Predatory publishers in the search environment: web traffic evidence for scholarly communication

Predatory publishing is usually studied through publication output, deceptive editorial practices, author behaviour and citation traces. Less is known about its reader-side digital footprint: how visible predatory publishers are on the open web, how users reach their websites and whether traffic resembles sustained scholarly use or shallow exposure. This research article analyses Semrush web-traffic estimates for a random sample of 213 predatory publisher records drawn from a master list of 1,361 entries, complemented by a benchmark group of ten established publishers. Data covered monthly organic traffic from January 2019 to December 2024, acquisition channels, engagement indicators and country-level traffic origins for high-traffic sites. The sampled predatory publishers received an estimated 244.1 million organic visits over six years. Traffic was extremely concentrated: ten records accounted for 85.6% of observed visits and the Gini coefficient was 0.94. Compared with established publishers, predatory publishers relied more heavily on organic search, had less referral traffic and generated shallower engagement. The findings show that predatory publishing is not only a problem of manuscript solicitation and publication output, but also of search-mediated discoverability within the scholarly communication landscape.

Keywords
predatory publishing, scholarly communication, open access, webometrics, discoverability, research integrity



RUBÉN COMAS-FORGAS
Professor
University of the Balearic Islands (UIB)
Spain



MERCÈ MOREY-LÓPEZ
Assistant Professor
University of Balearic Islands (UIB)
Spain



MANOS PAVLAKIS
Assistant Professor
Frederick University
Cyprus



HIROMI M. YOKOYAMA
Professor
University of Tokyo
Japan

Introduction

Predatory publishing remains one of the most persistent problems in the contemporary scholarly communication landscape. The issue has been debated for more than a decade, including over how predatory journals should be defined and distinguished (Beall, 2012; Grudniewicz et al., 2019). It is commonly discussed as an author-side risk: researchers receive flattering invitations, submit work to weakly controlled journals and sometimes pay article processing charges to publishers that provide little or no meaningful peer review. This view is important, but incomplete. Predatory publishers are also reader-facing web infrastructures. Their websites are indexed, searchable and visually similar to legitimate platforms. They can appear in the same open web environments in which researchers,

students, librarians, policy analysts and evidence users search for journals, articles, editorial policies and publication guidance.

Previous research has established the growth of predatory open access publishing, compared the characteristics of questionable and legitimate journals and examined why authors publish in suspect outlets (Demir, 2018; Frandsen, 2019). Other studies have shown that predatory journals can enter indexing systems and citation databases, raising concerns about the contamination of scholarly records (Strinzel et al., 2019). These studies have substantially improved our understanding of the supply side of predatory publishing: the journals, publishers, authors and articles involved.

The reader-side dimension is less visible. Citation studies nevertheless show that articles in predatory journals can circulate within scholarly and policy evidence environments (Albert et al., 2025; Björk et al., 2020). An article may receive few citations while still being found, downloaded, skimmed or used in teaching, policy preparation or preliminary literature searching. A publisher website may not be trusted by experts but may still rank in search results, display digital object identifiers, claim indexing, show editorial boards and provide article PDFs that look formally academic. In these situations, the problem is not only whether an author submits to a deceptive journal, but whether unreliable scholarly-looking content remains discoverable within routine search and discovery workflows.

'Citation studies … show that articles in predatory journals can circulate within scholarly and policy evidence environments'

Insights has previously addressed this challenge from the perspective of the scholarly communication community. Shen and Shah (2023) describe predatory publishing as a critical problem for researchers and emphasize that researchers should not rely solely on watchlists but should develop skills to detect predatory practices. Dobson (2016) presents Think.Check.Submit. as a cross-sector campaign designed to help researchers make informed decisions about publication venues and notes the role of librarians and scholarly communication professionals in disseminating guidance. Related reviews have also synthesized definitions and checklists for identifying potentially predatory journals (Cobey et al., 2018; Cukier et al., 2020). The present article contributes a complementary evidence layer: instead of asking only how to recognize predatory publishers, it asks how visible these publishers are on the web and through which pathways users encounter them.

Web traffic is not a measure of quality, credibility or scientific impact. It cannot tell us whether a visit was made by a prospective author, a student, a librarian, a reviewer, a search engine crawler or a casual reader. Nevertheless, traffic estimates are useful when direct data are unavailable, particularly when interpreted as comparative indicators rather than audited usage counts (Jansen et al., 2022). Predatory publishers rarely release audited usage data, transparent submission data or verified readership statistics. Carefully interpreted third-party webometrics therefore provide one of the few scalable ways to study exposure, discovery pathways and engagement at domain level.

