LIBER QUARTERLY
The Journal of the Association of European Research Libraries
Vol. 34, (2024) 1–11 | e-ISSN: 2213-056X

Determinants of Downloads as Demand for Hybrid Journals: Rationale for Bundling Services

Sumiko Asai

Meiji University, Tokyo, Japan, asai@meiji.ac.jp, orcid.org/0000-0002-6663-7798

Abstract

Although Big Deal contracts that provide access rights to all electronic journals published by the publisher initially gained favor with university libraries, some libraries have terminated these contracts owing to increased charges since the 2010s. Consequently, they are faced with the problem of selecting journals for purchase within their limited budgets. This study investigates the factors affecting the number of downloads, representing journal demand, to provide libraries with guidance on journal selection. The download equation for 1,485 hybrid journals published by Springer Nature is formulated using ordinary least squares. The results found that 5% and 50% of the 1,485 journals generated approximately 30% and 85% of the downloads in 2022, respectively. Downloads are concentrated in fewer journals, although the Pareto principle does not apply to hybrid journals. Demand concentration implies that libraries do not need to maintain access rights to all journals. Recently, a few leading publishers have provided access rights to almost all electronic journals based on transformative agreements aiming to promote open access. Therefore, this study's findings raise the issue of the rationale for bundling electronic journals in transformative agreements, which is similar to Big Deal. Moreover, the results of the download estimation reveal that hybrid journals with more open access articles, larger citation scores, and longer histories acquire more downloads. These findings indicate that open access accelerates the dissemination of research.

Keywords: download; journal demand; Pareto principle; Big Deal; transformative agreement

Uopen Journals | http://liberquarterly.eu/ | DOI: 10.53377/lq.18689

1. Introduction

In the 2000s, many university libraries signed Big Deal contracts—the bundling services of printed and all electronic journals—with leading publishers, such as Elsevier and Springer Nature. However, some libraries have terminated their services since the 2010s owing to increased charges (Olsson et al., 2020; Pedersen et al., 2014; Rodríguez-Bravo et al., 2021). Instead of acquiring access rights to all electronic journals based on Big Deal contracts, they have reverted to the practice of subscribing to individual journals, purchasing individual articles called the pay-per-view model, and relying on interlibrary loans (Pedersen et al., 2014). Nevertheless, as subscription prices have steadily increased, university libraries face the dilemma of determining which journals to subscribe to in order to fulfill researchers' needs within their constrained budgets. Transformative agreements aiming to shift hybrid journals to fully open access journals have recently become pervasive in the academic journal market. A few leading publishers provide research institutions that sign transformative agreements with the bundling service of almost all journals. The services based on the transformative agreements, which provide access to a large number of electronic journals by adding small costs, including article processing charges, have similarities with Big Deal services. Hence, the issue of the rationale for bundling services persists today. Emrani et al. (2010) and Zhu and Xiang (2016) showed that the Pareto principle, which suggests that 80% of demand originates from 20% of supply, applies to the academic journal market. Contrastingly, Loan and Mushtaq (2023) found that the Pareto principle does not apply to oncology journals, despite the concentration of demand in a small number of journals. We examine the usefulness of bundling services by investigating the distribution of downloads, including the issue of whether the Pareto principle applies to target journals.

As for-profit publishers have not released their circulation data since approximately 2000 (Bergstrom, 2001), we cannot precisely investigate the subscription demand for individual journals. In contrast, Springer Nature announces the number of article downloads and annual number of journal downloads on its journals' websites. Asai (2022), Davis (2011), and Wang et al. (2015) found a higher number of downloads of open access articles than of non-open access articles. Furthermore, Moed (2005) revealed that approximately 40% of article downloads occurred within six months from the release. Several other studies have also demonstrated that downloads are concentrated shortly after an article is released (Asai, 2022; Davis, 2013; Ding et al., 2021; Schloegl

