EVALUATING THE USAGE AND IMPACT OF E-JOURNALS IN THE UK

JOURNAL SPENDING, USE AND RESEARCH OUTCOMES: A UK INSTITUTIONAL ANALYSIS

CIBER WORKING PAPER 2

November 2008

Contents

Tables and figures

Tables

Figures

Summary of key findings

1	UK universities and colleges spent in excess of £79.8m licensing electronic journals during 2006/07. The level of use of these resources is enormous: almost 102 million articles were downloaded that year, at an average (direct) cost of £0.81.	*Table 1, p.11 and Table 6. p. 12*
2	Spending on e-serials correlates strongly with use, as measured by COUNTER, ScienceDirect or Oxford Journals downloads. This effect is not a function of the size of the universities concerned, and is very unlikely to have arisen by chance.	*Table 7, p.12; Tables 8-9, p. 13*
3	This finding may seem intuitively obvious but (a) this relationship has not been demonstrated before, and (b) alternative hypotheses were possible. For example, big deals were not delivering, or libraries were primarily focused on historic collection building rather than consumer demand. Whatever the explanation for these relationships, the evidence points to an efficient match between supply and demand.	
4	Research-intensive universities spend more per library user on electronic serials and their users make much more intensive use of these resources than is the case for the new, largely teaching-led, institutions. As a result, the unit cost per download is significantly less.	*Table 11, p.15 and Tables 12-13, p.16-17*
5	Within the 115 universities examined, a group of 10 institutions stands out as `super users'. These universities consume very much more than the others and are all drawn from the research elite. They are characterized by high spend on e-serials and powerful research outcomes.	*Table 15, p.20*
6	Use of electronic journals, as measured by article downloads, correlates very strongly with three measures of research outcome: numbers of PhD awards, research grants and contracts income, and numbers of research papers published. This statistical relationship cannot easily be explained by chance and it is scale-independent. The relationships may be pure co-incidence but this does not seem intuitively plausible.	*Tables 16-21, pp 22-25*

7 CIBER has developed a model that describes the relationship between article consumption (COUNTER downloads) and `downstream' research outcomes. The model cannot be explained by chance, it fits the data very well, and it has strong predictive value.

Tables 22-24, pp 28-29.

8 Application of the model suggests that benefits (positive research outcomes) are strongly geared to spending on and use of electronic serials. These may even be quantifiable: the model suggests, for example, that a five per cent cut in e-serials budgets across the UK HEI sector would very seriously compromise this country's research capacity.

Tables 26-27, p. 31 and Figure 15. p.30

9 It is important to note that this report does not claim to show causal relationships between journal spending, use and research outcomes. The data are wide open to interpretation. The apparent links between these elements may be entirely independent of one another and their apparent correlation may simply be a co-incidence. Nor is much light shown on directionality. If the links between spending and use, for example, are real, is it the case that investment in e-resources helps to create demand, or that long running success in research creates the conditions that allow library budgets to flourish. This is not an arcane statistical question: are good libraries one of the pre-conditions for a university's success in research terms, or are they simply status symbols, invested in after the event, the academic equivalent of the salesman's new BMW paid for out of the Christmas bonus?

Context and scope

The aim of this working paper is to stimulate debate and further research into two critical questions for the future health, even the long-term viability, of academic libraries. Does spending on electronic serials represent good value for money, and what evidence do we have that use is in any way related to `downstream' research outcomes? These are deceptively simple questions but since they impact hard upon librarians, journal publishers, academic managers and the taxpayer alike, we feel they deserve a serious and concerted attempt to find new ways to express the beginnings of a way forward. Further work is already underway to probe these findings by means of a richer, longitudinal analysis.

Research design

The research method that generated the findings presented in this paper was simply a blending of widely available statistical information into a single unified database. CIBER compiled data on electronic serial use at British universities from three sources: the annual returns made to SCONUL for all article downloads using COUNTER standards (for the academic year 2006/07) together with data supplied in confidence by Elsevier Science Direct and Oxford Journals (for calendar 2006). In the case of Elsevier and OUP, downloads are indexed so that downloads by the single biggest institutional user are recorded as 100 and all other universities as a proportion of that level of use. This is simply to protect the market positions of these two publishers.

CIBER also included a range of other data from the SCONUL database for 2006/07 (see Annex B for a detailed description) such as total registered library users expressed as FTEs, numbers of electronic serial titles, and library spending on those resources.

Data from the Higher Education Statistics Agency (HESA) rounds off the picture by providing two key indicators of `downstream' research outcomes (for 2006/07): research grants and contracts income, and numbers of PhD awards. The SCOPUS database was mined for a final outcomes indicator: the number of research papers entering the international literature during the 2007 publication year. This was collected university by university as for all the other indicators. Only `citeable' document types were included, i.e. research articles, notes and reviews.

To our great surprise, no one seems to have thought of harvesting such easily available `low hanging fruit' before. CIBER's intention was simply to investigate what kinds of associations, if any, might be demonstrable between inputs to the research process (represented here by article downloads and electronic serials budgets) and research outcomes while controlling for a range of institutional factors, including size and tribal membership, such as Russell Group universities.

The source data were all harvested online and compiled into a single database using the SPSS (Statistical Package for the Social Sciences) application.

Limitations

While the data we use all come from reputable official sources, there are many limitations and qualifications that need to be aired. Firstly, the concept of `use' that we know to be a highly complex set of activities is reduced to a single variable, the article download. Secondly, we have had to take the published data at face value, we are simply in no position to be able to question its validity, and the data includes some major gaps, especially around SCONUL returns for the number of COUNTER downloads[1]. Thirdly, the exercise is largely atheoretical, we are simply looking for any patterning in the data at this stage not presenting a rounded vision of the (very long) journey that begins with an article reading and ends with a successful research proposal, publication or doctoral award. Our justification is that in this report we are simply taking an aerial view of the UK scholarly landscape at one point in time. The very fact that we are able to show some intriguing relationships between journal spend, journal use, and research outcomes perhaps shifts our thinking in a significant way: maybe there is something to be gained from looking at those questions about value for money, efficiency and effectiveness by looking at a whole national system and conceiving of it as a scholarly ecosystem.

[1] SCONUL data provides COUNTER-compliant downloads for 98 UK universities during 2006/07. Unfortunately, several very major players did not file this data with SCONUL (Cambridge, UCL and Nottingham, for example) and so after much careful thought and consultation, CIBER has used the expedient of drawing on the Missing Values Analysis function within SPSS (v14) to estimate COUNTER downloads for 17 missing institutions. The rationale for this is that the statistical associations between, for example, known COUNTER downloads and other measures were so strong that this seemed a viable way forward. Without it, major research players would have to be excluded from this report, blunting its `UK scholarly ecosystem' message. In deriving COUNTER estimates for the missing universities, *all* the existing variables in the database were used. The convention used in this report is that actual SCONUL data is labeled as `COUNTER raw' while CIBER's estimates are labeled `COUNTER adjusted'.

Journal spending and use

Data compiled from SCONUL statistical returns for 115 British universities show that almost £79.8m was spent on providing electronic serials during 2006/07 out of a total serials budget of £112.7m. Total article downloads for that period are estimated by CIBER at an astonishing 101.9m. On average each library user (students, faculty and other registered users) consumed 42 downloads at a direct cost of 81 pence each.

Within this broad canvas there is of course considerable variation: library users in the research-led Russell Group universities each consumed 84.3 downloads, compared with 30.2 for the `new' universities, and made considerably greater user per title than the rest of the sector: 213.2 downloads per title for the Russell Group, 117.0 for the new universities.

The major finding in this section however relates to the close statistical correlation between spending on electronic serials and their use. This is perhaps unsurprising, but we do not believe that anyone has attempted to look at this association before and it should not be taken for granted. Indeed it is conceivable to marshal arguments that suggest otherwise: for instance that large-scale bundling of titles might be loosening the bonds between consumer demand and provision, or that libraries had a responsibility to maintain long historic runs of material at the same time as satisfying current demand.

The evidence from Figure 1 and Tables 7-9 show that electronic serials budgets are very much in line, quite literally, with use. Strong positive linear correlations are shown in the case of COUNTER, Elsevier ScienceDirect and Oxford Journals and these are associations are statistically significant at the 1 per cent level – in other words it is exceptionally unlikely that they have arisen by chance. There is a size effect here: one would expect larger universities to spend more and to consume more electronic product, all things being equal. For this reason, we have tried to discount effects due to scale by controlling for total FTE users, academic staff FTE and total institutional expenditure, and even for the presence or absence of a medical school (Tables 7-9). Significant associations still persist.