This study uses a longitudinal webometric design to analyse the digital footprint of predatory publishers between 2019 and 2024. It addresses four questions: how large and concentrated is estimated organic traffic to sampled predatory publishers, how did that traffic change over time, how do acquisition channels and engagement patterns compare with those of established publishers and what do country-level traffic patterns suggest about the global distribution of exposure? The article is written for the scholarly communication community, including librarians, publishers, research offices, indexing services and those who support researchers in navigating open access and publication choices.

'This study uses a longitudinal webometric design to analyse the digital footprint of predatory publishers'

Methods

Design and data source

The study used a longitudinal webometric design. Webometrics refers to the quantitative analysis of web-based structures, links, visibility, traffic and interactions (Björneborn & Ingwersen, 2004; Thelwall, 2009). This approach is particularly appropriate for predatory

publishing because many relevant processes are mediated through websites but are difficult to observe using conventional bibliometric data alone. Publisher websites are the interface through which users encounter journal titles, indexing claims, article PDFs, author guidelines, editorial-board information and publication fees.

Web-metric data were collected using Semrush between March and April 2025. Semrush estimates website traffic through proprietary combinations of clickstream, keyword, search and panel data. The values reported in this article are therefore not server logs and are not equivalent to COUNTER-compliant usage statistics. They should be interpreted as estimates of relative magnitude and pattern. For this reason, the analysis emphasizes aggregate trends, rank concentration, channel structure and group comparisons rather than exact claims about individual websites.

'the analysis emphasizes aggregate trends, rank concentration, channel structure and group comparisons'

Sampling frame and benchmark group

The primary dataset contained 213 predatory publisher records selected randomly in November 2024 from a master list of 1,361 entries on the Predatory Journals website (Predatory Journals, 2024). The sample covered 15.65% of the available population. Assuming maximum variability (p = .50), a 95% confidence level and a finite population of 1,361 entries, the estimated margin of error was approximately +/- 6.17%. The classification of publishers as predatory relied on the external sampling frame; the study did not conduct a new case-by-case adjudication of each publisher. This choice enabled analysis at scale but means that results should be interpreted at the aggregate level rather than as legal or editorial judgements on individual entities.

A benchmark group of ten established publishers was included as an interpretive reference point: Springer Nature, MDPI, Elsevier, Taylor & Francis, Wiley, Sage, Oxford University Press, Emerald, Frontiers and BMJ. This benchmark is not a matched causal control group. The publishers differ from the predatory sample in size, reputation, platform architecture, field coverage and discovery infrastructure. Its purpose is to provide context for interpreting the structure of web traffic, especially the balance between organic search, direct traffic and referral traffic.

'A benchmark group of ten established publishers was included as an interpretive reference point'

Variables and analysis

Publisher-level variables included number of journals, journal-portfolio category, declared country, conference activity, book publishing activity and whether the publisher combined conferences and books. Webometric variables included monthly organic traffic estimates for 72 months from January 2019 to December 2024, total organic traffic, acquisition channels (direct, referral, organic search, paid search, organic social, paid social, e-mail and display advertising) and engagement indicators (pages per visit, average visit duration and bounce rate).

Country-level traffic data were available for the high-traffic subset, defined as sites with more than 100,000 accumulated organic visits between 2019 and 2024. Country traffic was analysed for March 2023 to March 2025. Visits were aggregated by country across site records and weighted averages were calculated for pages per visit and bounce rate. An exploratory visits-per-researcher normalization was also examined using data from the UNESCO Institute for Statistics and World Bank indicators, but it is treated as interpretive because researcher-denominator values vary in year, coverage and quality across countries.

Traffic distributions were highly skewed, so the analysis used medians, ranked shares, cumulative concentration and the Gini coefficient. Mann-Whitney U tests compared predatory and benchmark publishers for traffic and engagement indicators, and Kruskal-Wallis tests compared predatory publishers by journal-portfolio category. No personal data were collected, no cookies were placed by the researchers and no individual-level browsing traces were accessed. Ethics approval was not required because the study used publicly visible website information and aggregate third-party traffic estimates.