& Gorraiz, 2010; Watson, 2007). Moreover, several studies examined the relationship between downloads and citations (Congleton et al., 2022; Ding et al., 2021; Fernández-Ramos et al., 2023; Moed, 2005; Schlögl et al., 2014; Vaughan et al., 2017). Vaughan et al. (2017) demonstrated that the correlation coefficients between downloads and citations ranged from 0.57 in medicine to 0.88 in engineering and the humanities using a Chinese database of academic journals. Fernández-Ramos et al. (2023) found a positive correlation between downloads and citations using COUNTER Journal Report 1. Ding et al. (2021) identified a significant causal relationship between the downloads and citations of articles in *The Lancet* using a vector autoregressive model. Schlögl et al. (2014) calculated the correlation coefficients between downloads, citations, and readership data and found positive relationships. Congleton et al. (2022) examined the relationship between citations and downloads for a journal published by Springer Nature using ordinary least squares (OLS) and found that the two variables have positive impacts on each other. Thus, several studies have focused on the relationship between citations and downloads; few have examined the influence of factors other than citations on journal demand. This study explores the number of downloads as a variable representing journal demand, instead of the number of circulations, to identify the determinants of demand. As this study investigates only Springer Nature journals owing to limited data availability, the results cannot be generalized to overall academic journals. Nevertheless, findings on the distribution of downloads, including the Pareto principle, and the download determinants could provide guidance to libraries entering into subscription contracts with publishers.

2. Methodology

This study targeted 1,485 hybrid journals published by Springer Nature in 2022. All journals were indexed in Scopus, and at least one open access article was published. In addition, this study constructed a download equation using OLS. OLS is the most basic regression model and determines the estimators that minimize the sum of squared residuals. The download equation is formulated as follows:

$$\begin{aligned} ln\ Download_{2022} = \ & \mathrm{f}\,(Constant\ term,\ ln\ Article_{2022},\ ln\ OpenRatio_{2022},\\ & ln\ CiteScore_{2021},\ ln\ Issue_{2022},\ ln\ Year,\ Society,\\ & Academic\ disciplines) \end{aligned}$$

where *ln* represents the natural logarithm. Taking the natural logarithm of the variables allows us to interpret the estimators as elasticities of download to the respective variables. This study used *Article*, *OpenRatio*, *CiteScore*, *Issue*, *Year*, *Society*, and *Academic Discipline* as the independent variables in the *Download* equation. These independent variables were selected based on empirical studies exploring the determinants of subscription prices (Chressanthis & Chressanthis, 1994; Coomes et al., 2017; Petersen, 1992). Previous studies found that downloads are concentrated around the time of article release, as mentioned in the Introduction. Therefore, in this study, *Article* represents the number of articles published in 2022, although readers can download articles published before 2022. *OpenRatio* refers to the proportion of open access articles to the number of articles in 2022, measured as a percentage. *CiteScore* for 2021 is the number of citations from 2018 to 2021 divided by the number of articles published during the same period. *CiteScore* for 2021 is the latest citation score available when downloading articles in 2022. *Issue* denotes the number of issues published in the year 2022. *Year* is the number of years since the journal's inception and is calculated by subtracting the inception year from 2022. Springer Nature publishes journals on behalf of academic societies, universities, and other research institutions, in addition to journals Springer Nature launched independently. When a journal is published on behalf of an academic society, university, or other research institution, Springer Nature provides information on the research institution that commissions publications on the journal's website. If this information was lacking, we deemed the journal to have been launched independently by Springer Nature. *Society* is set to 1 if the journal is published on behalf of an academic society, university, or other research institution, and 0 otherwise.

This study used the Scopus classification for academic disciplines, including *Agriculture*, *Arts and humanities*, *Chemistry*, *Computer science*, *Earth and planetary science*, *Engineering*, *Environmental science*, *Material science*, *Mathematics*, *Medicine*, *Physics*, and *Social sciences*. The variable *Agriculture* is set to 1 if the journal is in the agricultural and biological sciences, and 0 otherwise. Other variables representing academic disciplines are established similarly. As *Material science* formed the base group when formulating the *Download* equation, this variable does not appear in the equation. This study sourced the number of open and non-open access articles, CiteScores, and academic disciplines from Scopus in August 2023. The number of downloads, the year of the journal's inception, and whether the journal was published on behalf of

the associated research institution were available from the journals' websites in the same month.