It is very important to be clear that these findings do not provide any direct evidence of cause and effect. Indeed, it is at least theoretically possible that e-journals spend and use are purely co-incidental, or that they are independent of one another but jointly associated with a third dependent variable. Even if cause and effect could be demonstrated, the thorny issue of direction remains: does spending on resources drive use, or is spending a function of expressed consumer demand? Or a bit of both?

We leave these judgments to the reader of this report: CIBER's view is that although more work is needed, a strong prima facie case exists here, at least for the proposition that rational purchasing decisions are being taken. All we can say for the moment is that the doomsday hypothesis, that spending and use are randomly distributed is clearly not the case. At the UK national level, it would appear that the journals system is working efficiently and that resources are being deployed where they are needed.

Tables and graphics

Table 1: Electronic serial subscriptions and article downloads (COUNTER): descriptive statistics, 2006/07 (UK universities)

	n	Huber's M[2]	Mean	S.D.	Sum
Electronic serial subscriptions	115	£598,616	£856,974	864,250	£76,378,184
COUNTER article downloads [SCONUL]	98	480,070	664,158	638,711	83,983,439
COUNTER article downloads [CIBER estimate]	115	627,177	885,908	897,664	101,879,432

Table 2: COUNTER downloads (adjusted) per registered library user, 2006/07 (*n*=115 UK universities)

	95% confidence intervals					
Mean	*lower bound*	*upper bound*	Huber's M	St. dev.	Minimum	Maximum
47.2	41.3	53.1	42.0	31.8	0.6	149.0

Table 3: COUNTER downloads (adjusted) per registered library user by university sector, 2006/07 (*n*=112 UK universities)

			95% confidence intervals				
	n	Mean	*lower bound*	*upper bound*	St. dev.	Minimum	Maximum
Russell Group	21	84.3	75.0	93.6	20.5	41.1	140.4
Old universities (excl. above)	38	50.1	40.3	60.0	30.0	5.5	149.0
New universities	53	30.2	25.3	35.1	17.9	1.4	84.5

ANOVA F=41.59, d.f=2, the difference between sectors is significant at the 1% level

Table 4: COUNTER downloads (adjusted) per title, 2006/07 (*n*=115 UK universities)

	95% confidence intervals					
Mean	*lower bound*	*upper bound*	Huber's M	St. dev.	Minimum	Maximum
132.8	107.1	158.6	106.5	139.1	1.0	1,133.0

[2] Huber's-M is an estimation of the median value that takes account of highly skewed distributions.

Table 5: COUNTER downloads (adjusted) per title, 2006/07 by university sector (*n*=112 UK universities)

			95% confidence intervals				
	n	Mean	*lower bound*	*upper bound*	St. dev.	Minimum	Maximum
Russell Group	21	213.2	160.4	266.1	116.0	42	568
Old universities (excl. above)	38	111.2	91.6	142.2	77.0	5	319
New universities	53	116.9	65.3	157.1	166.4	3	1,113

ANOVA F=4.81, d.f.=2, no difference between the sectors (p=0.1)

Table 6: Direct cost per COUNTER download (adjusted) by university sector, 2006/07 (*n*=112 UK universities)

	n	Huber's M
Russell Group	21	£0.69
Old universities (excl. above)	38	£1.04
New universities	53	£0.69
All universities	112	£0.80

ANOVA F=2.18, d.f.=2, no difference between the sectors

Table 7: Correlation between total COUNTER downloads (adjusted) and electronic serial spend, 2006/07 (*n*=115 UK universities)

	Bivariate correlation (no control)	*Partial correlation, controlling for FTE library users*	*Partial correlation, controlling for FTE academic staff*	*Partial correlation, controlling for total institutional expenditure*	*Partial correlation, controlling for presence of a medical school*
Spearman	0.809**	-	-	-	-
Pearson	-	0.730**	0.346**	0.351**	0.642**

***Correlations are significant at the 1% level (2-tailed)*

Table 8: Correlation between Elsevier ScienceDirect downloads and electronic serial spend, 2006/07 (*n*=115 UK universities)

	Bivariate correlation (no control)	*Partial correlation, controlling for FTE library users*	*Partial correlation, controlling for FTE academic staff*	*Partial correlation, controlling for total institutional expenditure*	*Partial correlation, controlling for presence of a medical school*
Spearman	0.796**	-	-	-	-
Pearson	-	0.727**	0.280**	0.269**	0.632**

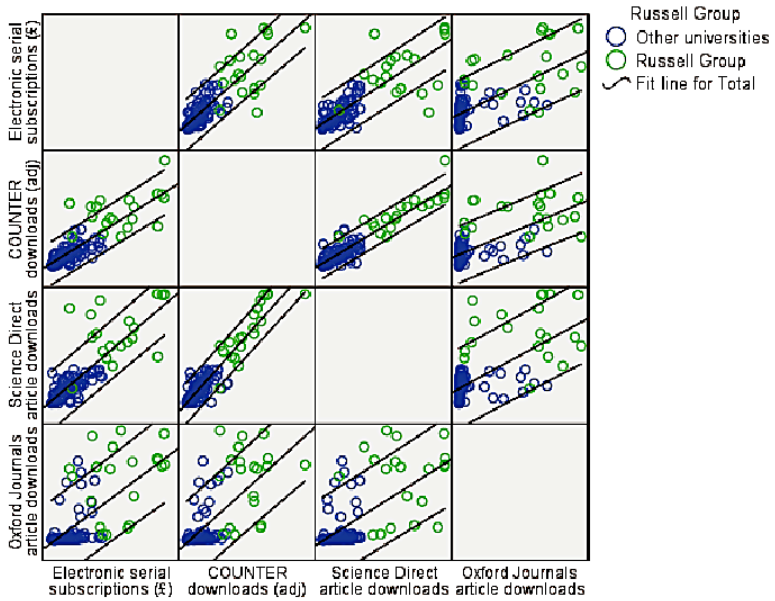
***Correlations are significant at the 1% level (2-tailed)*

Table 9: Correlation between Oxford Journals downloads and electronic serial spend, 2006/07 (*n*=115 UK universities)

	Bivariate correlation (no control)	*Partial correlation, controlling for FTE library users*	*Partial correlation, controlling for FTE academic staff*	*Partial correlation, controlling for total institutional expenditure*	*Partial correlation, controlling for presence of a medical school*
Spearman	0.771**	-	-	-	-
Pearson	-	0.570**	0.282**	0.214*	0.420**

**Correlations are significant at the 1% level (2-tailed)
*Correlation is significant at the 5% level (2-tailed)

Figure 1: Electronic serial spend, COUNTER (adjusted), Elsevier ScienceDirect, and Oxford Journals downloads: matrix scatterplot (fitted linear regression trendlines with 95% confidence intervals)



Patterns of usage

In this section, we will be looking a little more closely at patterns of use before turning to the 64 thousand dollar question, is there any evidence of a link between use and research outcomes?

Table 11 is interesting in that it demonstrates very clear blue water between teaching-led and research-led universities in respect of the amount of resource they deploy per library user in procuring electronic journals. The differences are massive.

Russell Group universities not only spend more per user (Table 10), but they use more (Table 12). A consequence is that the unit cost per download is much cheaper (Table 6). This takes us back to the issue of directionality. Assuming that the links between use and spend are real, rather than co-incidental, are the new universities held back in research terms by relatively poor provision, or is there simply not enough demand pull through to encourage them to invest more?

CIBER was once posed a question by a senior journal publisher which we now feel a little more able to answer. Are super users of journal materials also super producers?

To approach this problem, we divided all 115 universities into groups on the basis of their reported COUNTER downloads for 2006/07 using the estimations for missing cases explained earlier. The groups were identified on the basis of standard deviations from the mean for the whole population. `Moderate users' are defined as those institutions whose total downloads fall one or two standard deviations below the mean: `High users' fall into the two standard deviations above the mean, while `Super users' are those whose total download activity is more than two standard deviations above the mean. The universities are detailed in Table 16.

