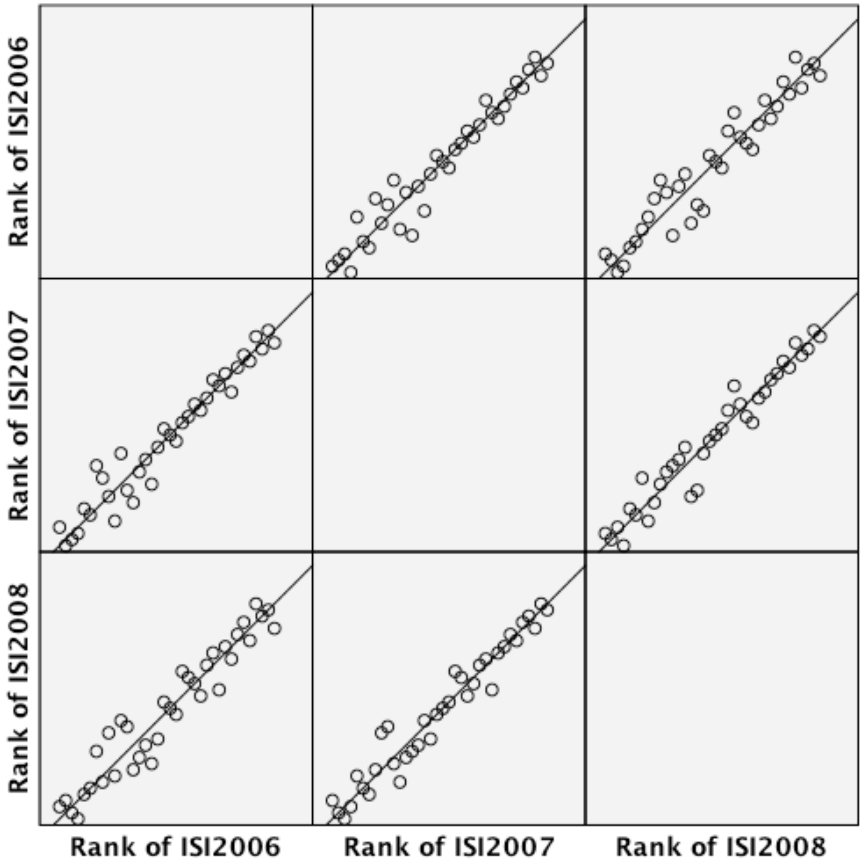
2008 vs 2007: Spearman's *rho* = 0.973, $p < 0.01$

2007 vs 2006: Spearman's *rho* = 0.971, $p < 0.01$

2008 vs 2006: Spearman's *rho* = 0.959, $p < 0.01$

How to read this graphic

We start by putting the journals into ranked order by ISI impact factor for each of the three years 2006-2009. These charts show how these ranked orderings compare across different years. For example, the middle chart in the right hand column compares 2007 and 2008. If there were no volatility in journal citation ranking, all the journals would like on the diagonal. Journals below the diagonal have fallen down the pecking order.



Stability of the usage factor

Context: year-on-year changes in rankings for Elsevier SNIP metrics

All subjects (n=278 journals) Elsevier SNIP metric

CIBER Research Limited

This slide repeats the previous one, this time using Elsevier's SNIP (Source Normalised Impact per Paper) as the measure of citation impact. Unlike the ISI impact factor, SNIP is field-independent, so we can legitimately combine data for all the subjects used in this study rather than just present the results for medicine and the life sciences.

Journal rankings based on Elsevier's SNIP metric are also consistent year on year, although they are a little more volatile than the ISI impact factor:

Rank order correlations

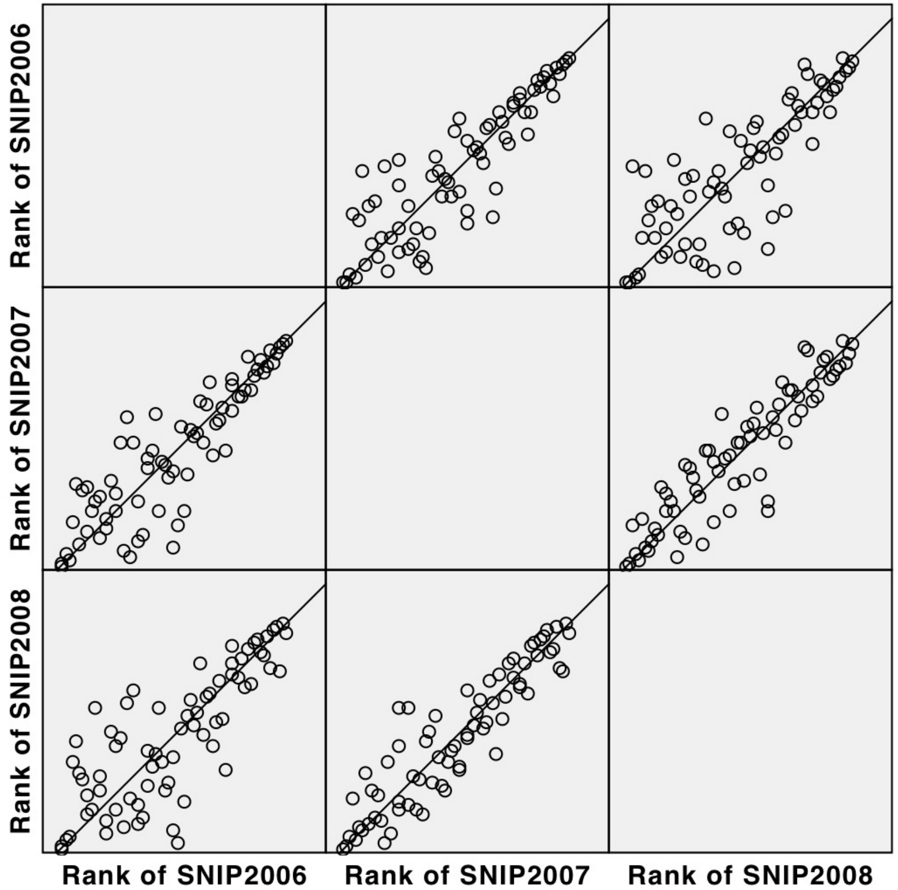
2008 vs 2007: Spearman's *rho* = 0.901, $p < 0.01$

2007 vs 2006: Spearman's *rho* = 0.909, $p < 0.01$

2008 vs 2006: Spearman's *rho* = 0.828, $p < 0.01$

Key finding

Measures of citation impact deliver journal rankings that are broadly consistent year-on-year, with just enough variation to keep them interesting.