Results

Publisher structure

The sample contained both small operators and large portfolio-based publishers. Excluding seven missing cases, 29.1% of publishers reported fewer than five journals, 18.4% reported six to ten journals, 24.8% reported 11 to 50 journals and 27.7% reported more than 50 journals. Just over half of the publishers with available data therefore reported more than ten journals. Declared countries were concentrated in a small number of locations, especially the United States and India, but 18.3% of records lacked a usable country value.

Additional services were less common than might be expected. A total of 40 publishers (18.8%) offered conference organization and 49 (23.0%) offered book publishing. Sixteen publishers (7.5%) offered both conferences and books, 57 (26.8%) offered one of the two services and 140 (65.7%) offered neither. Service diversification may be important for particular predatory operators, but it was not a universal feature of the sample. Table 1 summarizes the structural characteristics of the sample.

Dimension	Category	n	%
Number of journals	Fewer than 5	60	29.1
	6–10	38	18.4
	11–50	51	24.8
	More than 50	57	27.7
Country declared	United States	54	25.4
	India	54	25.4
	United Kingdom	19	8.9
	Canada	7	3.3
	Pakistan	6	2.8
	Other countries	34	16.0
	Missing/undisclosed	39	18.3
Additional services	Conferences: yes	40	18.8
	Book publishing: yes	49	23.0
	Both conferences and books	16	7.5
	Neither service	140	65.7

Table 1. Structural characteristics of sampled predatory publishers (n = 213)

Traffic scale, time trend and concentration

Across the full 2019–2024 period, the 213 predatory publisher records accumulated an estimated 244.1 million organic visits. Annual traffic rose from 16.2 million visits in 2019 to 26.5 million in 2020 and 65.1 million in 2021. It then declined to 41.6 million in 2022, increased again to 65.5 million in 2023 and fell to 29.3 million in 2024. Mean monthly visits per active publisher followed the same pattern: 9,213 in 2019, 14,521 in 2020, 35,244 in 2021, 22,406 in 2022, 34,914 in 2023 and 15,505 in 2024 (see Table 2 and Figure 1).

The temporal pattern indicates that predatory publisher visibility intensified during the Covid-19 pandemic period but remained volatile. The increase in 2020 and especially 2021 coincided with Covid-19 restrictions that limited access to laboratories, field sites and other research spaces for many researchers, potentially increasing time devoted to desk-based research, literature searching and manuscript preparation. This temporal co-occurrence should not be interpreted as causal. The decline in 2022, rebound in 2023 and fall in 2024 suggest that visibility was shaped by multiple factors, including search algorithms, domain availability, publisher activity, de-indexing, user awareness and Semrush coverage, rather than by a simple linear growth process.

Traffic was extremely concentrated. The median total traffic per predatory publisher record was only 31,785 visits across six years, while the mean was 1.15 million. The top publisher record accounted for 25.4% of all observed traffic, the top three for 55.6%, the top ten for 85.6% and the top 20 for 92.9%. The Gini coefficient was 0.94. Exposure to predatory publisher websites is therefore not evenly distributed across the long tail of suspect publishers; it is dominated by a small number of visible domains and brands (see Figure 2).

'Exposure to predatory publisher websites … is dominated by a small number of visible domains and brands'

Year	Annual visits	Mean monthly visits	Mean monthly visits per active publisher	Mean active records
2019	16,200,171	1,350,014	9,213	146.6
2020	26,451,342	2,204,278	14,521	151.6
2021	65,131,783	5,427,649	35,244	154.0
2022	41,577,177	3,464,765	22,406	154.8
2023	65,455,135	5,454,595	34,914	156.0
2024	29,296,538	2,441,378	15,505	157.5

Table 2. Organic traffic to predatory publisher records, 2019–2024

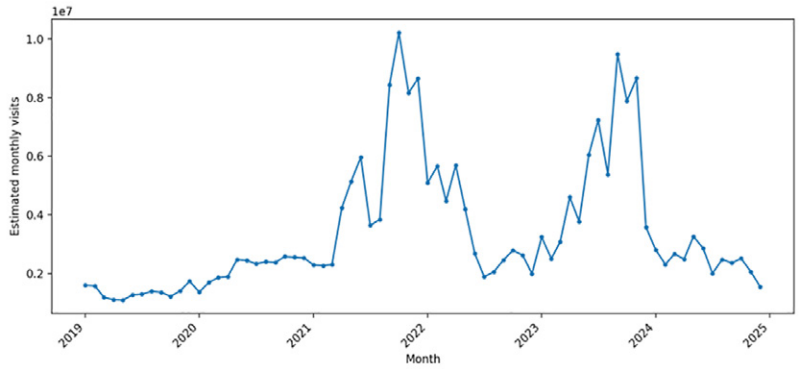

Figure 1. Monthly aggregate organic traffic to sampled predatory publisher records, January 2019 – December 2024