Tables and graphics

Table 10: Library serial performance indicators: descriptive statistics, 2006/07 (UK universities)

	n	Median	Mean	S.D.	Range
Electronic serial spend per user	115	£30.27	£37.51	29.79	223.45
Inter-library loan requests per user	115	0.29	0.34	0.26	2.10
Journals as a % of university spend	115	0.52%	0.53%	0.18	0.93
Article downloads per title	98	100.26	132.51	143.50	1,131.65
% of serial titles available in e-format	115	87.3%	83.6%	11.7	68.2
Electronic serial titles per user	115	0.43	0.56	0.42	2.04

Table 11: Electronic serial spend per library user by university sector, 2006/07 (*n*=112 UK universities)

		95% confidence interval for mean	
	Mean	*lower*	*upper*
Russell Group	£58.44	£46.39	£70.49
Other old universities	£52.20	£40.57	£63.82
New universities	£19.50	£16.82	£22.19
Sector	**£37.90**	£32.28	£43.51

ANOVA F=28.88, d.f.=111, p<0.001

Figure 2: Electronic serial spend per library user by university sector, 2006/07

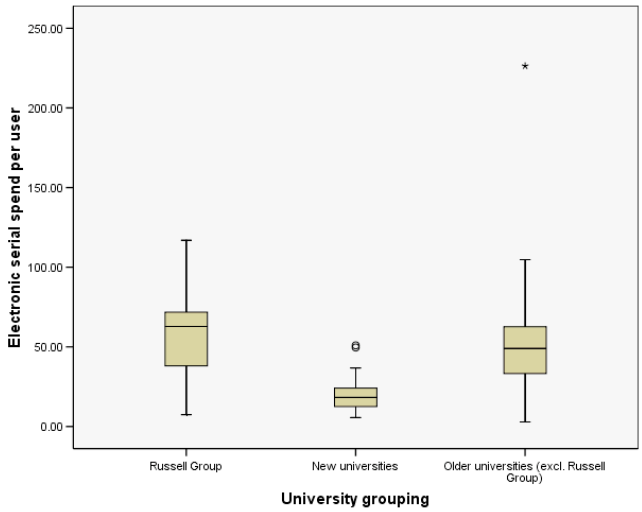


Table 12: Article downloads (COUNTER adjusted) per library user by university sector, 2006/07
(*n*=115 UK universities)

		95% confidence interval for mean	
	Mean	*lower*	*upper*
Russell Group	83.2	71.6	94.8
Other old universities	51.5	40.8	62.1
New universities	30.5	25.0	35.9
Sector	**47.0**	40.9	53.2

ANOVA F=31.52, d.f.=95, p<0.001

Figure 3: Article downloads (COUNTER adjusted) per library user by university sector, 2006/07

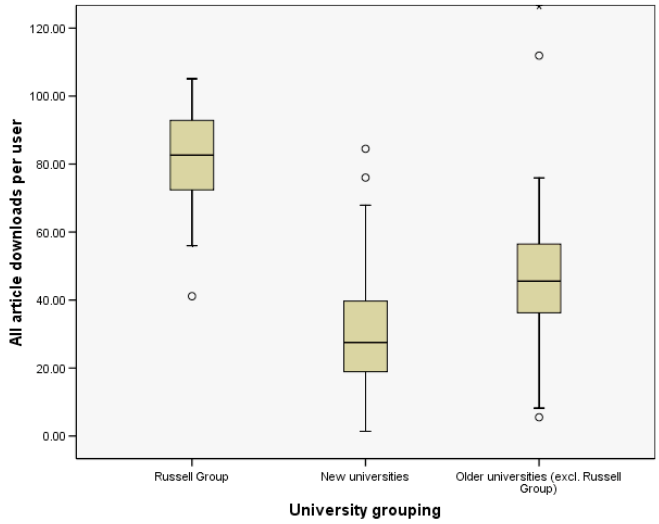


Figure 4: Electronic serial spend per article download (COUNTER adjusted) by university sector, 2006/07

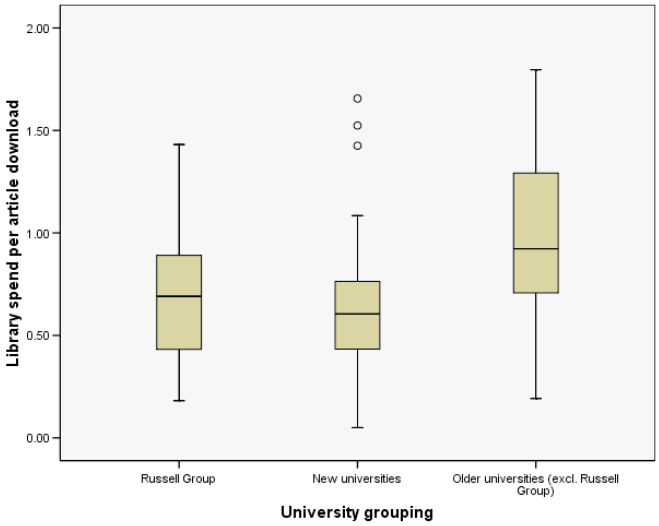


Table 13: UK universities by use intensity grouping (*n*=115 UK universities)

	n
Moderate users	80
High users	25
Super users	10

Table 14: List of UK universities by use intensity

Moderate users (80)

Abertay
Aberystwyth
Anglia Ruskin
Aston
Bangor
Bath
Bedfordshire
Birkbeck
Bolton
Bournemouth
Bradford
Brighton
Brunel
Canterbury Christ Church
Chester
City
Coventry
Cranfield
De Montfort
Derby
Dundee
East London
Edge Hill
Essex
Glamorgan
Glasgow Caledonian
Gloucestershire
Goldsmiths
Greenwich
Heriot-Watt
Hertfordshire
Hull
Institute of Education
Keele
Kent
Kingston
Leeds Metropolitan
Lincoln
Liverpool Hope
London Metropolitan
London School of H&TM
Loughborough
Manchester Metropolitan
Middlesex
Napier
Northampton
Nottingham Trent
Oxford Brookes
Paisley
Plymouth
Portsmouth
Queen Margaret
Reading
Robert Gordon
Roehampton
Royal Holloway
Salford
School of Oriental & African Studies
Sheffield Hallam
South Bank
Southampton Solent
St Andrews
Stirling
Sunderland
Sussex
Swansea

Teesside
Thames Valley
Ulster
University of Wales Institute Cardiff
University of the Arts
University of Wales College Newport
West of England
Westminster
Winchester
Wolverhampton
Worcester

High users (25)

Aberdeen
Bath Spa
Bristol
Cardiff
Central Lancashire
Durham
East Anglia
Exeter
Glasgow
Huddersfield
Lancaster
Leicester
Liverpool
Liverpool John Moores
London School of Economics
Newcastle
Northumbria
Queen Mary London
Queen's Belfast
Sheffield
Southampton
Strathclyde
Surrey
Warwick

Super users (10)

Birmingham
Cambridge
Edinburgh
Imperial College
King's College London
Leeds
Manchester
Nottingham
Oxford
University College London

Table 15: Characteristics of UK universities by use intensity, analysis of variance, 2006/07
(*n*=115 UK universities)

		Mean	95% confidence interval for mean		ANOVA	
			lower	*upper*	*F*	*Sig.*
Electronic serial spend per user	Moderate users	£29.79	£23.53	£36.04	10.24	.000
Significant difference between groups	High users	£50.89	£37.37	£64.41		
	Superusers	£58.93	£45.93	£71.91		
Journals as a % of university spend	Moderate users	0.50%	0.46%	0.54%	3.31	.040
Significant difference between groups	High users	0.60%	0.51%	0.68%		
	Superusers	0.59%	0.52%	0.66%		
Article downloads per title	Moderate users	120.6	81.3	159.8	2.27	.109
No difference between groups (very large standard deviations are evident within this data)	High users	123.2	90.2	156.2		
	Superusers	214.1	164.2	264.1		
% of serial titles available in e-format	Moderate users	84.6%	81.7	87.5%	1.13	.326
No difference between groups	High users	82.6%	78.9	86.4%		
	Superusers	79.9%	75.6	84.3%		
Library spend per article downloaded	Moderate users	£0.95	£0.69	£1.22	0.37	.690
No difference between groups	High users	£0.91	£0.72	£1.11		
	Superusers	£0.70	£0.53	£0.88		
Electronic serial titles per user	Moderate users	0.58	0.48	0.67	0.20	.819
No difference between groups	High users	0.51	0.32	0.70		
	Superusers	0.54	0.31	0.77		

Reference

Judy Luther (2008). *University Investment in the Library: What's the Return? A Case Study at the University of Illinois at Urbana-Champaign.* San Diego, CA: Elsevier Library Connect, White Paper Number 1.