Stability of the usage factor

Year-on-year changes in rankings for the usage factor

Medical and life science journals (n=48) all items, all versions

CIBER Research Limited

Having seen how rankings based on citations compare over time, we now turn to the usage factor itself. Is it stable enough to bear comparison with the established citation metrics?

The usage factors reported here are based on a single publication year with use being measured in months 1-24 and include the whole journal (all items and all versions).

Usage factor rankings deliver high and statistically significant correlations year-on-year:

Rank order correlation

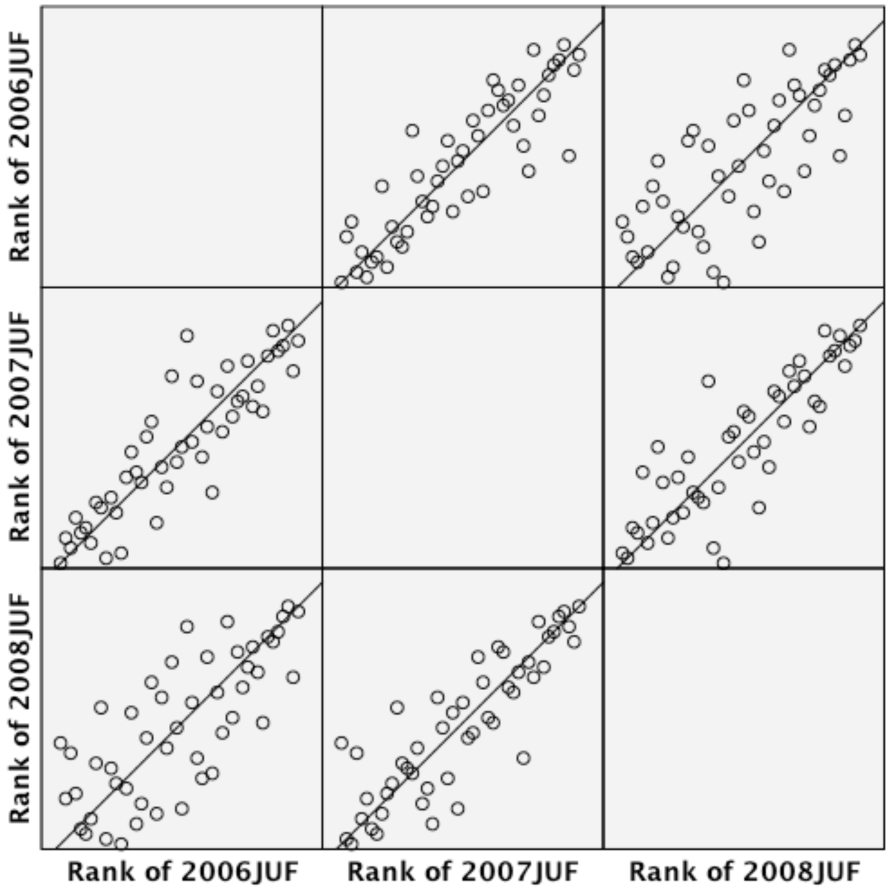
2008 vs 2007: Spearman's *rho* = 0.886, $p < 0.01$

2007 vs 2006: Spearman's *rho* = 0.862, $p < 0.01$

2008 vs 2006: Spearman's *rho* = 0.755, $p < 0.01$

Key findings

These correlations are smaller than those for the impact factors in the earlier slides, but they are still high and statistically very significant. This analysis shows that usage factors are more volatile than impact factors and any journal rankings based on them will show greater churn year on year. But, broadly speaking, they do a similar job.



Stability of the usage factor

Year-on-year changes in rankings for the usage factor

Medical and life science journals (n=36) articles only, final version

CIBER Research Limited

Perhaps a fairer comparison between usage and citation metrics would exclude editorial and other non-research material.

If we do this and restrict the analysis to articles only in their final version, we find a marked uplift in correlation:

Rank order correlations

2008 vs 2007: Spearman's *rho* = 0.915, $p < 0.01$

2007 vs 2006: Spearman's *rho* = 0.909, $p < 0.01$

2008 vs 2006: Spearman's *rho* = 0.878, $p < 0.01$

Key finding

It may or may not be practicable to construct usage factors that can be disaggregated by document type. The limited data presented in this report suggests that usage factors deliver rankings that are consistent and comparable with those generated from citation data.

Stability of the usage factor

JUF months 1-12 and months 1-24 compared by subject

2008 usage factor (articles only, final versions, months 1-12 and 1-24 (n=320 journals)

CIBER Research Limited

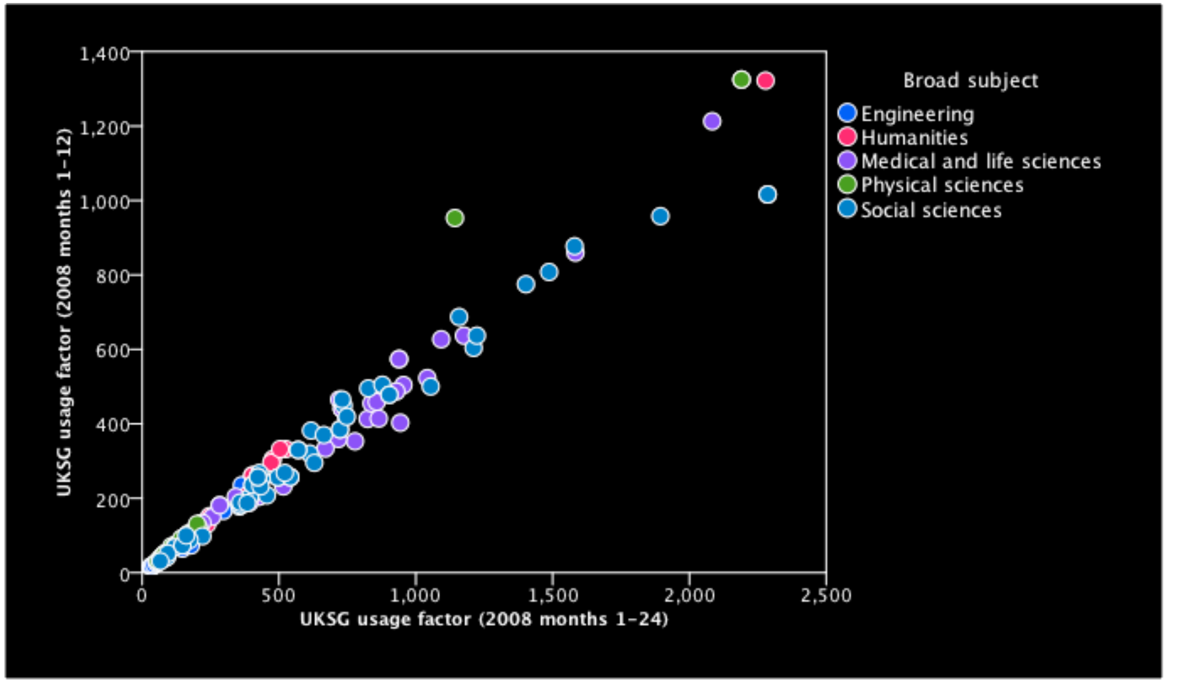
Another dimension of the usage factor's stability relates to the length of the time window within which downloads are captured. Do we lose useful information if we go for a shorter as opposed to a longer time window?