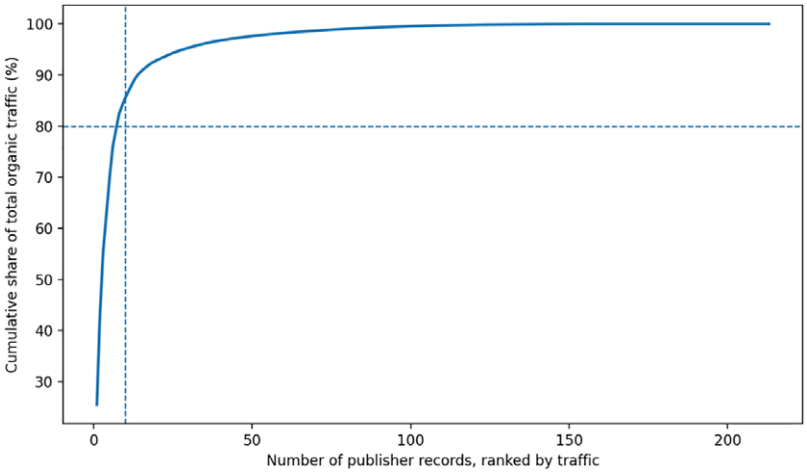

Figure 2. Cumulative traffic concentration by ranked predatory publisher records

Acquisition channels and engagement

Acquisition-channel data showed that predatory publisher traffic was strongly search-mediated. Organic search represented 60.6% of acquisition traffic for predatory publishers, compared with 43.1% for the benchmark publishers. Direct traffic represented 27.4% of predatory publisher acquisition traffic and 32.9% for benchmark publishers. Referral traffic was much lower for predatory publishers (9.7%) than for benchmark publishers (21.5%). Paid search, paid social, email and display advertising were marginal in both groups.

Engagement indicators also differed. Predatory publishers had a lower weighted average of pages per visit (1.71) than benchmark publishers (2.38), a shorter weighted average visit duration (304 seconds compared with 488 seconds) and a higher weighted bounce rate (74.7% compared with 60.0%). Non-parametric comparisons were statistically significant for pages per visit (p = .008), visit duration (p = .004) and bounce rate (p = .007). The pattern suggests that predatory publishers attract substantial web exposure but generate shallower interaction than established publishers.

The benchmark comparison is central to interpretation. Established publishers generated much larger traffic volumes overall, as expected, but they also displayed a different traffic structure: more referral traffic and deeper engagement. This is consistent with established publishers being embedded in library systems, indexing services, journal platforms, institutional links, references and alerts. Predatory publishers appear to depend more heavily on open-web discoverability (see Table 3 and Figure 3).

'Predatory publishers appear to depend more heavily on open-web discoverability'

Indicator	Predatory publishers	Established publishers
Direct traffic share	27.4%	32.9%
Referral traffic share	9.7%	21.5%
Organic search share	60.6%	43.1%
Weighted pages per visit	1.71	2.38
Weighted visit duration	304 seconds	488 seconds
Weighted bounce rate	74.7%	60.0%
Median total traffic per publisher, 2019–2024	31,785	250,507,510

Table 3. Acquisition channels and engagement comparison

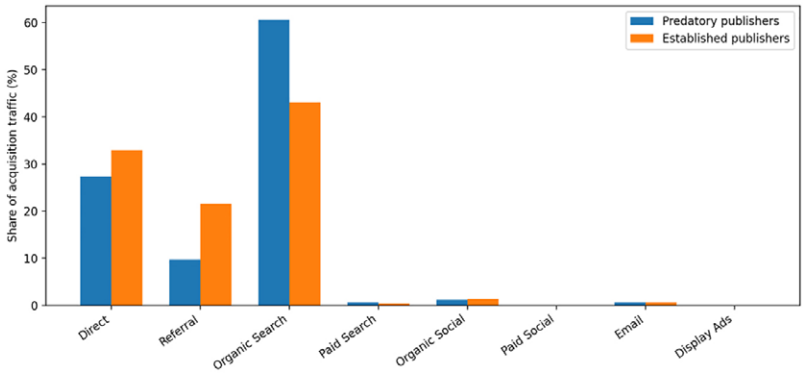

Figure 3. Distribution of traffic acquisition channels in predatory and established publisher groups