Journal use and research outcomes

The inability of the library community to clearly demonstrate a positive return on investment is a defining theme at this point in the profession's fortunes. Studies are few and far between, and often highly abstract and theoretical, or unconvincing and unconvincing (e.g. Luther 2008).

CIBER's approach for this RIN study has been to tackle the problem simultaneously top-down, using easily available official statistics, and bottom-up using case studies. In both instances, we have been working with unusually comprehensive data sets and with a minimum of a priori hypothesizing.

In this phase of the report, we simply link three measures of research outcome: numbers of PhD awards, research grants and contracts income (from HESA sources) and numbers of research papers entering the international literature (from SCOPUS) with useage data (from SCONUL, Elsevier and Oxford Journals), all compiled at the institutional level.

A remarkably consistent picture emerges. Journal use is positively correlated with all three research outcomes and linear modeling suggests that (a) the statistical association is both very significant, and (b) that our models offer a good fit using the R-squared indicator, at least for all article downloads (COUNTER) and those from the ScienceDirect platform (Tables 16-18). Again, these associations appear to be scale-independent, at least when normalized per academic FTE. In the case of Oxford Journals, we find that the relationship between use and research outcomes is significant but the data is more scattered and offers less goodness-of-fit. This is perhaps explained by the comparatively small set of journals offered on that platform (around 230) and their particular subject characteristics, with a strong emphasis on medical titles.

Using our earlier groupings, it becomes clear that intensity of use is a very good predictor of research outcomes and that high and, especially, super users seem to reap disproportionately high rewards (Table 21).

Again, we need to remind ourselves that these correlations may be co-incidental and that cause and effect should not be imputed to our findings at this early stage. There are clearly many other factors at play here but what we can however say with great confidence is that the correlations cannot easily be explained away by chance, nor that they are simply a function of institutional size. While not wishing to advance `no smoke without fire' arguments, we do wonder whether these findings might give cost cutting academic administrators pause for thought! We will return to the question of these correlations in the next section, where we present a rather more sophisticated model.

Tables and graphics

Table 16: COUNTER article downloads (raw) and three research outcome measures: bivariate correlation, 2006/07 (*n*=86 UK universities, outliers excluded)

		Papers per academic	RGC income per academic	PhD awards per 100 academic FTE
COUNTER	Spearman's rho	0.443	0.320	0.301
	Significance (2-tailed)	<0.01	<0.01	<0.01
	R^2 goodness-of-fit	0.771	0.610	0.645

Table 17: COUNTER article downloads (adjusted) and three research outcome measures: bivariate correlation, 2006/07 (*n*=109 UK universities, outliers excluded)

		Papers per academic	RGC income per academic	PhD awards per 100 academic FTE
COUNTER	Spearman's rho	0.619	0.565	0.569
	Significance (2-tailed)	<0.01	<0.01	<0.01
	R^2 goodness-of-fit	0.751	0.648	0.683

Figure 5: Correlations between COUNTER article downloads (adjusted) and three research outcome measures

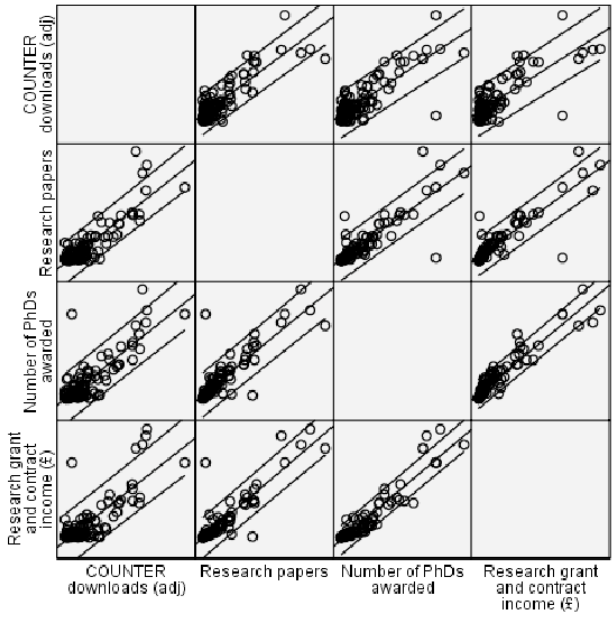


Table 18: Elsevier ScienceDirect article downloads and three research outcome measures: partial correlation controlling for number of library users, 2007 (n=110 UK universities, outliers excluded)

		Research papers	RGC income	PhD awards
ScienceDirect	Pearson coefficient	0.691	0.599	0.235
	Significance (2-tailed)	<0.01	<0.01	<0.01
	R^2 goodness-of-fit	0.839	0.722	0.730

Figure 6: Correlations between Elsevier ScienceDirect article downloads and three research outcome measures

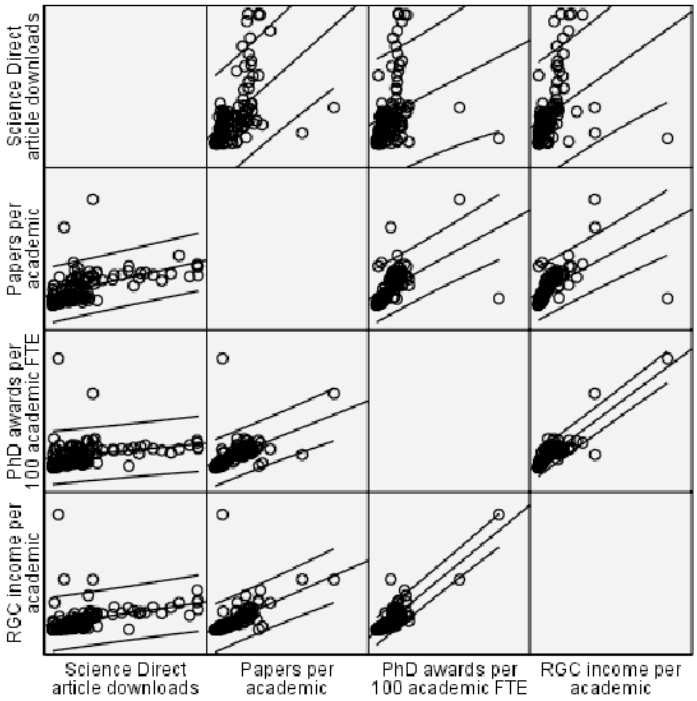


Table 19: Oxford Journals article downloads and three research outcome measures: partial correlation controlling for number of library users, 2007 (n=109 UK universities, outliers excluded) (*n*=109)

		Research papers	RGC income	PhD awards
Oxford Journals	Pearson coefficient	0.512	0.480	0.433
	Significance (2-tailed)	<0.01	<0.01	<0.01
	R^2 goodness-of-fit	0.516	0.427	0.448

Figure 7: Correlations between Oxford Journals article downloads and three research outcome measures

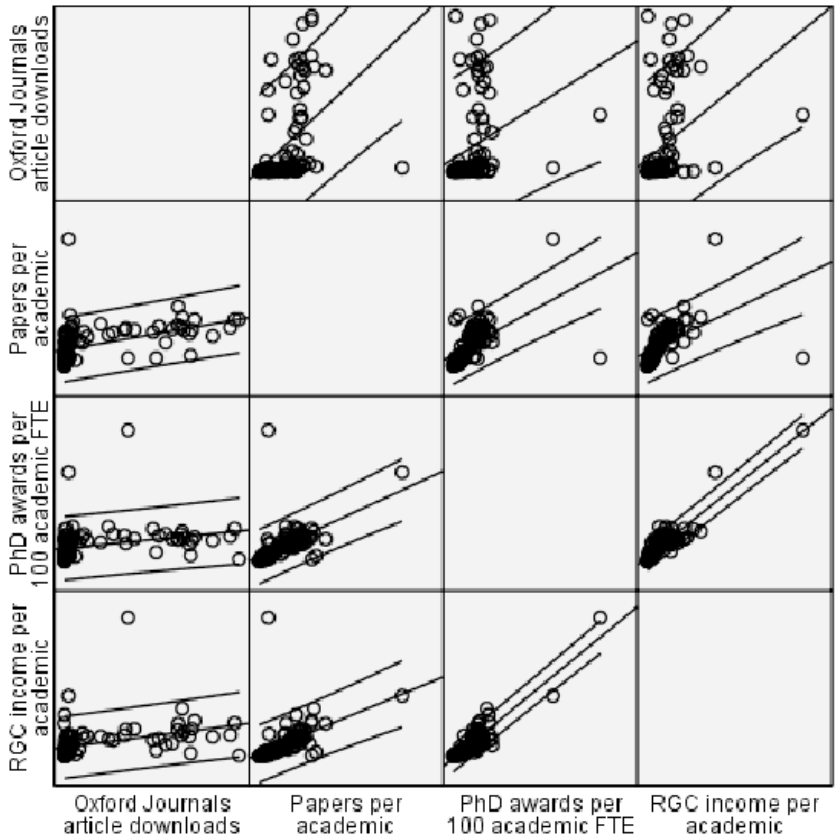


Table 20: Electronic serial spend per library user and three research outcome measures: bivariate correlation, 2006/07 (*n*=115 UK universities, outliers excluded)