3. Findings

A total of 1,485 journals collectively published 225,595 articles in 2022, and these articles were downloaded a total of 505 million times. Figures 1 and 2 depict the number of journal downloads, ranked in descending order. A small number of journals have a large number of downloads.

Table 1 presents the relationship between the proportions of journal titles and downloads ranked in descending order. Column (a) shows the distribution of the 1,485 journals across all disciplines, whereas column (b) focuses on 355 journals in medicine with the most journals. Table 1 (a) indicates that 5% of the journals (74 of 1,485) account for 29.9% of downloads. Additionally, 20% of the journals (297 of 1,485) account for 60.8% of downloads, which is less than 80%. The results in column (b) are almost identical to those in column (a). The results demonstrate that the Pareto principle does not apply to hybrid journals published by Springer Nature, although downloads are concentrated in a small number of journals.

Fig. 1: Number of downloads.

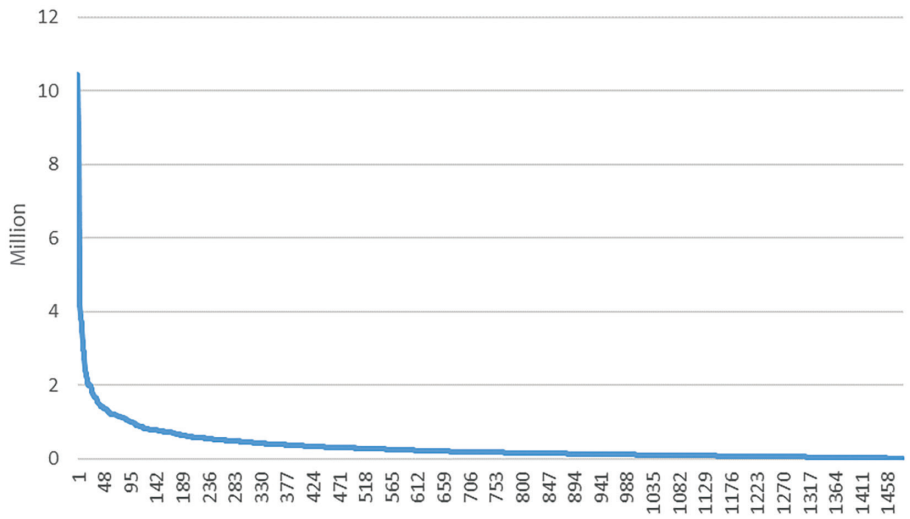

Fig. 2: Cumulative number of downloads.

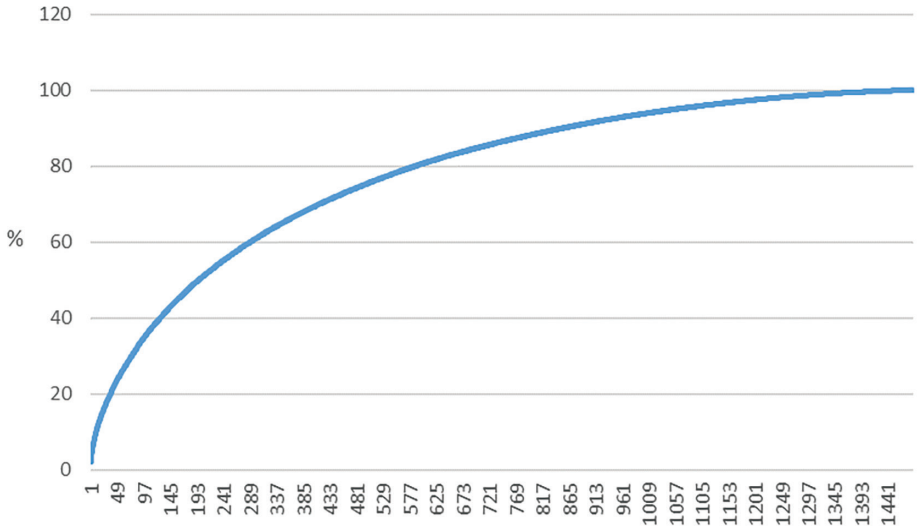

Table 1: Proportion of titles and downloads (%).