Country-level exposure

Country-level data for high-traffic predatory sites (more than 100,000 accumulated organic visits between 2019 and 2024) revealed a global pattern. The United States generated the largest absolute volume of traffic, with 7.8 million visits between March 2023 and March 2025, followed by India (6.3 million), Brazil (6.0 million), China (3.0 million) and the Philippines (2.2 million). Kenya, Indonesia, the United Kingdom, Canada and Australia completed the top ten. The top five countries accounted for more than half of country-attributed traffic in the high-traffic predatory subset. Table 4 summarizes the top ten country origins, which are also shown in Figure 4.

These figures should not be interpreted as evidence that researchers in those countries publish in, endorse or intentionally use predatory journals. Traffic is not authorship. Country-level visits may include authors, readers, students, librarians, editors, automated agents, VPN-routed traffic and search engine crawlers. VPN use can shift the apparent country of origin because a visit may be geolocated to the VPN exit server rather than to the user's physical location. As the United States is a common location for VPN exit infrastructure, some traffic attributed to the United States may originate from

users elsewhere; analogous distortions are possible for other common exit locations. Nevertheless, the results show that exposure is not confined to a single region or to countries with weak research infrastructures. Predatory publisher websites are visible from both established and emerging research systems.

'Predatory publisher websites are visible from both established and emerging research systems'

Normalizing web traffic to predatory publishing platforms by the number of researchers per country reveals substantial variations in how intensely research communities engage with these publishers. This indicator, visits per researcher, offers a complementary lens for assessing exposure alongside absolute traffic volumes. The number of researchers per country used for this analysis is based on the latest available statistics from the UNESCO Institute for Statistics (UIS) and World Bank Open Data. Specifically: UNESCO Institute for Statistics (UIS) – The primary source for data on full-time equivalent (FTE) researchers across countries. These data are compiled from national statistical offices. UNESCO Institute for Statistics data were used as the primary source (UNESCO Institute for Statistics, 2025). Where these data were outdated or incomplete, the World Bank's Researchers in R&D (per million people) indicator was used to estimate researcher numbers by multiplying the reported rate by the country's total population and dividing by one million (World Bank, 2025). Kenya appears as the most exposed case in the exploratory ranking, with an estimated 139 visits per researcher between March 2023 and March 2025. The Philippines follows with around 104 visits per researcher. Brazil, India and Ethiopia occupy the next group, with values in the range of approximately 19 to 27 visits per researcher. These values should be treated cautiously because denominator data differ in year and quality, but they suggest that relative exposure may be particularly intense in some smaller or fast-growing research systems.

Rank	Country	Visits	Share of country-attributed predatory traffic	Weighted pages/visit	Weighted bounce rate
1	United States	7,799,847	17.4%	1.45	81.1%
2	India	6,263,102	14.0%	3.24	66.6%
3	Brazil	5,994,407	13.4%	2.87	53.1%
4	China	3,026,470	6.8%	3.14	50.6%
5	Philippines	2,184,744	4.9%	1.44	77.7%
6	Kenya	1,668,568	3.7%	1.62	76.5%
7	Indonesia	1,443,391	3.2%	1.91	71.5%
8	United Kingdom	1,362,630	3.0%	1.49	80.9%
9	Canada	988,094	2.2%	1.40	81.8%
10	Australia	870,828	1.9%	1.44	81.5%

Table 4. Main country origins for high-traffic predatory publisher websites, March 2023 – March 2025