		Research papers per academic	RGC income per academic	PhD awards per 100 academics
Electronic serial spend per user	Spearman's rho	0.599	0.682	0.442
	Significance (2-tailed)	<0.01	<0.01	<0.01
	R^2 goodness-of-fit	0.370	0.195	0.048

Table 21: Four research outcome measures by use intensity grouping: analysis of variance, 2006/07 (*n*=115 UK universities)

		Mean	95% confidence interval for mean		ANOVA	
			lower	*upper*	*F*	*Sig.*
Times research rating	Moderate users	2.2	1.8	2.6	26.86	.000
Significant difference between groups	High users	3.8	3.1	4.6		
	Superusers	5.1	4.7	5.4		
Research papers per academic	Moderate users	0.4	0.3	0.5	16.78	.000
Significant difference between groups	High users	0.8	0.5	1.1		
	Superusers	1.0	0.9	1.2		
RGC income (£000s) per academic	Moderate users	12.7	6.6	18.9	9.32	.000
Significant difference between groups	High users	29.0	18.7	39.4		
	Superusers	39.7	31.1	48.3		
PhD awards per 100 academics	Moderate users	9.1	6.3	11.8	6.05	.003
Significant difference between groups	High users	17.5	10.9	24.1		
	Superusers	17.4	14.6	20.1		

Figure 8: Times research rating by use intensity grouping

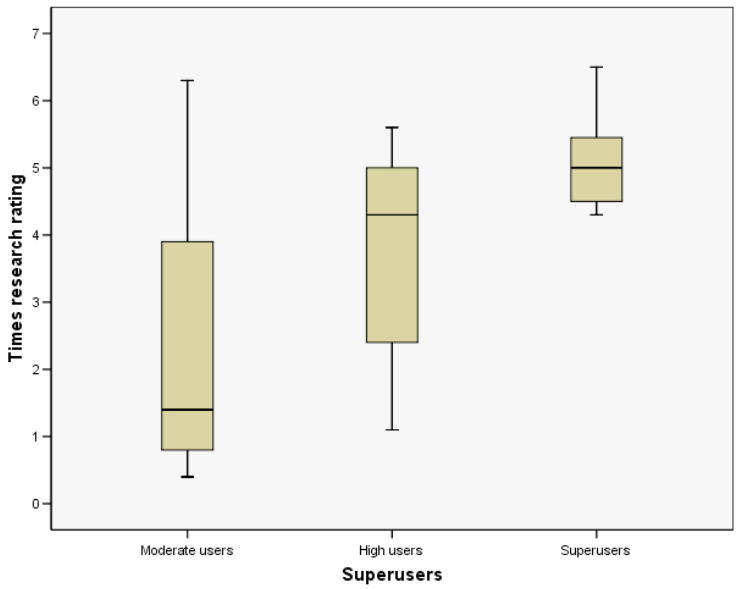


Figure 9: Papers per academic by use intensity grouping

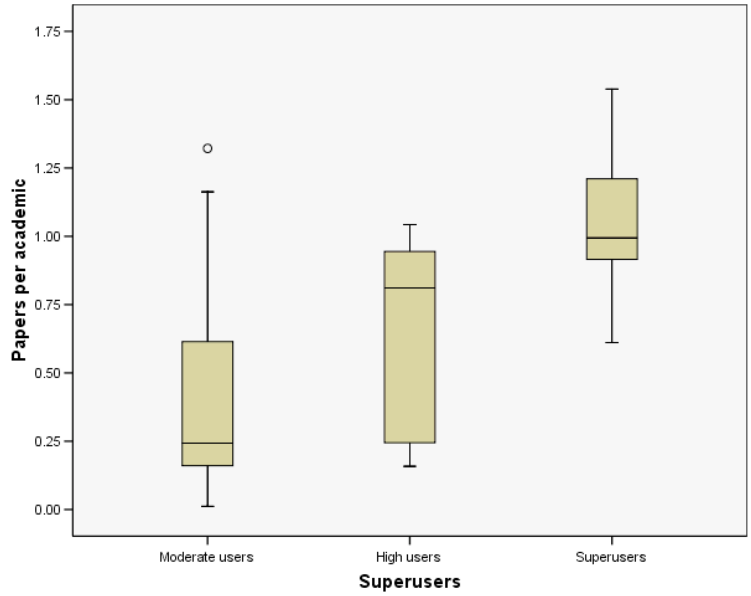


Figure 10: Research grant and contract income by university usage grouping

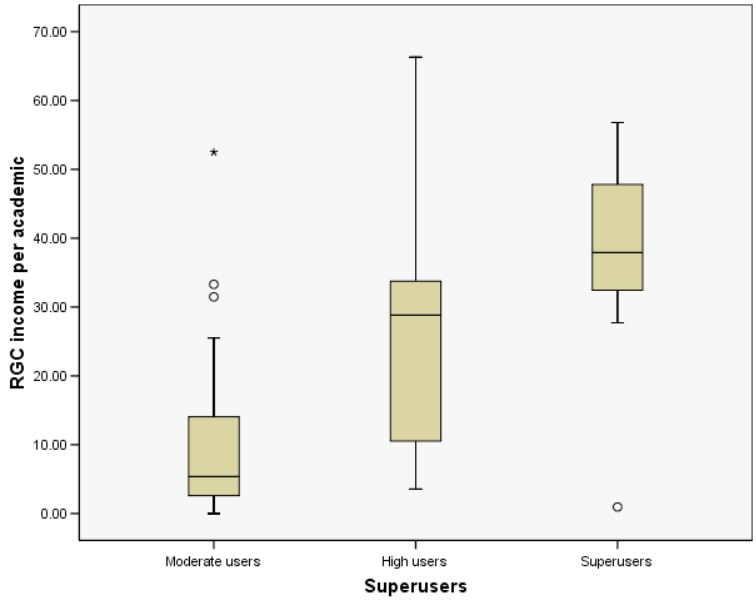
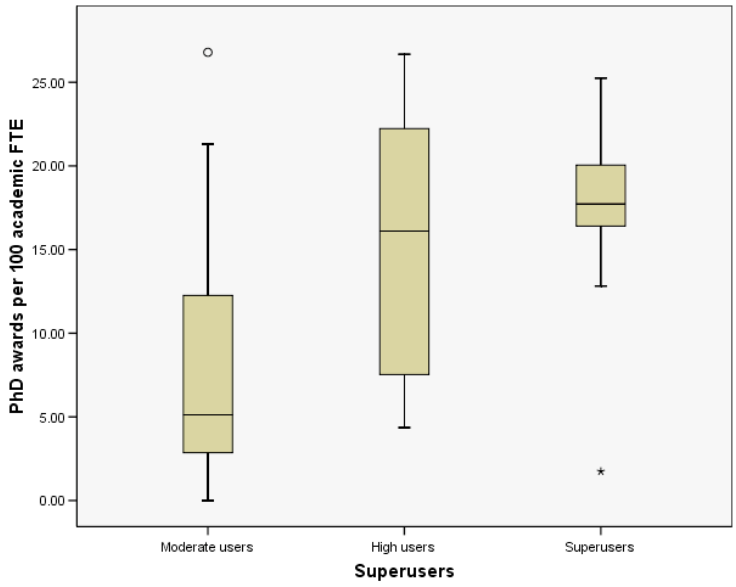


Figure 11: PhD awards per 100 academics by use intensity grouping



Modelling journal spend, usage and research outcomes

In this final section we attempt another first. The starting point is that we know from earlier in this report that journal use correlates strongly with research outcomes – there is plenty of evidence of good and significant fits. COUNTER downloads (adjusted by CIBER) looks the most promising candidate for a more sophisticated model: after all, it reflects use of all e-journals at each university.