Titles	(a) Total	(b) Medicine
5	29.9	27.8
10	43.6	40.1
20	60.8	56.1
30	72.1	68.3
40	80.4	77.2
50	86.4	84.3
60	91.1	89.7
70	94.7	93.8
80	97.3	96.7
90	99.1	98.8
100	100.0	100.0

Table 2 outlines the statistics of the variables for the 1,485 hybrid journals, excluding the binary variables. The coefficients of variation for *Download* and *Article* indicate large variations. Moreover, the skewness values of the two variables are 8.32 and 12.00, respectively, demonstrating that the distributions have long right tails. The CiteScore mean is 4.4. Springer Nature publishes frequently cited journals. The mean *OpenRatio* is 23.2, implying that many articles in hybrid journals remain non-open access.

Table 2: Summary of statistics for variables.

	Download	*Article*	*OpenRatio*	*CiteScore*	*Issue*	*Year*
Mean	339,954	152	23.2	4.4	6.4	34.0
Median	179,356	83	20.4	3.9	6.0	29.0
SD	564,698	263	16.6	3.2	4.2	25.6
CV (%)	166	173	71.5	72.2	65.0	75.4
Skewness	8.32	12.00	0.85	3.84	3.59	2.14

SD: standard deviation; CV: coefficient of variation; *Download*: number of downloads in 2022; *Article*: number of articles published in a journal in 2022; *OpenRatio*: proportion of open access articles to total articles published in a journal in 2022 (%); *CiteScore*: CiteScore in 2021; *Issue*: number of issues in 2022; *Year*: number of years since the journal's inception.

Table 3 reports the estimation results of *Download* using OLS. Columns (a) and (b) display the results for journals across all disciplines and medicine only, respectively. The coefficient of *Article* (0.6297) in column (a) is positive at the 1% significance level, as expected. The coefficient of *OpenRatio* (0.2522) is

Table 3: Estimation results.

Variables	Coefficients	
	(a) Total	(b) Medicine
Constant	6.8471 (0.1330)***	7.6901 (0.1543)***
Article	0.6297 (0.0185)***	0.5486 (0.0321)***
OpenRatio	0.2522 (0.0163)***	0.2259 (0.0344)***
CiteScore	0.5172 (0.0242)***	0.5375 (0.0427)***
Issue	0.2036 (0.0328)***	0.2175 (0.0553)***
Year	0.1757 (0.0200)***	0.1583 (0.0396)**
Society	−0.1198 (0.0283)***	−0.1000 (0.0507)**
Agriculture	0.2453 (0.1060)**	
Arts and Humanities	0.5039 (0.1161)***	
Chemistry	0.2895 (0.1171)**	
Computer science	−0.0603 (0.1161)	
Earth and planetary science	−0.0492 (0.1182)	
Engineering	−0.0140 (0.1066)	
Environmental science	0.1687 (0.1264)	
Mathematics	−0.3703 (0.1064)***	
Medicine	0.3840 (0.1009)***	
Physics	−0.0708 (0.1239)	
Social sciences	0.5277 (0.1034)***	
Adjusted R^2	0.7709	0.7899
Number of observations	1,485	355

Standard errors are presented in parentheses.
***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

also positive at the 1% significance level. This result implies that open access articles are more frequently downloaded than non-open access articles, confirming the findings of prior studies (Asai, 2022; Davis, 2011; Wang et al., 2015).

The coefficient of *CiteScore* (0.5172) is positive at the 1% significance level. The positive value suggests that citation score is one factor in selecting journals for reading. The coefficient of *Year* (0.1757) is positive at the 1% significance level. This finding has two implications. First, journals with long publishing histories tend to have many loyal readers, leading to a large number of downloads. Second, articles published before 2022 can be downloaded in 2022. Journals that have existed for many years tend to publish numerous articles. Therefore, *Year* represents the number of downloadable articles in addition to journal popularity, although the two effects cannot be separated. The coefficient of *Society* is negative at the 1% significance level. The findings indicate that journals published independently by Springer Nature have more downloads than those published on behalf of academic societies and other research institutions. For journals published independently by Springer Nature, the publisher may carefully assess demand before deciding whether to launch a journal. The coefficients of *Arts and Humanities*, *Medicine*, and *Social sciences* are positive at the 1% significance level, indicating a large number of downloads in a journal. Conversely, the number of downloads in a mathematical journal is small due to the negative coefficient of *Mathematics*. The estimated results for medical journals are similar to those across all disciplines.