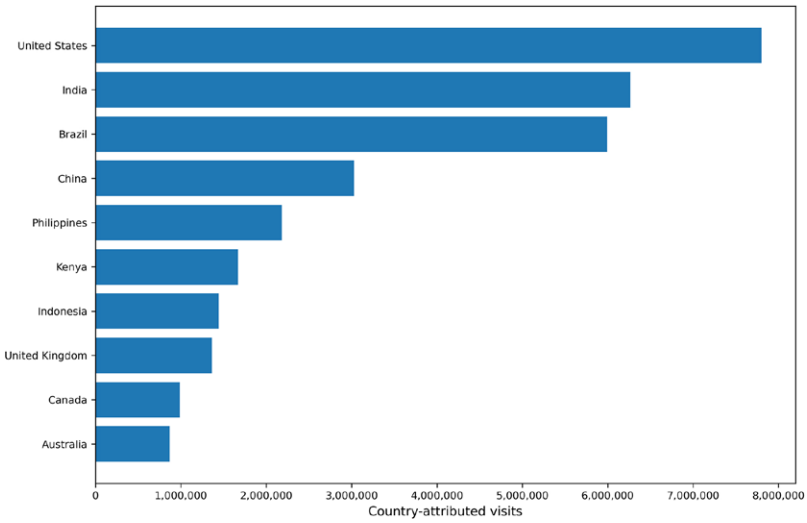

Figure 4. Top country origins for high-traffic predatory publisher websites, March 2023 – March 2025

Discussion

The findings broaden the evidence base on predatory publishing in three ways. First, they show that predatory publishers have a measurable and substantial web presence. The sampled records accumulated an estimated 244.1 million organic visits over six years. Even allowing for uncertainty in Semrush estimates, this scale indicates that predatory publisher websites are not marginal within the online scholarly information environment.

Second, exposure is highly concentrated. A small number of publisher records account for the overwhelming majority of observed traffic. This matters for the scholarly communication community because interventions can be prioritized. General education remains necessary, but domain-level monitoring, search engine quality signals, publisher-authentication tools and targeted guidance may be most effective if they address the highly visible operators that dominate exposure.

'A small number of publisher records account for the overwhelming majority of observed traffic'

Third, the study identifies organic search as the principal acquisition channel for predatory publishers. This finding shifts attention from e-mail solicitation alone to ordinary discovery behaviour. A researcher looking for a submission venue, a student searching for an article PDF or a policy analyst searching for evidence may encounter predatory publisher content through the same search interface used to reach legitimate scholarly sources. This is a discoverability problem, not only an author-deception problem.

The results complement earlier practical guidance in *Insights*. Dobson (2016) described Think.Check.Submit. as an effort to empower researchers to evaluate publication venues rather than rely solely on lists. Shen and Shah (2023) similarly argued that researchers should develop skills to detect predatory practices and use practical resources for assessing journals. The present data suggest why such skills remain necessary: predatory publishers are not invisible. They are discoverable at scale, often through organic search, and they can imitate the web conventions of legitimate publishing.

'researchers should develop skills to detect predatory practices and use practical resources for assessing journals'

The benchmark comparison clarifies the nature of the problem. Established publishers attract much larger volumes of traffic, but they also receive more referral traffic and generate deeper engagement. This suggests stronger integration into scholarly infrastructure: library systems, indexing platforms, citations, institutional links and recognized journal platforms. Predatory publishers, by contrast, are more dependent on open-web discovery and attract shallower engagement. They are not simply smaller versions of established publishers; they have a different digital profile.

The country mapping also requires a careful interpretation. High absolute traffic from the United States, India, Brazil, China and the Philippines indicates that exposure is global. It does not demonstrate authorship or culpability. The exploratory researcher-normalized mapping suggests that relative exposure may be especially high in some settings where research systems are expanding and where publication guidance, mentoring and trusted open-access routes may be uneven. Interventions should therefore be differentiated: large research systems need safeguards because absolute exposure is high, while smaller or fast-growing systems may need stronger capacity building because relative exposure may be high.

For publishers, indexing services and infrastructure providers, the results point to the importance of trust signals. Referral traffic appears to be a marker of integration into legitimate scholarly infrastructures. Strengthening reliable metadata, persistent identifiers, journal-authentication services, link resolver quality and transparent editorial information may help users distinguish between credible and deceptive platforms. For libraries and research offices, the practical task is not only to warn authors about suspect invitations, but to teach source evaluation in the search environment itself.

'For publishers, indexing services and infrastructure providers, the results point to the importance of trust signals'

This means that training should cover publisher landing pages, indexing claims, editorial board information, article processing charges, journal portfolios, metrics badges, article PDFs and domain-level credibility. It also means that scholarly communication support should be connected to evidence synthesis, doctoral training, open-access advice and research integrity programmes. The problem arises before publication decisions, during routine searching and reading, when users form first impressions about whether a source appears trustworthy.

Limitations and future research

The study has limitations. Semrush provides estimated traffic rather than audited server logs. Estimates may be less accurate for small sites and may change retrospectively according to proprietary methods that cannot be fully audited. The figures reported here should therefore be treated as indicators of relative magnitude, concentration and pattern rather than exact visit counts.

