



Do Journal Downloads Reliably Measure Usage? Trusting the Fox to Count Your Hens?

Joint work with Ted Bergstrom and Alex
Wood-Doughty

Douglas G. Steigerwald
UC Santa Barbara
December 2018

Influence of Academic Research

- Award grant funding
- Rank journals, departments, universities
- Hire and promote faculty
- **University librarians purchasing decisions**

Influence Measures

- Citation Counts
 - Science Citation Index 1964
 - Journal Citation Reports 1975
- Download Counts
 - electronic storage of research
 - more immediate
 - **local measure of usage**
 - downloads by university campus

Downloads Predict Citations

- Previous studies confirm
 - narrow data sets – arXiv or Caltech
- Our data is broad
 - 10 UC Campuses
 - 8,000 journals across a wide set of fields
- Downloads are a good predictor
 - “predicted” values of citations capture nearly 80% of the variation in citations

Factors in the Relationship

The relationship between downloads and citations depends on

1. Academic field
 1. ratio of downloads to citations is high for humanities
 2. ratio of downloads to citations is low for physical sciences
2. Journal (prestige, size)
3. Publisher?

Publisher Incentives

Davis and Price (2006)

publishers may artificially inflate downloads by requiring users to view HTML versions before PDF linking (CrossRef) directly to article (not abstract)

CEO Elsevier (2004) testimony on pricing

“The biggest single factor is usage ... the cost per article download is coming down... that is fantastic value”

Are download measures reliable?

- Downloads should not depend on publisher
- If downloads do depend on publisher:
Are downloads accurately measured?
- Download counts
 - gathered by publishers, not libraries
 - given in summary form to libraries

Data Sources

- Downloads – California Digital Library
 - handles subscriptions for all 10 UC campuses
 - downloads occurring in 2010-2016
 - 7 publishers – 7,724 journals
 - 425 million recorded downloads
- **Publishers attempt to restrict public access**
 - UC did not agree to restriction clause

Publishers

- 7 publishers selected because they provided detailed download reports
 - commercial publishers
 - Elsevier Springer Taylor & Francis Wiley
 - professional societies
 - American Chemical (ACS) Electrical Engineers (IEEE) Nature (NPG)
- Publishers report total downloads
 - **exclude downloads from open access journals**

Publisher Field Coverage

Table 1: Number of Subscription Journals by Research Field and Publisher

	Arts and Humanities	Health Sciences	Life Sciences	Physical Sciences	Social Sciences	Number of Journals
American Chemical Society	0	7	9	31	1	47
Elsevier	28	808	405	681	314	2235
IEEE	3	2	5	192	10	212
NPG: Nature-branded	0	11	16	7	1	34
NPG: Other	0	19	17	1	0	36
Springer	63	370	314	809	310	1865
Taylor & Francis	283	259	189	377	947	2054
Wiley	81	320	235	278	324	1238
Total	457	1796	1190	2375	1906	7724

Download & Citation Patterns

per article, average over journals

	Downloads	Citations (Impact Factor)	Ratio
Life Sciences	13.9	9.4	1.5
Health Sciences	9.7	7.4	1.3
Social Sciences	5.6	4.4	1.3
Physical Sciences	5.5	6.8	0.8
Arts & Humanities	5.3	2.7	2.0

citations are a poor measure of usage for Arts & Humanities

Download & Citation Patterns

per article, average over journals

	Downloads	Citations (Impact Factor)	Ratio
Commercial			
Elsevier	12.8	9.1	1.4
Springer	4.6	4.9	0.9
Taylor & Francis	3.0	0.9	3.3
Wiley	7.4	7.3	1.0

Elsevier reports a high ratio of downloads to citations – true for all their journals

Taylor & Francis has a small number of journals with very high download rates (their typical journal has a ratio of 0.9)

Modeling Downloads

- Journal downloads depend on
 1. Size – number of articles
 2. Prestige – number of citations
 3. Year – more downloads in recent years
 4. Field – witness Arts & Humanities
 5. Publisher – this should not matter

Downloads and Prestige

- compare 2 journals, one of which is more prestigious
 - in the same field, from the same publisher, of the same size, in the same download year
- for the journal with 10% more citations

		Download Increase
Highly Responsive	Life Sciences	14%
Neutral	Physical Sciences	10%
Less Responsive	Health Sciences	9%
	Social Sciences	7%
	Arts & Humanities	5%

Download Year

- average growth rate over 2011-2016
 - accounting for prestige, size, publisher

		Annual Growth Rate
Upward Trend	Health Sciences	3.1%
	Life Sciences ¹	2.4%
	Social Sciences	1.6%
None	Physical Sciences	
	Arts & Humanities	

- ¹ high variation across journals

Downloads and Field

- predict downloads based on
 - prestige, size, publisher, year of download
 - \hat{D} predicted downloads for a journal in Social Science

	Predicted Downloads
Arts & Humanities	$2.5 * \hat{D}$
Life Sciences	\hat{D}
Health Sciences	$0.8 * \hat{D}$
Physical Sciences	$0.4 * \hat{D}$

- Arts & Humanities 2.5 times the downloads in Social Science
- Physical Science 40% of the downloads in Social Science
- **more detailed predictions by field are in the paper**

Publisher Differences?

publisher should not matter

- compare reported downloads from 2 publishers
 - for 2 prestigious Health Sciences journals, that each publish 100 articles a year
 - Elsevier reports twice as many downloads at Taylor & Francis
 - Elsevier reports 40% more downloads than Springer or Wiley
- **Elsevier reports far more downloads**
 - precisely measured effects, unchanged if we use narrower field classifications
 - not specific to any given year of downloads

Publisher Platforms

- reported downloads include
 - HTML and PDF
 - publisher platform may automatically download HTML and provide an option to download PDF
 - artificially inflates downloads
- occurs when accessing an article through a search engine or Crossref
- **if this varies systematically across publishers, we should see publisher differences in the proportion of pdf downloads**

Publisher Differences

- examine reported total downloads
 - for every 100 pdf downloads:

	Reported Downloads
Elsevier	260
Springer	140
Taylor & Francis	140
Wiley	150

- similar in magnitude to publisher effects
- **local usage may be overstated for all publishers**
 - to “correct” Elsevier reported downloads, divide them in half

Key Takeaways

- significant differences in the way publishers report downloads
 - overcount actual usage
 - duplicates and bulk downloads
- ensure access to data so that accurate measures of local usage can be developed
 - do not agree to clauses restricting use of download data
 - demand transparency and obtain log files – the actual record of downloads – to provide detailed information and a check on reported numbers