The analysis opposite suggests not necessarily.

The figure shows the relation between 2008 usage factors based on 12- and 24-month windows. Statistically, the correlation is very high (Spearman's *rho* = 0.99, $p < 0.01$).

Key finding

On the basis of this analysis, there is little to be gained by a 24- over a 12-month time window. If would appear that the former does not add significant information.



Stability of the usage factor

Coefficients of variance

All subjects, all items, all versions, publication year 2007, months 1-24 (n=28,667 items)

CIBER Research Limited

An annual usage factor is but a one digit summary of a lot of underlying and rather complex behavioural information. Over a period of time, many people will visit a journal online. They will make quick decisions on whether to stay, what they want to look at. They are not constrained in their behaviour in the same way they are when filling in their online tax return. They can come and go and do as they please. The result is that the usage factor is a mixture of random noise and clear signal.

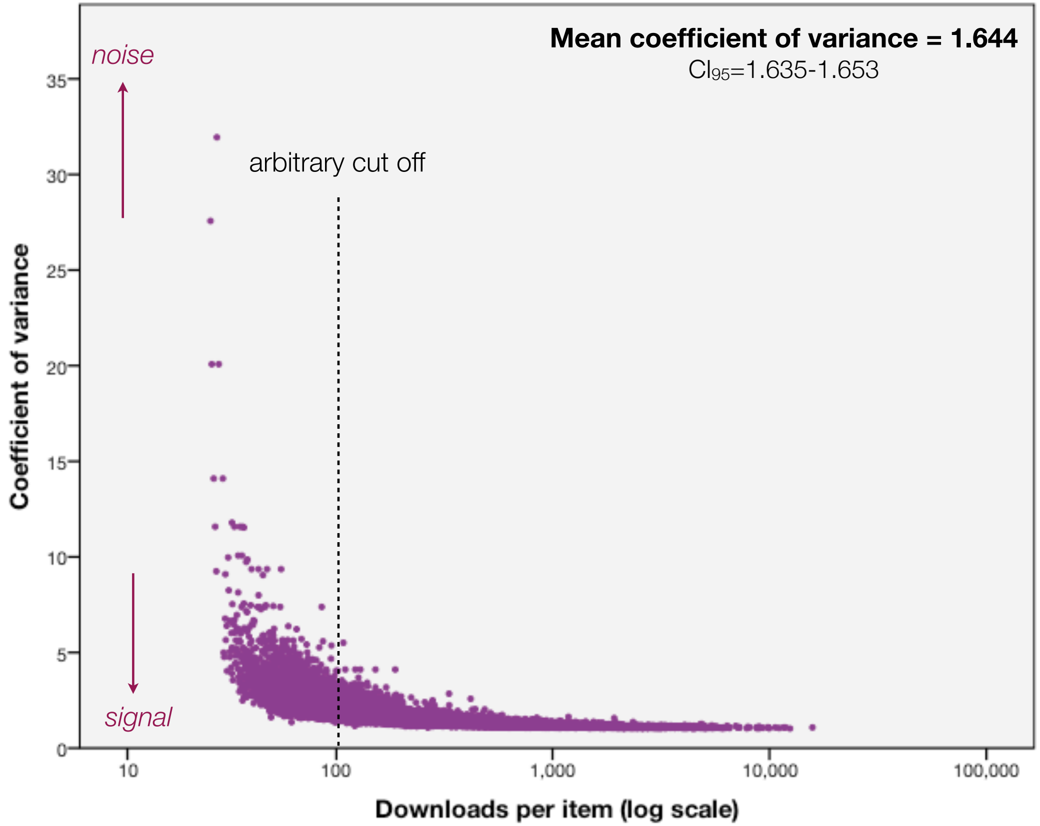
The figure opposite shows `noise' and strength of `signal' on the vertical. As the number of downloads per item increases, there is less noise and a clearer signal.

Key finding

The point at which the signal becomes clear and rises above the noise floor is arbitrary, but this analysis strongly suggests that usage factors of less than 100 are not likely to be accurate. Their values are highly volatile and they are not likely to accurately reflect `real use'.

How to read this graphic

The vertical axis on this figure shows the coefficient of variance: the standard deviation divided by the average (in this case the mean and the median are the same). The horizontal axis shows the number of downloads per item.