The sampling frame relied on an external predatory publisher list. Such lists are valuable for constructing empirical samples, but they are contested, dynamic and heterogeneous. Some publishers change domains, disappear, rebrand or modify practices. The study therefore does not make a legal or editorial judgement about each individual record. Its claims concern the aggregate web footprint of a sampled population of publisher records classified as predatory by an external registry.

The publisher record was used as the main unit of analysis, while web traffic is technically domain- or URL-based. Some records share identical or related domains and some groups operate multiple brands. Sensitivity checks that collapsed duplicate URLs reduced aggregate traffic but did not alter the main conclusions about concentration, search dependence or country ranking. Future research should refine the distinction between publisher, journal, domain, subdomain and article-level URL.

‘Future research should refine the distinction between publisher, journal, domain, subdomain and article-level URL’

Country attribution reflects apparent IP-based traffic origin, not nationality, institutional affiliation, discipline or user role. VPN use can cause a visit to be attributed to the exit server’s location rather than the user’s physical location. Visits may come from authors, readers, students, librarians, editors, automated agents or crawlers. Engagement indicators are also behavioural proxies: a high bounce rate may indicate shallow exposure, but it may also reflect article-level access in which the user obtains what they need from a single page. Future research should link web traffic to article-level data, indexing status, retractions, citations, policy-document citations and qualitative studies with users.

Conclusion

Predatory publishing should be understood not only as a publication-market problem, but also as a search-mediated exposure problem. The sampled predatory publisher records accumulated an estimated 244.1 million organic visits from 2019 to 2024, with traffic highly concentrated among a small number of visible operators. Compared with established publishers, predatory publishers relied more strongly on organic search, received less referral traffic and generated shallower engagement.

For the scholarly communication community, the message is practical. Efforts to reduce the harms of predatory publishing should combine author education with stronger discovery, metadata and trust infrastructures. Researchers need support not only at the point of manuscript submission but also during searching, reading, citation and evidence appraisal. Publishers, librarians, research offices, indexers and platform providers all have a role in making trustworthy scholarly communication easier to recognize and deceptive publishing harder to encounter without warning.

‘Efforts to reduce the harms of predatory publishing should combine author education with stronger discovery, metadata and trust infrastructures’

Data accessibility statement

The data underlying this article are available from the corresponding author, Rubén Comas-Forgas, upon reasonable request.

Abbreviations and Acronyms

A list of the abbreviations and acronyms used in this and other Insights articles can be accessed here – click on the following URL and then select the 'full list of industry A&As' link: **http://www.uksg.org/publications#aa**.

Competing interests

The authors have declared no competing interests.

Declaration of AI-assisted technologies

During the preparation of this work, the authors used ChatGPT to support language editing, content organization and journal-format adaptation. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the manuscript.

Author contributions (CRediT)

Rubén Comas-Forgas: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Visualization, Writing – original draft, Writing – review & editing, Project administration. Mercè Morey-López: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Visualization, Writing – original draft, Writing – review and editing. Manos Pavlakis: Validation, Writing – review & editing. Hiromi M. Yokoyama: Validation, Writing – review and editing. All authors reviewed and approved the final manuscript.

References

Albert, M. A., Lalu, M. M., & Grudniewicz, A. (2025). Investigating the trustworthiness of research evidence used to inform public health policy: A qualitative interview study on the use of predatory journal citations in policy documents. *Health Research Policy and Systems*, *23*, 7. **https://doi.org/10.1186/s12961-024-01282-9** (accessed 12 October 2026).

Beall, J. (2012). Predatory publishers are corrupting open access. *Nature*, *489*(7415), 179. **https://doi.org/10.1038/489179a** (accessed 12 October 2026).

Björk, B.-C., Kanto-Karvonen, S., & Harviainen, J. T. (2020). How frequently are articles in predatory open access journals cited? *Publications*, *8*(2), 17. **https://doi.org/10.3390/publications8020017** (accessed 14 October 2026).

Björneborn, L., & Ingwersen, P. (2004). Toward a basic framework for webometrics. *Journal of the American Society for Information Science and Technology*, *55*(14), 1216–1227. **https://doi.org/10.1002/asi.20077** (accessed 14 October 2026).

Cobey, K. D., Lalu, M. M., Skidmore, B., Ahmadzai, N., Grudniewicz, A., & Moher, D. (2018). What is a predatory journal? A scoping review. *F1000Research*, *7*, 1001. **https://doi.org/10.12688/f1000research.15256.2** (accessed 15 October 2026).