In this experiment, CIBER modeled the relationship between COUNTER downloads and three sets of research outcomes (research papers, PhD awards, research grant and consultancy income). A range of curves were fitted to the data (using SPSS curve estimation). The best fit, by some distance, was a cubic model:

$y = b_0 + (b_1x) + (b_2x^2) + (b_3x^3)$

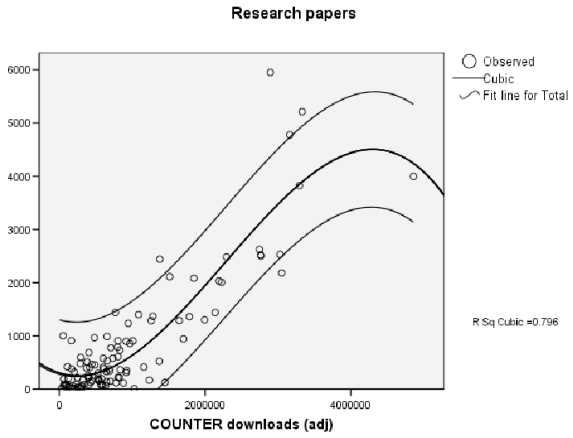
This offers a better description of the relationship between article consumption (as measured by downloads) and research outcomes and it is more credible than others, especially linear or power models in that the cuboid model becomes saturated. The parameter estimates are given in the next three tables.

The associated graphs show an excellent fit (all show high R^2 values, low probability – significant at the 1 per cent level, and few outliers).

Table 22: Curve estimation for research papers, cubic model (*n*=115 UK universities)

Cubic model		*Parameter estimates*				
R-squared	Independent variable:	Sig.	Constant	b1	b2	b3
0.796	COUNTER downloads (adj)	<0.01	304.58	<0.01	<0.01	<0.01

Figure 12: Cubic model for research papers



Note: The graphs show the fitted cubic model with 95% confidence intervals

Table 23: Curve estimation for PhD awards, cubic model (*n*=115 UK universities)

Cubic model		*Parameter estimates*				
R-squared	Independent variable:	Sig.	Constant	b1	b2	b3
0.721	COUNTER downloads (adj)	<0.01	76.19	<0.01	<0.01	<0.01

Figure 13: Cubic model for PhD awards

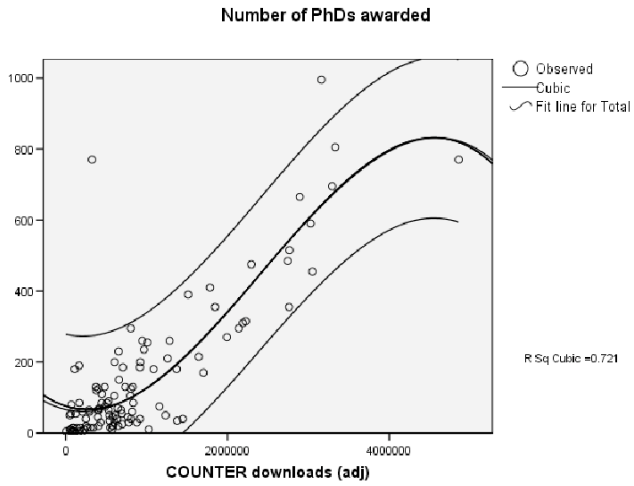
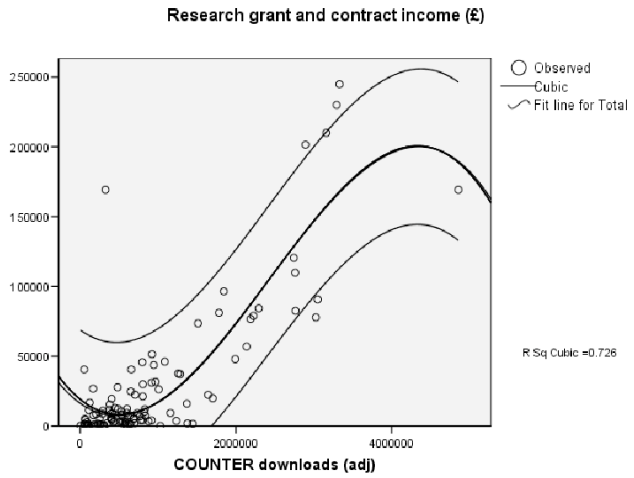


Table 24: Curve estimation for research grants and contracts (*n*=115 UK universities)

Cubic model		*Parameter estimates*				
R-squared	Independent variable:	Sig.	Constant	b1	b2	b3
0.726	COUNTER downloads (adj)	<0.01	19,189.40	-0.05	<0.01	<0.01

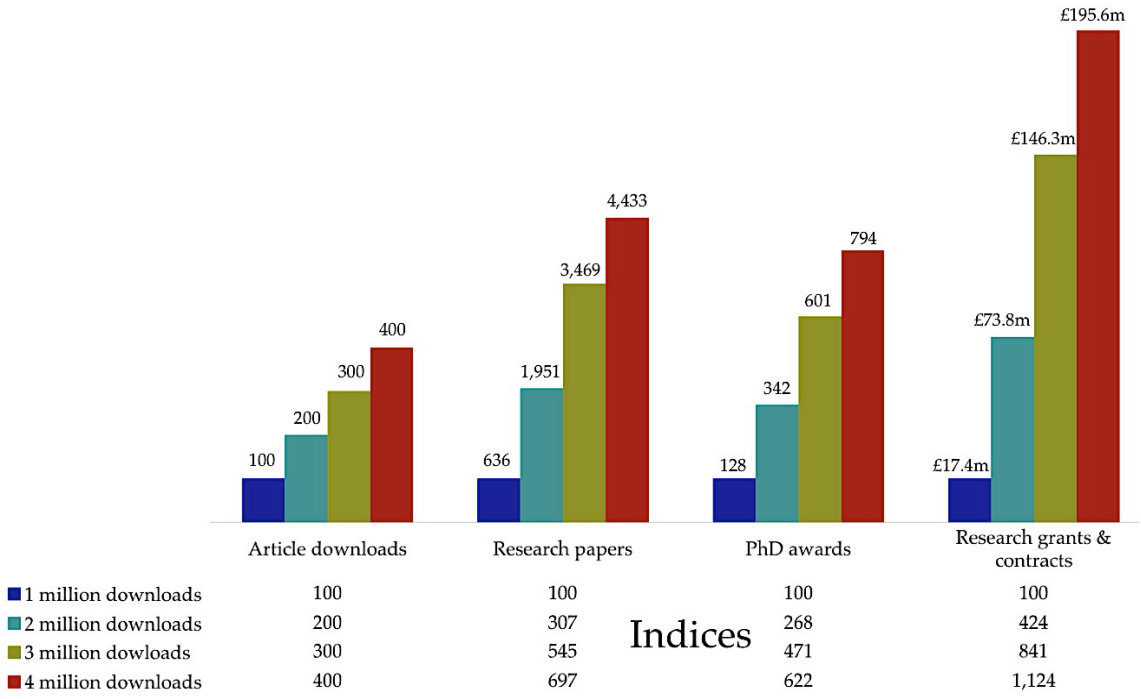
Figure 14: Cubic model for research grants and contacts income



The graphic below shows the model in action. What are the associated increases in research outcomes (in terms of the logic of model) for four hypothetical universities who download, 1, 2 3 or 4 million articles each year?

The data are presented as an index (1 million downloads = 100) and the actual values are shown on the bar.

Figure 15: Modeling the association between article downloads (COUNTER adjusted) and research outcomes, cubic model



Just to reassure us all that the model is robust, the next Table shows CIBER predictions (using the statistical model) and the actual values obtained for three groups of universities from the raw data supplied by HESA and SCONUL. The fit is very close and suggests that we can place a high degree of confidence in the model.

Table 25: Fit between model predictions and actuals

		Numbers of research papers		*Numbers of PhDs awarded*		*Research grants and contracts (£000s)*	
	n	CIBER prediction	Actual	CIBER prediction	Actual	CIBER prediction	Actual
Moderate users	80	263	250	66	56	7,518	5,602
High users	25	1,079	1,157	200	222	35,078	36,202
Super users	10	3,528	3,330	612	623	149,197	152,384

Why go to this much trouble? Well, if the model is any good, and it looks to be, we could use it to test out the implications of an across-the-board cut in e-serials budgets, as in the next Table. Cause and effect cannot be ascribed to the model, but it does represent a kind of `wisdom of crowds': 112 UK universities that have probably settled on a fairly stable set of

relationships with their faculty, students and paymasters. The system has an inertia and dynamics of its own.