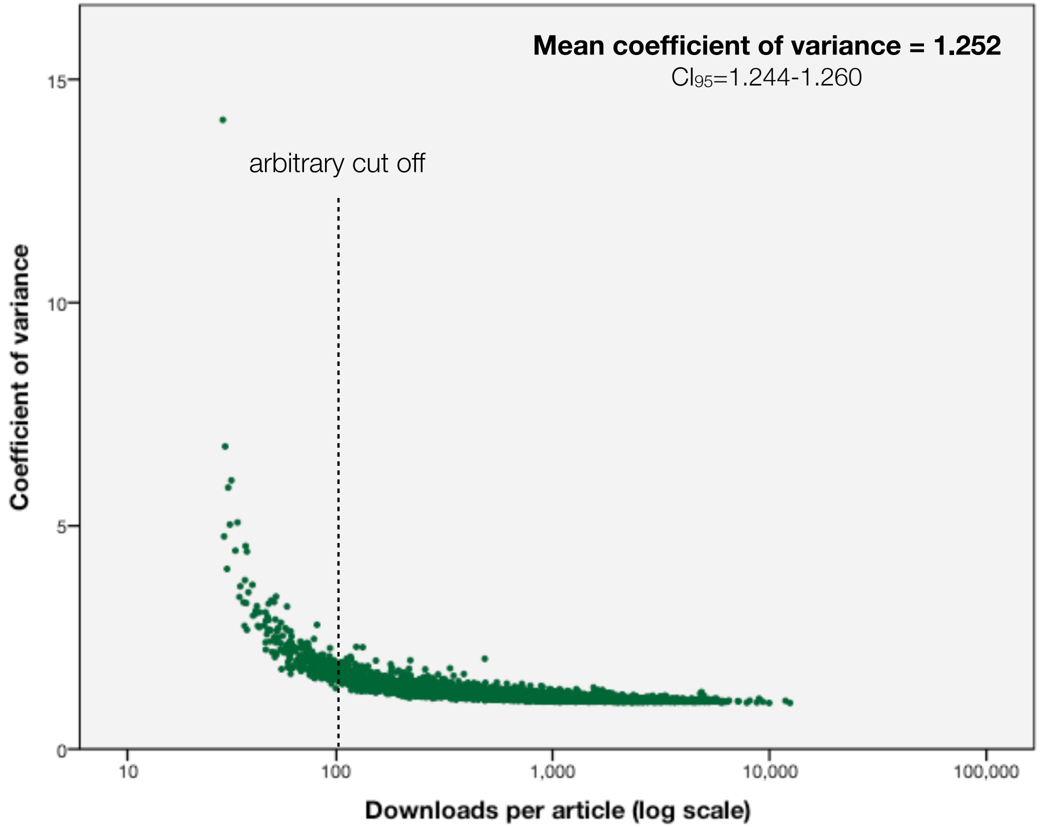
Stability of the usage factor

Coefficients of variance

All subjects, articles only, final version, publication year 2007, months 1-24 (n=6,324 articles)

CIBER Research Limited

In this analysis, we restrict the data to articles only in their final version. The overall level of noise is greatly reduced and the signal is generally clearer and sharper. But again, a lower limit of 100 for the usage factor seems easily justified.



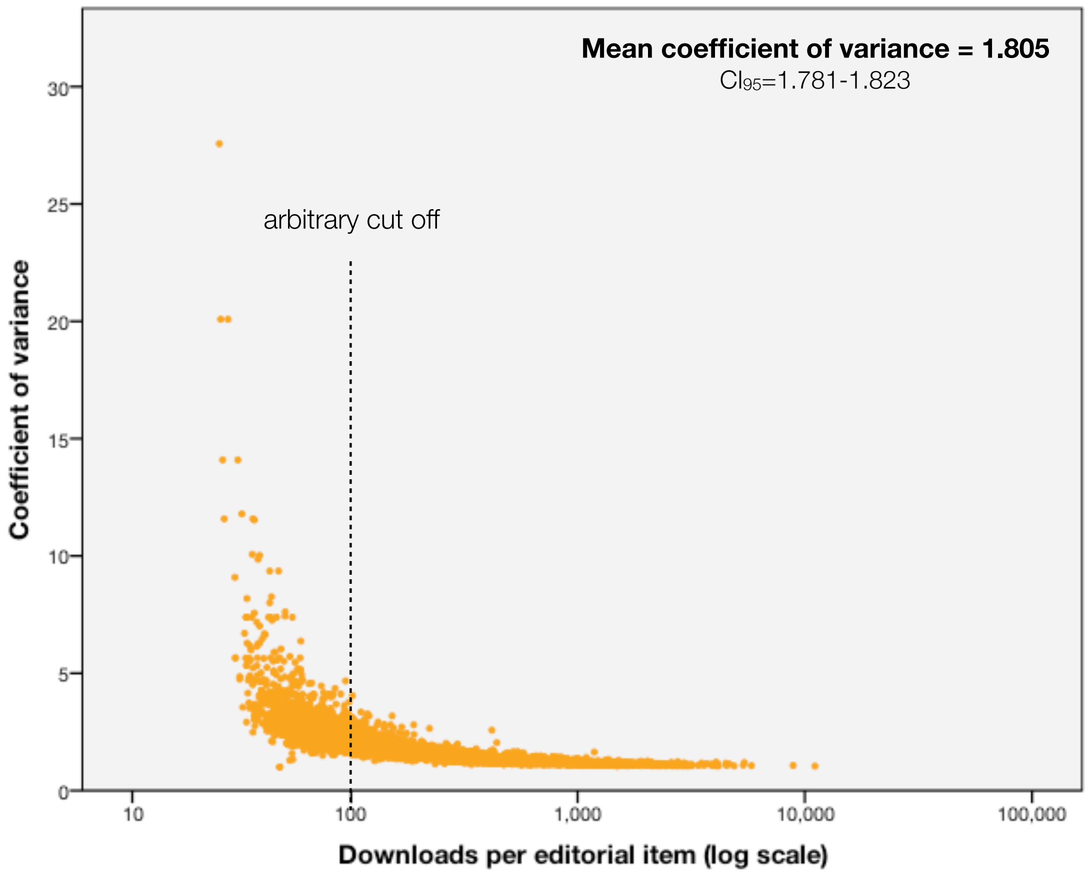
Stability of the usage factor

Coefficients of variance

All subjects, editorial matter only, all versions, publication year 2007, months 1-24 (n=6,480 items)

CIBER Research Limited

And a similar conclusion obtains when we only consider editorial material.



Stability of the usage factor

Preliminary conclusions

CIBER Research Limited

Even more so than in earlier sections, the research presented here is merely illustrative of some of the issues that need to be considered if and when the usage factor is brought into full scale production. We need more consistent data over a longer period to be really sure, but these are some useful early pointers.

The themes emerging so far are that there are sufficiently large differences in usage factor between titles to suggest they may be useful in separating sheep and goats. Journal rankings based on usage are a little less stable than those based on citation impact, but year-on-year variation is relatively modest and certainly comparable with the rankings generated by ISI and SNIP citation metrics. In any case, impact and usage factors need to be interpreted within margins of error of around plus or minus 22 per cent. For this reason alone, the use of three decimal places of precision to record ISI impact factors is questionable: usage factors in this report are presented as integers.

Usage factors with a value below 100 should not be trusted to provide an accurate reflection of real use. At and below this level, random noise is a major issue. As well as being unreliable, low usage factors can be more easily gamed, an issue that we will return to in Section 6.

PART FIVE
Patterns of use and citation

This section further compares the usage factor with measures of citation impact to see what, if any association there is between the two.