Cukier, S., Helal, L., Rice, D. B., El-Saadany, S., Ahmadzai, N., Wilson, M., & Moher, D. (2020). Checklists to detect potential predatory biomedical journals: A systematic review. *BMC Medicine*, *18*, 104. **https://doi.org/10.1186/s12916-020-01566-1** (accessed 14 October 2026).

Demir, S. B. (2018). Predatory journals: Who publishes in them and why? *Journal of Informetrics*, *12*(4), 1296–1311. **https://doi.org/10.1016/j.joi.2018.10.008** (accessed 14 October 2026).

Dobson, H. (2016). Think.Check.Submit.: the campaign helping researchers navigate the scholarly communication landscape. *Insights*, *29*(3), 228–232. **https://doi.org/10.1629/uksg.323** (accessed 14 October 2026).

Frandsen, T. F. (2019). Why do researchers decide to publish in questionable journals? *Learned Publishing*, *32*(3), 217–222. **https://doi.org/10.1002/leap.1214** (accessed 14 October 2026).

Grudniewicz, A., Moher, D., Cobey, K. D., Bryson, G. L., Cukier, S., Allen, K., Ardern, C., Balcom, L., Barros, T., Berger, M., Ciro, J. B., Cugusi, L., Donaldson, M. R., Egger, M., Graham, I. D., Hodgkinson, M., Khan, K. M., Mabizela, M., Manca, A., . . . Lalu, M. M. (2019). Predatory journals: No definition, no defence. *Nature*, *576*, 210–212. **https://doi.org/10.1038/d41586-019-03759-y** (accessed 14 October 2026).

Jansen, B. J., Jung, S., & Salminen, J. (2022). Measuring user interactions with websites: A comparison of two industry standard analytics approaches using data of 86 websites. *PLOS One*, *17*(5), e0268212. **https://doi.org/10.1371/journal.pone.0268212** (accessed 19 October 2026).

Predatory Journals. (2024). *The Predatory publishers list*. Publishers. **https://www.predatoryjournals.org/the-list/publishers** (accessed 16 October 2026).

Shen, C., & Shah, L. (2023). Predatory publishing practices: what researchers should know before submitting their manuscript. *Insights*, *36*, 19. **https://doi.org/10.1629/uksg.631** (accessed 14 October 2026).

Strinzel, M., Severin, A., Milzow, K., & Egger, M. (2019). Blacklists and whitelists to tackle predatory publishing: A cross-sectional comparison and thematic analysis. *mBio*, *10*(3), e00411–19. **https://doi.org/10.1128/mBio.00411-19** (accessed 19 October 2026).

Thelwall, M. (2009). *Introduction to webometrics: Quantitative web research for the social sciences*. Springer International Publishing. **https://link.springer.com/book/10.1007/978-3-031-02261-6** (accessed 19 October 2026).

UNESCO Institute for Statistics. (2025). UIS.Stat database. **https://data.uis.unesco.org/index.aspx?queryid=66** (accessed 26 March 2025).

World Bank. (2025). Researchers in R&D (per million people). **https://data.worldbank.org/indicator/SP.POP.SCIE.RD.P6** (accessed 26 March 2025).

Article copyright: © 2026 Rubén Comas-Forgas, Mercè Morey-López, Manos Pavlakis and Hiromi M. Yokoyama. This is an open access article distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use and distribution provided the original author and source are credited.

Corresponding author:
Rubén Comas-Forgas
Professor
University of the Balearic Islands (UIB)
Spain
E-mail: rubencomas@uib.es
ORCID ID: https://orcid.org/0000-0002-8885-753X

Co-authors:
Mercè Morey-López
ORCID ID: https://orcid.org/0000-0002-2050-6602

Manos Pavlakis
ORCID ID: https://orcid.org/0000-0001-9814-9222

Hiromi M. Yokoyama
ORCID ID: https://orcid.org/0000-0003-4760-762X

To cite this article:
Comas-Forgas R, Morey-López M, Pavlakis M, Yokoyama HM, "Predatory publishers in the search environment: web traffic evidence for scholarly communication," *Insights*, 2026, 39: 23, 1–11; DOI: https://doi.org/10.1629/uksg.796

Submitted on 5 July 2026 Accepted on 3 September 2026 Published on 24 November 2026

Published by UKSG in association with Ubiquity Press.