4. Discussion and Conclusion

Table 1 indicates a concentration of downloads in a small number of journals, although they do not follow the Pareto principle. It is reasonable to consider whether to continue the Big Deal contract by calculating the cost per download and comparing it with the pay-per-view and interlibrary loans (Chamberlain, 2022; Pedersen et al., 2014). On the one hand, a publisher's marginal costs for distributing an article are nearly zero. Even if publishers permit electronic access to all journals, the incremental costs of distribution are small. On the other hand, libraries that subscribe to individual journals tend to decrease the number of subscription journals owing to the increased

subscription prices. However, as transformative agreements predetermine the payments during the contract period, publishers can maintain their revenues through transformative agreements. Moreover, libraries and researchers in the institutions may favor the bundling services of electronic journals based on the transformative agreements at present because the number of journals available increases significantly. However, if the prices specified in transformative agreements continue to increase, libraries will face a similar problem as experienced in Big Deal—the difficulties in purchasing books and other publishers' journals owing to the constrained library budgets other than the bundling services (Edlin & Rubinfeld, 2004). Although the initial purpose of transformative agreements is to promote open access, bundling services would stiffen library budgets. Consequently, sales of other publishers without transformative agreements decline, and leading publishers with transformative agreements may enhance their market power in the academic journal market. Therefore, we should carefully monitor the prices of the transformative agreements and examine the usefulness of the bundling services.

Previous studies demonstrated that the number of downloads of open access articles was higher than that of non-open access articles (Asai, 2022; Davis, 2011; Wang et al., 2015). Our findings also indicate that open access increases downloads. Therefore, open access is beneficial for disseminating research. Regarding the 1,485 hybrid journals examined in this study, the proportion of open access articles ranged from 0.4% to 88.9%. Thus, the progress of open access varies across journals. Even if universities canceled subscriptions to hybrid journals with many open access articles, researchers in the institutions would not have great difficulty accessing the articles. Therefore, it is reasonable for university libraries to decide to subscribe to journals based on the level of progress of open access in addition to the fitness of the research domains and the frequency of citations.

This study has some limitations. Our findings cannot be generalized to overall journals, as this study investigated hybrid journals published only by Springer Nature. Second, it considered the distribution of journal demand without taking into account subscription payments because some research institutions purchase journals at unpublished prices that differ from list prices. The relationship between demand and price remains unclear, and more research is warranted.

References

Asai, S. (2022). Access patterns of electronic articles: The case study of Scientometrics. *Publishing Research Quarterly*, *38*(2), 295–307. https://doi.org/10.1007/s12109-022-09887-0

Bergstrom, T. C. (2001). Free labour for costly journals? *Journal of Economic Perspectives*, *15*(4), 183–198. https://doi.org/10.1257/jep.15.4.183

Chamberlain, S. (2022). Lessons learned from reevaluating Big Deals with Unsub. *Serials Review*, *48*(3–4), 234–237. https://doi.org/10.1080/00987913.2022.2132090

Chressanthis, G. A., & Chressanthis, J. D. (1994). The determinants of library subscription prices of the top-ranked economics journals: An econometric analysis. *The Journal of Economic Education*, *25*(4), 367–382. https://doi.org/10.1080/00220485.1994.10844848

Congleton, R. D., Marsella, A., & Cardazzi, A. J. (2022). Readership and citations as alternative measures of impact. *Constitutional Political Economy*, *33*(1), 100–114. https://doi.org/10.1007/s10602-021-09333-x

Coomes, O. T., Moore, T. R., & Breau, S. (2017). The price of journals in geography. *The Professional Geographer*, *69*(2), 251–262. https://doi.org/10.1080/00330124.2016.1229624

Davis, P. M. (2011). Open access, readership, citations: A randomized controlled trial of scientific journal publishing. *The FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology*, *25*(7), 2129–2134. https://doi.org/10.1096/fj.11-183988