Patterns of use and citation

Usage and ISI impact factors compared by subject

2008 usage (articles only, final versions, months 1-12) and ISI impact factors compared

CIBER Research Limited

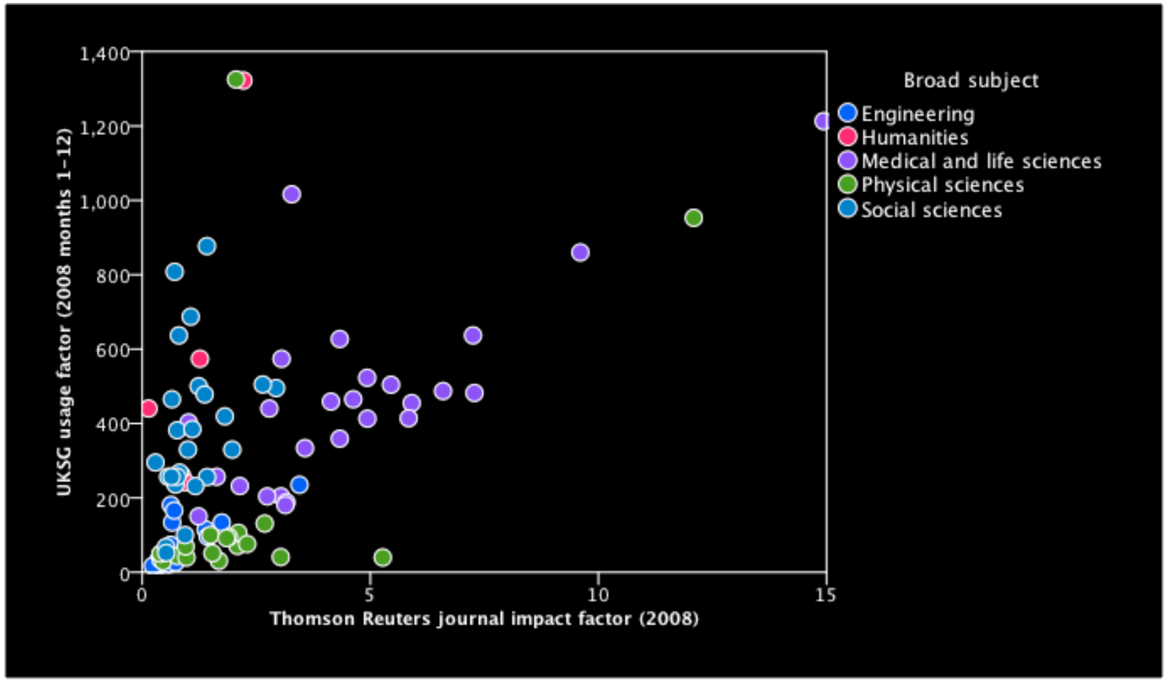
In this section we compare the usage factor directly with two measures of citation impact: the classic ISI impact factor and Elsevier's SNIP metric.

Our starting point is that we do not expect to see a clear correlation between them. They are measuring different things (`votes' by authors and readers) and the two populations may or may not be co-extensive.

This figure plots usage against the ISI impact factor for 148 journals drawn from all five subject areas.

Key finding

The 2008 usage and ISI impact factors are statistically independent (Spearman's *rho* = -0.07, *p*=0.28). In other words, knowledge of one does not enable one to predict the value of the other. Citation and reading behaviours are different.



How to read this graphic

The values for 2008 usage and ISI impact factors are plotted against each other. Each dot is a journal. Different subjects are picked out in different colours according to the key.

Patterns of use and citation

The JUF and Elsevier's SNIP compared by subject

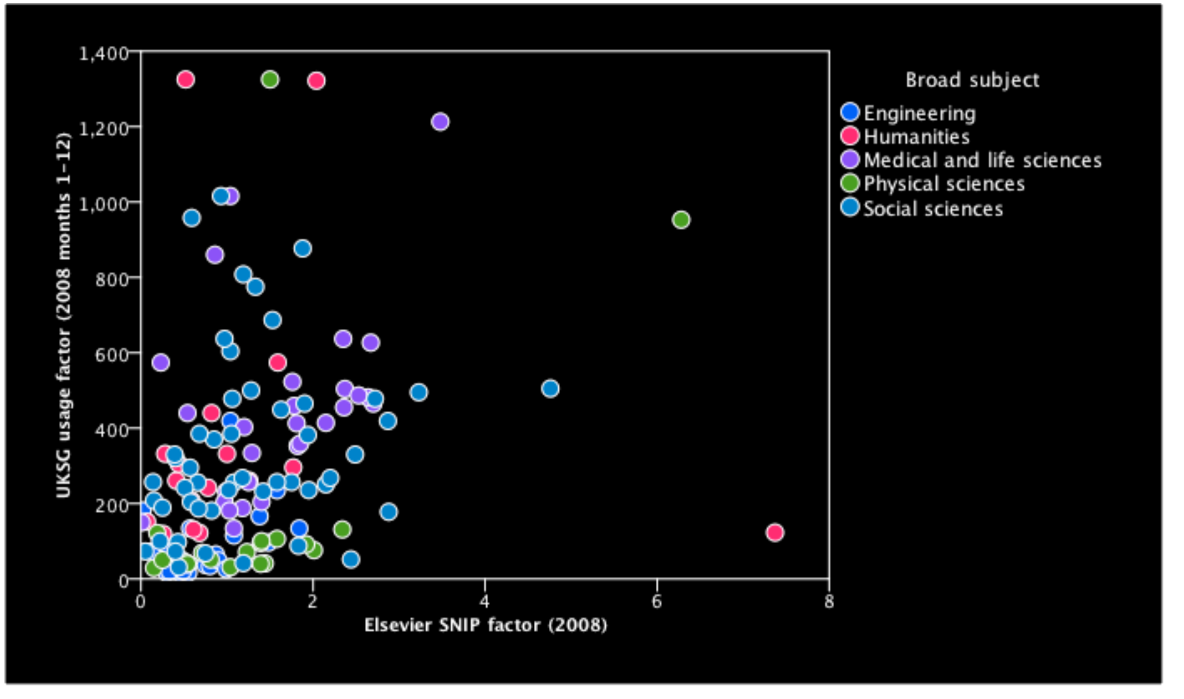
2008 JUF (articles only, final versions, months 1-12) and 2008 SNIP

CIBER Research Limited

Nor does the usage factor correlate meaningfully with Elsevier's SNIP metric (Spearman's *rho* = 0.11, *p*>0.05).

This is a blessing in a sense: if citation behaviour was strongly associated with use, there would be little or no point developing a usage indicator. As it stands, the two provide information about different kinds of behaviour.

There is no clear patterning by subject: we have examples here of relatively high usage and relatively low citation (and vice versa) in all the subject areas.