Table 26: Modelling a five per cent real terms *cut* in e-serial budgets, sectoral aggregates

	Cost savings	**Disbenefits**		
	E-serials budget (£millions)	*Research papers*	*PhD awards*	*RGC income (£millions)*
Russell Group (n=21)	+£1.7m	-3,914	-658	-£185.0m
Other old universities (n=38)	+£1.2m	-633	-105	-£12.8m
New universities (n=53)	+£1.3m	-336	-58	-£3.3m
Whole sector	**+£4.2m**	**-4,883**	**-831**	**-£201.1m**

This exercise suggests that the savings of £4.2m that would be achieved by cutting university e-serials by five per cent in real terms, might be achieved at serious cost to the fabric of UK research. The effects would seem likely to fall very, very hard indeed on the top research universities and to generally be counter-productive if, as the model possibly suggests, library services and income generation are causally entwined (see recent Elsevier White Paper).

Reversing the logic, in the next Table, we can see similar but slightly larger effects (the model is a curve, not a straight line), if e-serials budgets were to be increased by five per cent across the board.

Table 27: Modelling a five per cent real terms *increase* in e-serial budgets, sectoral aggregates

	Additional cost	**Benefits**		
	E-serials budget (£millions)	*Research papers*	*PhD awards*	*RGC income (£millions)*
Russell Group (n=21)	-£1.7m	+3,889	+658	+£185.3m
Other old universities (n=38)	-£1.2m	+679	+112	+£15.4m
New universities (n=53)	-£1.3m	+371	+64	+£1.4m
Whole sector	**-£4.2m**	**+4,939**	**+843**	**+£202.0m**

This is an artificial exercise in all sorts of ways, but it does offer more of an illustrated sense of how journal use and research outcomes are intertwined in a statistical if not an immediately explicable way.

Three more models are included as Annex C to this report using different assumptions (for example, normalization for size of institution). CIBER's intention is to stimulate debate and further work in this new area, which looks to be very promising indeed.

Annex A

List of universities

Aberdeen
Abertay
Aberystwyth
Anglia Ruskin
Aston
Bangor
Bath
Bath Spa
Bedfordshire
Birkbeck
Birmingham
Bolton
Bournemouth
Bradford
Brighton
Bristol
Brunel
Cambridge
Canterbury Christ Church
Cardiff
Central Lancashire
Chester
City
Coventry
Cranfield
De Montfort
Derby
Dundee
Durham
East Anglia
East London
Edge Hill
Edinburgh
Essex
Exeter
Glamorgan
Glasgow
Glasgow Caledonian
Gloucestershire
Goldsmiths
Greenwich
Harper Adams
Heriot-Watt
Hertfordshire
Huddersfield
Hull
Imperial College
Institute of Education
Keele
Kent
King's College London
Kingston
Lancaster
Leeds
Leeds Metropolitan
Leicester
Lincoln
Liverpool

Liverpool Hope
Liverpool John Moores
London Metropolitan
London School of Economics
London School of Hygiene and Tropical Medicine
Loughborough
Manchester
Manchester Metropolitan
Middlesex
Napier
Newcastle
Northampton
Northumbria
Nottingham
Nottingham Trent
Open
Oxford
Oxford Brookes
Paisley
Plymouth
Portsmouth
Queen Margaret
Queen Mary London
Queen's Belfast
Reading
Robert Gordon
Roehampton
Royal Holloway
Salford
School of Oriental & African Studies
Sheffield
Sheffield Hallam
South Bank
Southampton
Southampton Solent
St Andrews
Stirling
Strathclyde
Sunderland
Surrey
Sussex
Swansea
Swansea Institute
Teesside
Thames Valley
Ulster
University College London
University of Wales Institute Cardiff
University of the Arts
West of England
University of Wales College Newport
Warwick
Westminster
Winchester
Wolverhampton
Worcester
York

Annex B

Indicators used

Dependent variables

Code	Variable	Notes	Source	Date	Units
COUNTER	Article downloads	COUNTER-compliant statistics for all journals (raw and adjusted versions)	SCONUL	2006/07	Numbers of downloads
NCOUNTER	Article downloads [N]	COUNTER-compliant statistics for all journals	SCONUL	2006/07	Numbers of downloads per user
ELSEVIER	Article downloads	ScienceDirect	Elsevier	2007	Index 100=Oxford
OXFORD	Article downloads	Oxford Journals	CIBER	2007	Index 100=Newcastle

Basic independent variables

Code	Variable	Notes	Source	Date	Units
ACADEMICS	FTE academic staff	-	HESA	2006/07	FTE
ALLJSPEND	Total serial subscriptions		SCONUL	2006/07	£
ALLTITLES	Current serial titles	-	SCONUL	2006/07	Numbers of titles
EJSPEND	Electronic serial subscriptions	Includes electronic only and print/electronic subscriptions.	SCONUL	2006/07	£
ETITLES	Electronic serial titles	E-only and print/e-subscriptions.	SCONUL	2006/07	Numbers of titles
INTERLOANS	ILL applications made	-	SCONUL	2006/07	Numbers of applications
NETLIBSPEND	Total library expenditure	-	SCONUL	2006/07	£
NETUNISPEND	Total institutional expenditure	A useful proxy for size.	SCONUL	2006/07	£
PAPERS	Research papers	Articles, notes and reviews published in 2007.	SCOPUS	2007	Numbers of papers
PhD	Numbers of PhDs	Research outcome.	HEIDI	2006/07	Numbers of awards in 2006/07
PJSPEND	Print serial subscriptions	-	SCONUL	2006/07	£
PTITLES	Print only serial titles	-	SCONUL	2006/07	Numbers of titles
QR	Quality-related income	-	HEIDI	2006/07	£
RGC	Research grant and contract income.	Research outcome.	HESA	2006/07	£
STANDING	Times research rating	A composite indicator based on official sources.	TIMES	2008	Scale=0 to 7
STUDENTS	FTE students	-	HESA	2006/07	FTE
TOTALESPEND	Total electronic expenditure	Journals, databases, e-books.	SCONUL	2006/07	£
USERS	FTE library users	Includes staff, students and external registered users.	SCONUL	2006/07	FTE

Derived independent variables

Basic independent variables normalised for size of institution

Code	Notes	Source	Units
NALLJSPEND	Total serial subscriptions [N]	ALLJSPEND / USERS	£ per user
NALLTITLES	Current serial titles [N]	ALLTITLES / USER	Number of titles per user
NCOUNTER	COUNTER article downloads [N]	COUNTER / USERS	Number of downloads per user
NEJSPEND	Electronic serial subscriptions [N]	EJSPEND / USERS	£ per user
NETTITLES	Electronic serial titles [N].	ETITLES / USER	Number of titles per user
NINTERLOANS	ILL applications made [N]	INTERLOANS / USERS	Number of interloans per user
NNETLIBSPEND	Total library expenditure [N]	NETLIBSPEND / USERS	£ per user
NNETUNISPEND	Total institutional expenditure [N]	NETUNISPEND / USERS	£ per user
NPJSPEND	Print serial subscriptions [N]	PJSPEND / USERS	£ per user
NPTITLES	Print only serial titles [N]	PTITLES / USERS	Number of titles per user
NQR	Quality-related income [N]	QR / ACADEMICS	£ per user
NRGC	Research grant and contract income [N]	RGC / USERS	£ per user
NTOTALESPEND	Library commitment to e-resources.	TOTALESPEND / USERS	£ per user

Additional independent variables (also normalised for size of institution)

Code	Notes	Source	Units
BARGAIN	A value-for-money measure.	100*(ALLJSPEND/NETUNISPEND)	Serial spend as a % of total university spend.
CHOPMOUT	A productivity measure.	PAPERS / ACADEMICS	Research papers per academic FTE
DEPTH	Another measure of use.	COUNTER/ETITLES	Article downloads per serial
DIGITAL	An indicator of how far the library has moved to e-journal provision.	100*(ETITLES / ALLTITLES)	% of journal titles that are available in e-form
SUPERVIS	A research productivity measure.	PhD / ACADEMICS	PhD awards per 100 academic FTEs
VFMDOWNLD	A measure of value for money.	EJSPEND / COUNTER	Library £ per article download
VFMPAPERS	A research productivity measure.	QR / PAPERS	QR £ per research paper