Davis, P. M. (2013). Public accessibility of biomedical articles from PubMed Central reduces journal readership – Retrospective cohort analysis. *The FASEB Journal*, *27*(7), 2536–2541. http://doi.org/10.1096/fj.13-229922

Ding, Y., Dong, X., Bu, Y., Zhang, B., Lin, K., & Hu, B. (2021). Revisiting the relationship between downloads and citations: A perspective from papers with different citation patterns in the case of the Lancet. *Scientometrics*, *126*(9), 7609–7621. https://doi.org/10.1007/s11192-021-04099-3

Edlin, A. S., & Rubinfeild, D. L. (2004). Exclusion or efficient pricing? The 'Big Deal' bundling of academic journals. *Antitrust Law Journal*, *72*(1), 119–158.

Emrani, E., Moradi-Salari, A., & Jamali, H. R. (2010). Usage data, e-journal selection, and negotiations: An Iranian consortium experience. *Serials Review*, *36*(2), 86–92. https://doi.org/10.1080/00987913.2010.10765289

Fernández-Ramos, A., Rodríguez-Bravo, B., & Diez-Diez, Á. (2023). Use of scientific journals in Spanish universities: Analysis of the relationship between citations and downloads in two university library consortia. *Scientometrics*, *128*(4), 2489–2505. https://doi.org/10.1007/s11192-023-04670-0

Loan, F. A., & Mushtaq, R. (2023). Pareto 80/20 rule: A new principle for serial subscription. *Serials Review*, *49*(1–2), 10–14. https://doi.org/10.1080/00987913.2023.2179783

Moed, H. F. (2005). Statistical relationships between downloads and citations at the level of individual documents within a single journal. *Journal of the American Society for Information Science and Technology*, *56*(10), 1088–1097. https://doi.org/10.1002/asi.20200

Olsson, L., Lindelöw, C. H., Österlund, L., & Jakobsson, F. (2020). Swedish researchers' responses to the cancellation of the big deal with Elsevier. *Insights the UKSG Journal*, *33*, Article 25. https://doi.org/10.1629/uksg.521

Pedersen, W. A., Arcand, J., & Forbis, M. (2014). The Big Deal, interlibrary loan, and building the user-centered journal collection: A case study. *Serials Review*, *40*(4), 242–250. https://doi.org/10.1080/00987913.2014.975650

Petersen, H. C. (1992). The economics of economics journals: A statistical analysis of pricing practices by publishers (Research Note). *College and Research Libraries*, *53*(2), 176–181. https://doi.org/10.5860/crl_53_02_176

Rodríguez-Bravo, B., Fernández-Ramos, A., De-la-Mano, M., & Vianello-Osti, M. (2021) The evolution and revision of big deals: A review from the perspective of libraries. *El Profesional de la Información*, *30*(4), 1699–2407. https://doi.org/10.3145/epi.2021.jul.15

Schloegl, C., & Gorraiz, J. (2010). Comparison of citation and usage indicators: The case of oncology journals. *Scientometrics*, *82*(3), 567–580. https://doi.org/10.1007/s11192-010-0172-1

Schlögl, C., Gorraiz, J., Gumpenberger, C., Jack, K., & Kraker, P. (2014). Comparison of downloads, citations and readership data for two information systems journals. *Scientometrics*, *101*(2), 1113–1128. https://doi.org/10.1007/s11192-014-1365-9

Vaughan, L., Tang, J., & Yang, R. (2017). Investigating disciplinary differences in the relationships between citations and downloads. *Scientometrics*, *111*(3), 1533–1545. https://doi.org/10.1007/s11192-017-2308-z

Wang, X., Liu, C., Mao, W., & Fang, Z. (2015). The open access advantage considering citation, article usage and social media attention. *Scientometrics*, *103*(2), 555–564. https://doi.org/10.1007/s11192-015-1547-0

Watson, A. B. (2007). Measuring demand for online articles at the *Journal of Vision*. *Journal of Vision*, *7*(7), 1–3. https://doi.org/10.1167/7.7.i

Zhu, Q., & Xiang, H. (2016). Differences of Pareto principle performance in e-resource download distribution: An empirical study. *The Electronic Library*, *34*(5), 846–855. https://doi.org/10.1108/EL-05-2015-0068