Patterns of use and citation

Preliminary conclusions

CIBER Research Limited

This report finds no evidence that usage and citation impact metrics are statistically associated. This is hardly surprising since author and reader populations are not necessarily co-extensive. Indeed in the case of practitioner-facing journals, the overlap will be minimal.

As a result, the usage factor adds new evidence to our understanding of the structure and dynamics of reading. It also opens up the possibility of developing new ways of looking at scholarly communication, with different journals occupying very different niches within a complex ecosystem.

PART SIX
Gaming the journal usage factor

This section explores three scenarios to see what kinds of diagnostic might help to spot gaming attempts.

Gaming the journal usage factor

The three scenarios

CIBER Research Limited

Indicators systems affect the way people behave. Can the usage factor can be gamed and, more to the point, can deliberate attempts to inflate it be easily spotted? Our starting point lies in the natural behaviour of real users. In CIBER's long experience, usage data is almost always lognormally distributed when people are able to choose what they look at without constraint. This can be seen in the figure opposite, where a few items attract a huge amount of use while most items attract little interest.

In a thought experiment to explore gaming, we use a hypothetical journal with 1,000 items and a median usage factor of 324 as the test bed for three gaming scenarios:

Lone wolf

An author looking for promotion downloads his best paper ten, a hundred, a thousand, ten thousand or a hundred thousand times.

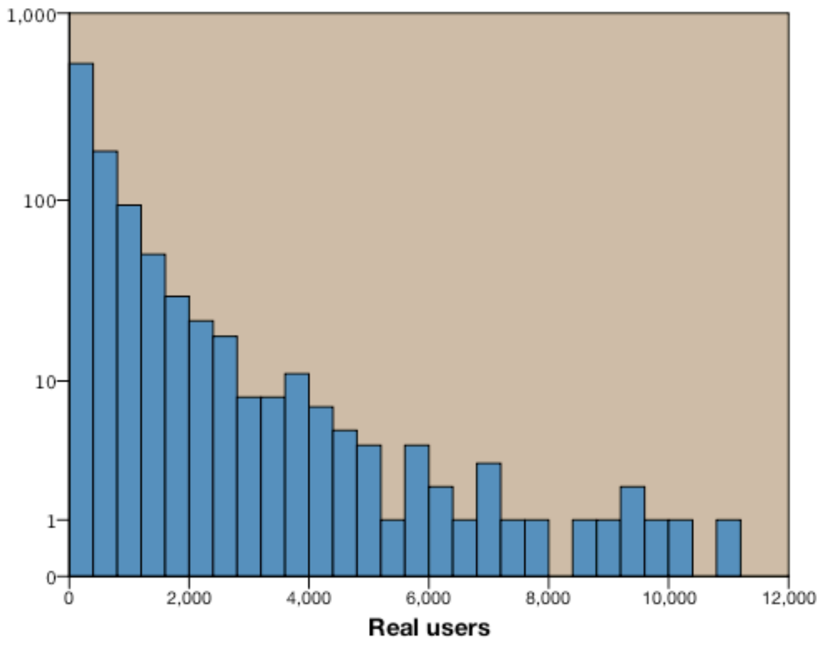
Random attack

A publisher engages a computer science student to develop a software agent that randomly downloads half of the items in the journal, *n* times as above.

Carpet bombing

A publisher seeking a higher usage factor for a journal develops a software agent that downloads *every* item in the journal *n* times.

Can these attacks be detected?



Gaming the journal usage factor

The three scenarios

Lone wolf

CIBER Research Limited

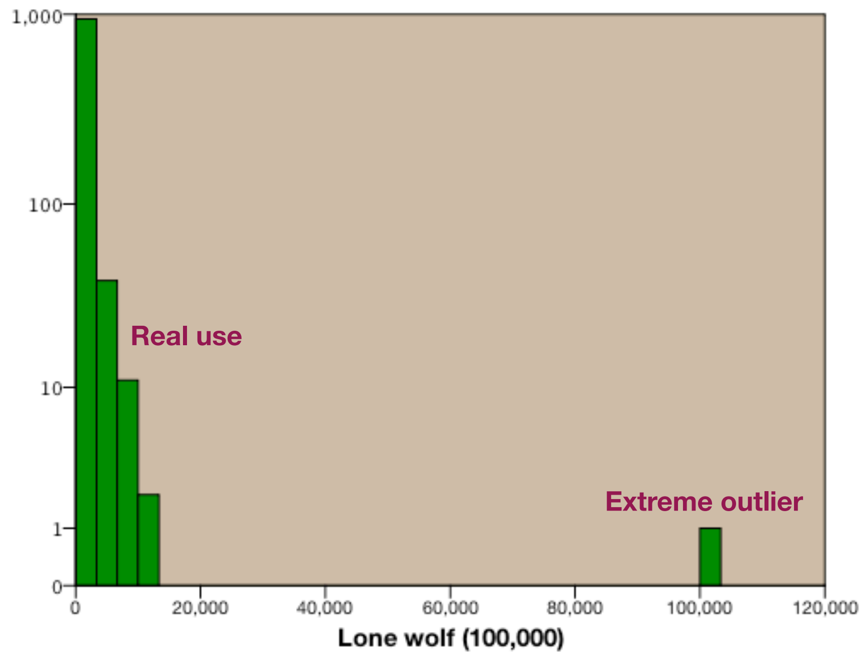
Lone wolf behaviour is easily spotted if it is extreme enough, as we can see opposite.

If we calculate the usage factor using the median, then it remains at its `correct' level of 324 no matter how many times the author downloads his paper: ten or a hundred thousand times.

If we were to use the arithmetic mean, the average rises from 806 to 906 in the extreme case of 100,000 additional downloads for a single item. This artificially inflates the usage factor by 12 per cent.

High intensity lone wolf behaviour be **easily detected** by screening the data for the kinds of extreme outliers shown opposite. The question could then be asked whether this was a genuine paper that had gone viral (like the Wakeford MMR research) or something more suspicious?

At low levels of intensity, it would not be possible to detect lone wolf activity from aggregate statistics. Over the time window chosen (possibly two years), this kind of behaviour is to be expected but, as we have seen, it makes no difference if we follow CIBER's strong recommendation and calculate using the median.