Descriptive and dummy variables

Code	Variable	Notes	Units
MEDICAL	Medical school	-	1=No 2=Yes
SECTOR	University grouping	-	1=Russell Group 2=Other old universities 3=Post-1994 universities 4=Other institutions
SUPERUSER	Superuser	Distribution of ELSEVIER in bands. The lower band is those cases one standard deviation below the mean. Heavy users are one standard deviation above the mean and superusers more than one standard deviation above.	1=Moderate user 2=Heavy user 3=Superuser

Annex C

Alternative models

MODEL 1: Electronic spend per user

Dependents = Number of research papers, Numbers of PhD awards, Total research grants and contract income (£)

Power model

$y = b_0 x^{b1}$

Table 28: Curve estimation for research papers, power model (*n*=115 UK universities)

Power model			*Parameter estimates*	
R-squared	Independent variable:	Sig.	Constant	b1
0.438	Electronic serial spend per user	<0.01	3.95	1.309

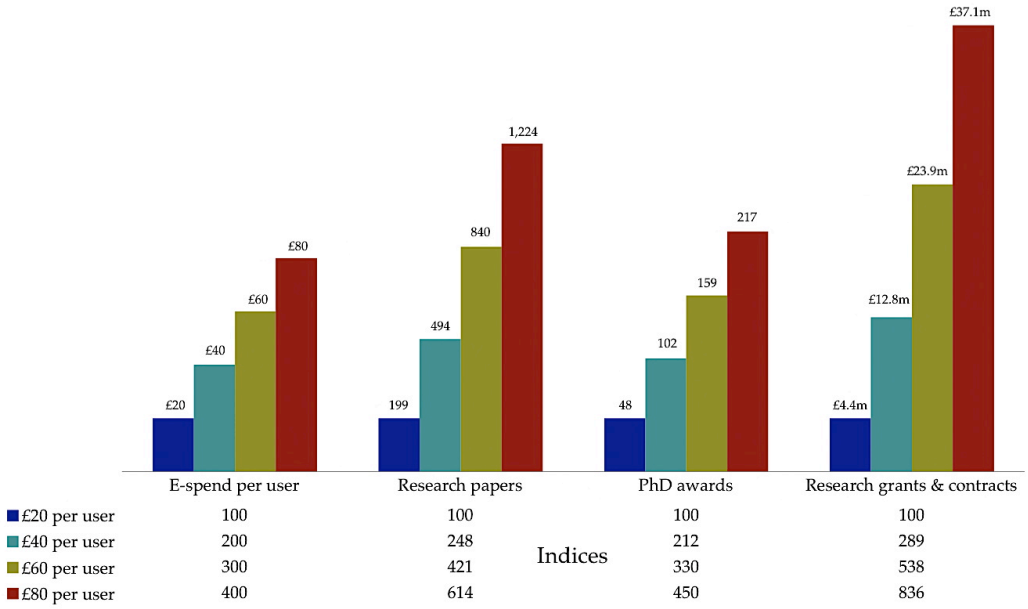
Table 29: Curve estimation for PhD awards, power model (*n*=115 UK universities)

Power model			*Parameter estimates*	
R-squared	Independent variable:	Sig.	Constant	b1
0.391	Electronic serial spend per user	<0.01	1.87	1.086

Table 30: Curve estimation for research grants and contracts income, power model (*n*=115 UK universities)

Power model			*Parameter estimates*	
R-squared	Independent variable:	Sig.	Constant	b1
0.438	Electronic serial spend per user	<0.01	45.09	1.532

Figure 16: Modelling the association between electronic serial spend per use and research outcomes, power model



Indices	E-spend per user	Research papers	PhD awards	Research grants & contracts
£20 per user	100	100	100	100
£40 per user	200	248	212	289
£60 per user	300	421	330	538
£80 per user	400	614	450	836

MODEL 2: Article downloads per user

Dependents = Number of research papers, Numbers of PhD awards, Total research grants and contract income (£)

Cubic model

$y = b_0 + (b_1x) + (b_2x^2) + (b_3x^3)$

Table 31: Curve estimation for research papers, cubic model (*n*=115 UK universities)

Cubic model		*Parameter estimates*				
R-squared	Independent variable:	Sig.	Constant	b1	b2	b3
0.486	Article downloads per user	<0.01	562.40	-47.39	1.36	-0.01

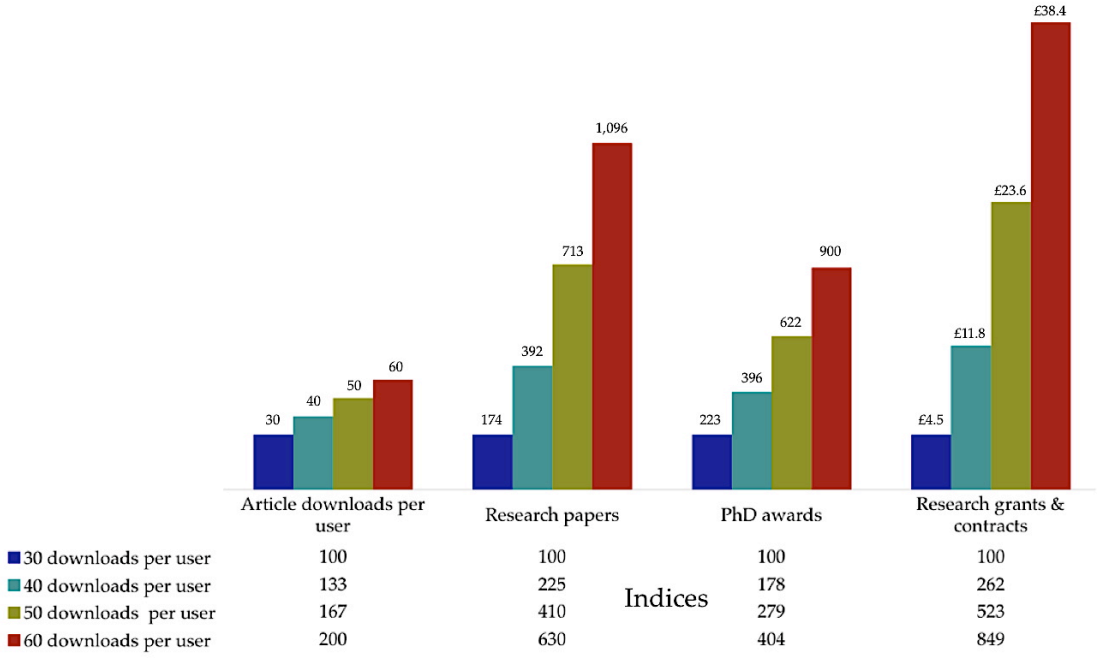
Table 32: Curve estimation for PhD awards, cubic model (*n*=115 UK universities)

Cubic model		*Parameter estimates*				
R-squared	Independent variable:	Sig.	Constant	b1	b2	b3
0.437	Article downloads per user	<0.01	130.81	-8.63	0.23	<0.01

Table 33: Curve estimation for research grants and contracts, cubic model (*n*=115 UK universities)

Cubic model		*Parameter estimates*				
R-squared	Independent variable:	Sig.	Constant	b1	b2	b3
0.367	Article downloads per user	<0.01	24,578.48	-2,021.1	52.64	-0.52

Figure 17: Modeling the association between article downloads per user and research outcomes, power model



MODEL III: Total electronic serials budget

Cubic model

$y = b_0 + (b_1x) + (b_2x^2) + (b_3x^3)$

Table 34: Curve estimation for research papers, cubic model (*n*=115 UK universities)

Cubic model		*Parameter estimates*				
R-squared	Independent variable:	Sig.	Constant	b1	b2	b3
0.625	Electronic serial spend	<0.01	-103.92	<0.01	<0.01	<0.01

Table 35: Curve estimation for PhD awards, cubic model (*n*=115 UK universities)

Cubic model		*Parameter estimates*				
R-squared	Independent variable:	Sig.	Constant	b1	b2	b3
0.604	Electronic serial spend	<0.01	-2.44	<0.01	<0.01	<0.01

Table 36: Curve estimation for research grants and contracts, cubic model (*n*=115 UK universities)

Cubic model		*Parameter estimates*				
R-squared	Independent variable:	Sig.	Constant	b1	b2	b3
0.629	Electronic serial spend	<0.01	-268.81	0.05	<0.01	<0.01

Figure 18: Modeling the association between electronic serial budgets and research outcomes, cubic model