Gaming the journal usage factor

The three scenarios

Random attack

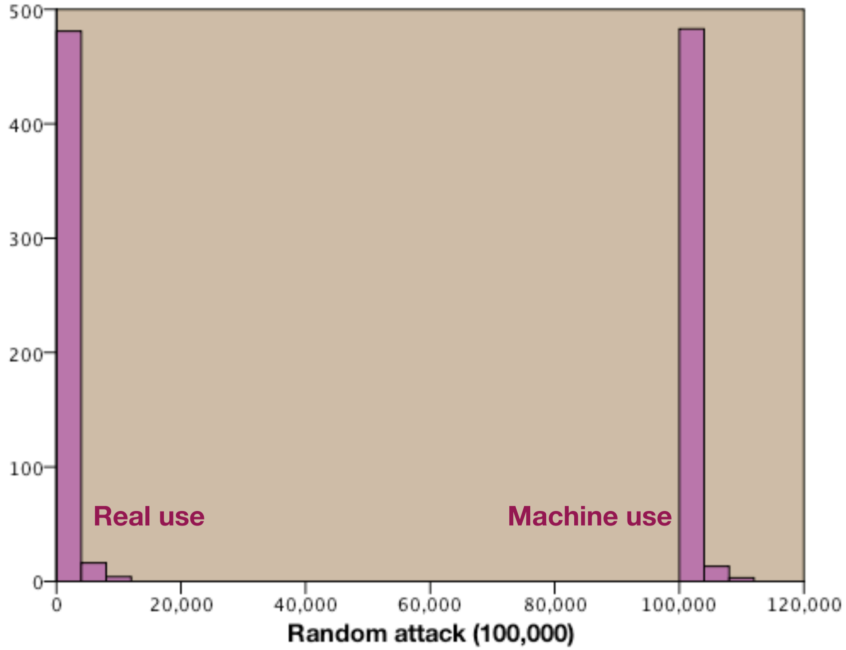
CIBER Research Limited

In this scenario, a software agent downloads randomly selected items many times. In the case opposite, half of the papers were selected for multiple downloads and the other half were left alone.

At the extreme end of things, such behaviour would be **very easy to spot**: two completely isolated distributions emerge: one the result of human activity and the other of machine activity.

If left undetected, random attack would have a massive impact on both mean (up more than 6,291 per cent to 50,706!) and median usage (up from 324 to 10,152).

The next slide shows that this kind of random attack becomes progressively less easy to detect at lower levels than the 100,000 downloads per item shown here.

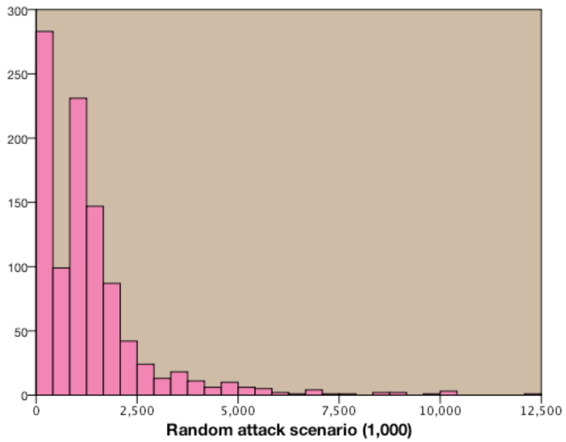


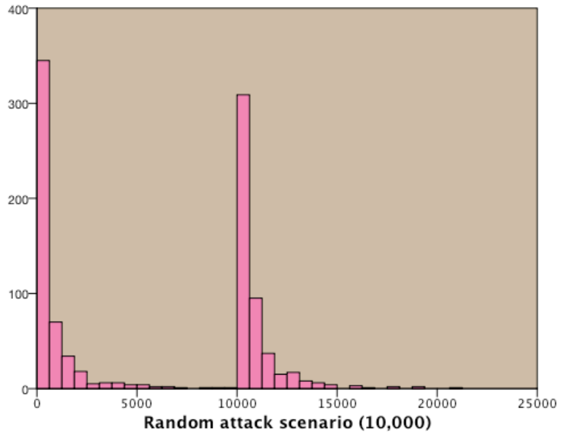
Gaming the journal usage factor

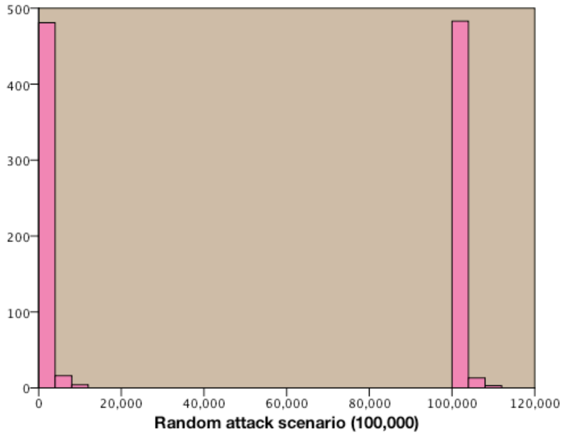
The three scenarios

Random attack

CIBER Research Limited







The figures above show that random attack is *very* easy to identify at the higher end (10,000 and 100,00 downloads). Very distinct patterns of human and machine activity are evident here.

At the 1,000 level, however, it is not really possible to spot this by eye alone: the distribution here does not look much different from unadulterated real use.

But random attack at this level (1,000) would still have a very dramatic effect on the usage factor (the median usage factor would rise from 324 to 1,324) and so we need a strategy. We need some statistical forensics!

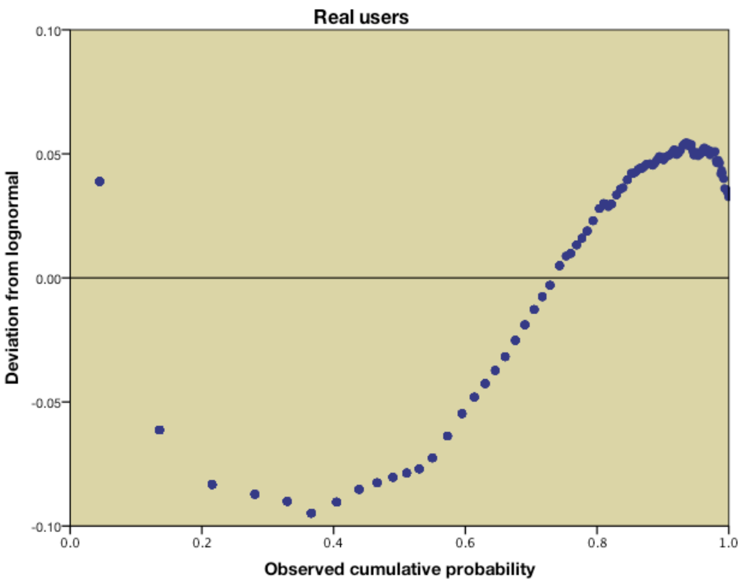
Gaming the journal usage factor

The three scenarios

Random attack

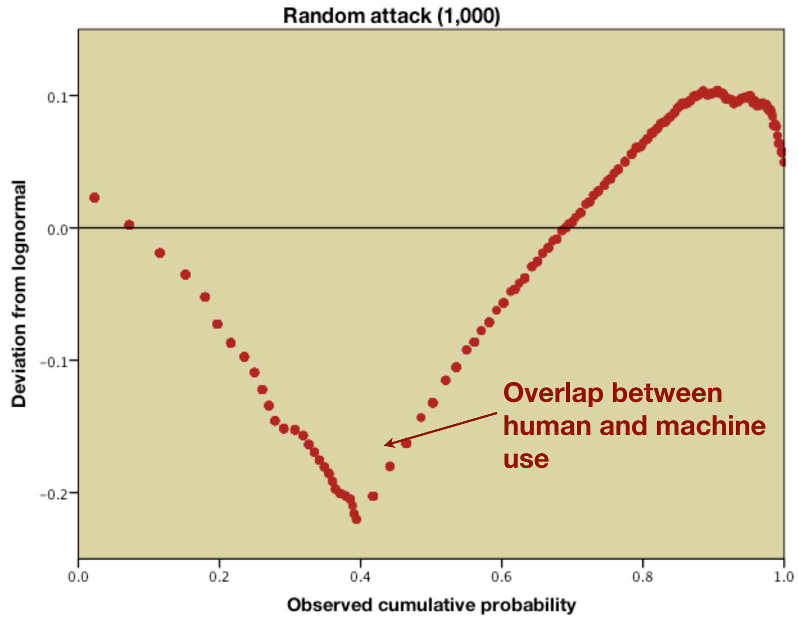
CIBER Research Limited

Real use
(expected behaviour)



This analysis compares real human use with a perfect lognormal distribution. The fit is approximate: most of the data fall on a straight line but humans tend to shun less propular items and to be drawn more strongly to popular items than a simple lognormal distribution would model.

Random attack
(observed behaviour)



In the figure on the previous page, it was very difficult to detect any difference between human and software agents at the 1,000 level. This analysis reveals a very clear break point between the two distributions and this approach might offer a useful forensic. This data is very different from the `natural' behaviour to the left. It looks *very* suspicious.

Gaming the journal usage factor

The three scenarios

Carpet bombing

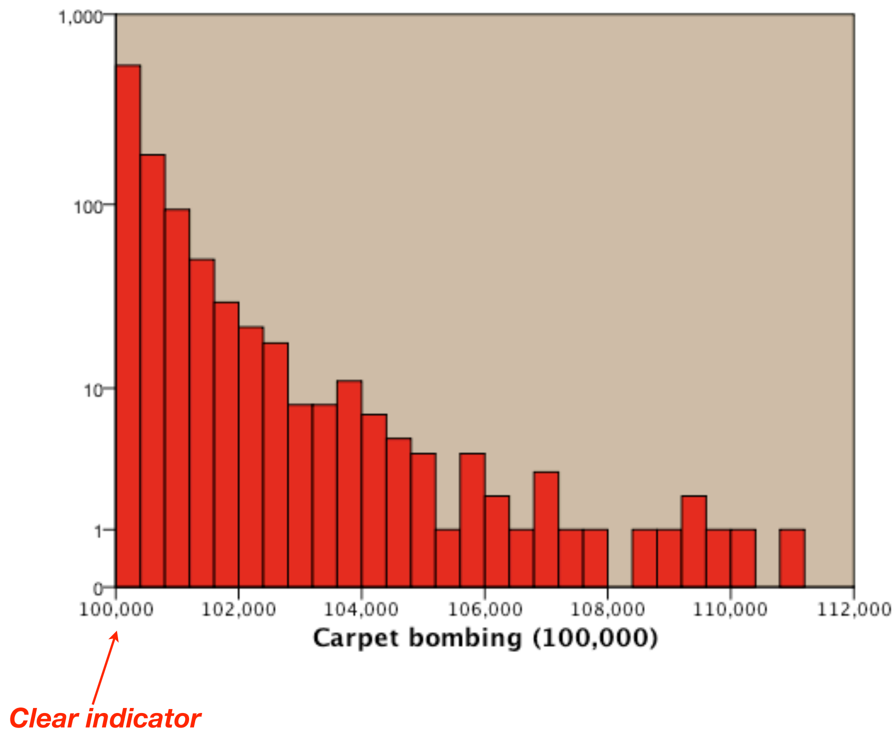
CIBER Research Limited

Carpet bombing is a scenario where someone develops a software agent that can fly under the radar and download *every* item *n* times. This would have a profound impact even at relatively low intensities.

If each item were downloaded a thousand times, the arithmetic mean would rise from its natural level of 806 to 1,806, inflating it by 224 per cent. This time, since every item is targeted, the median also rises (from 324 to 1,324).

Unlike the previous scenario, this time the fundamental shape of the curve stays the same - however few or however many additional machine downloads take place, so the same diagnostics will not be effective.

There is however a **clear indicator**. The distribution starts at 100,000. A journal that generated high levels of use for every article, and with no statistical noise below that level is simply not credible.



The implementation of the usage factor on a large scale would encourage the thought in some people's minds that they could extract some advantage by trying to game the system. This is inevitable and gaming issues are hardly unknown with respect to citation measures. But the perception will be that it is far easier to game a system which relies on simple acts of downloading than it is to write a paper with an unusually biased set of references and get it accepted. Ultimately, the credibility of the usage factor will stand or fall on the technical safeguards adopted: and these will need to move rapidly with new developments in spiders, bots and other software agents.

If these agents manage to fly under the radar, there are still mechanisms that can be used to filter and identify gross abuses of the system. We have coined the term `statistical forensics' in this report to refer to ways in which unusual patterns of non-human activity might be identified under three gaming scenarios. These three scenarios do not comprise a comprehensive list and many variants on these dishonest potential behaviours must be possible. But it strikes CIBER as unlikely that humans could really game the system to a significant extent. The `lone wolf' scenario is easily detectable, whether driven by human hands or by software. Human intervention alone would be a big ask for larger journals and would require many conspirators: and the deceit would have to be repeated, plus some, every year.

Random computer attacks or blanket bombing on any scale distort the fundamental rhythms that CIBER has observed for many years in usage data and they are again detectable by means of forensic statistics. This is by no means the end of the story, rather the beginning, but the gaming exercises here do begin to suggest that dishonest manipulation of the usage factor is a far more difficult problem than most people intuit when they are told how the factor is calculated.