

The costs of late quality signals in scholarly publishing

Abdelghani Maddi

`abdelghani.maddi@cnrs.fr`

Groupe d'Étude des Méthodes de l'Analyse Sociologique de la Sorbonne

Article

Keywords:

Posted Date: October 22nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11296438/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

The costs of late quality signals in scholarly publishing

Abdelghani Maddi*

* Sorbonne Université, CNRS, Groupe d'Étude des Méthodes de l'Analyse Sociologique de la Sorbonne, GEMASS, Paris, France

Correspondence to: abdelghani.maddi@cnrs.fr

Journal discontinuation for quality concerns acts on an already accumulated publication history, but the scale of scientific and financial exposure surrounding such decisions is poorly quantified. Here we link 722 journals recorded by Scopus as discontinued for quality issues to 1.13 million articles and reviews in OpenAlex, observed OpenAPC payments and annual article-processing-charge price data. Nearly one quarter (24.7%) of recovered works appeared after the recorded discontinuation year. Relative to matched journals that remained indexed, publication output changed little in the event year but diverged progressively thereafter, reaching -33.9% after one year, -60.9% after two years and -79.7% after five years; the mean post-event contrast was -62.5%. Observed payments cover only a small subset, but direct and context-supported price evidence documents APC exposure in the tens of millions of US dollars. Complete extrapolation is dominated by weak imputation and is therefore treated only as a sensitivity analysis. These results reveal a temporal asymmetry in journal-level quality signalling: substantial scientific and financial commitments can accumulate before a formal signal, while adjustment unfolds over several years.

Bibliographic databases such as Scopus provide structured coverage of scholarly journals and articles.^{1,2} Coverage decisions can change over time, but discontinuation acts on an already accumulated publication record: previously indexed articles remain in the scholarly

record, and journals may continue to publish. By the time a formal journal-level signal appears, substantial scientific and financial commitments may already exist, and the response to that signal may unfold only gradually.

Concerns about editorial standards, rapid publication growth and financial incentives have intensified in parts of scholarly publishing, particularly in open-access markets.^{3–5} Yet labels such as predatory or low quality remain contested and do not map cleanly onto individual articles.⁴ We therefore avoid assigning these labels ourselves. Our primary cohort is defined only by the wording of a historical Scopus record that marked journals as discontinued due to quality issues. The flag is used as a journal-level provenance signal, not as a judgement on the validity of every article the journals published.

To examine the scale and timing of this exposure, we combine the Scopus signal with OpenAPC payment records,⁶ annual APC price data^{7–10} and OpenAlex publication histories.¹¹ This allows us to ask three questions: how much scholarly output has accumulated by the event; how publication trajectories evolve afterwards relative to comparable journals that remain indexed; and how large the associated APC exposure is under increasingly broad price evidence. The resulting cohort comprises 722 journals and more than 1.1 million articles and reviews.

Large exposure precedes the formal signal

The August 2026 Scopus list contained 999 discontinuation events. Linking these records to the historical quality-issue field identified 722 journals, of which 296 also overlapped the stricter GhoS(t)copus publication-concerns cohort.³ Among the 722 journals, 640 matched an OpenAlex source and 628 contained at least one article or review. After source-level

recovery and deterministic deduplication, we retrieved 1,127,977 unique works, or 99.98% of the 1,128,194 records expected from OpenAlex source counts (Fig. 1).

Most exposure was already present by the event year, but publication continued afterwards. Of the recovered works, 630,519 preceded the event year, 219,085 appeared during it and 278,227 followed it; 146 could not be assigned a usable event-year relation. Overall, 24.7% of recovered works were post-event. This does not imply that authors submitted after learning of the indexing decision: publication years are coarse, submission dates were not consistently available and we do not observe when the discontinuation became known to prospective authors.

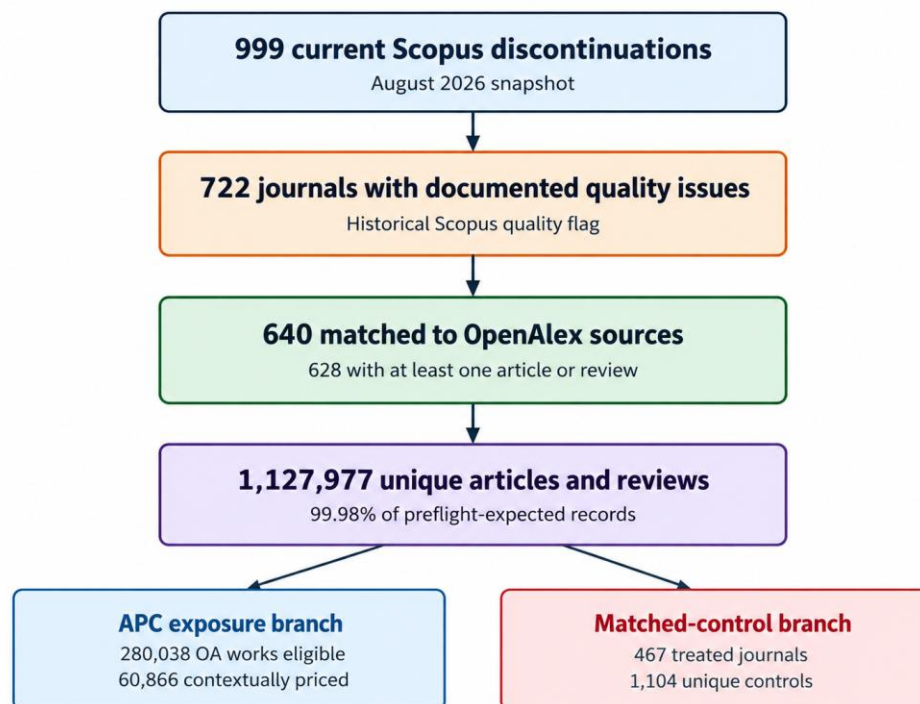


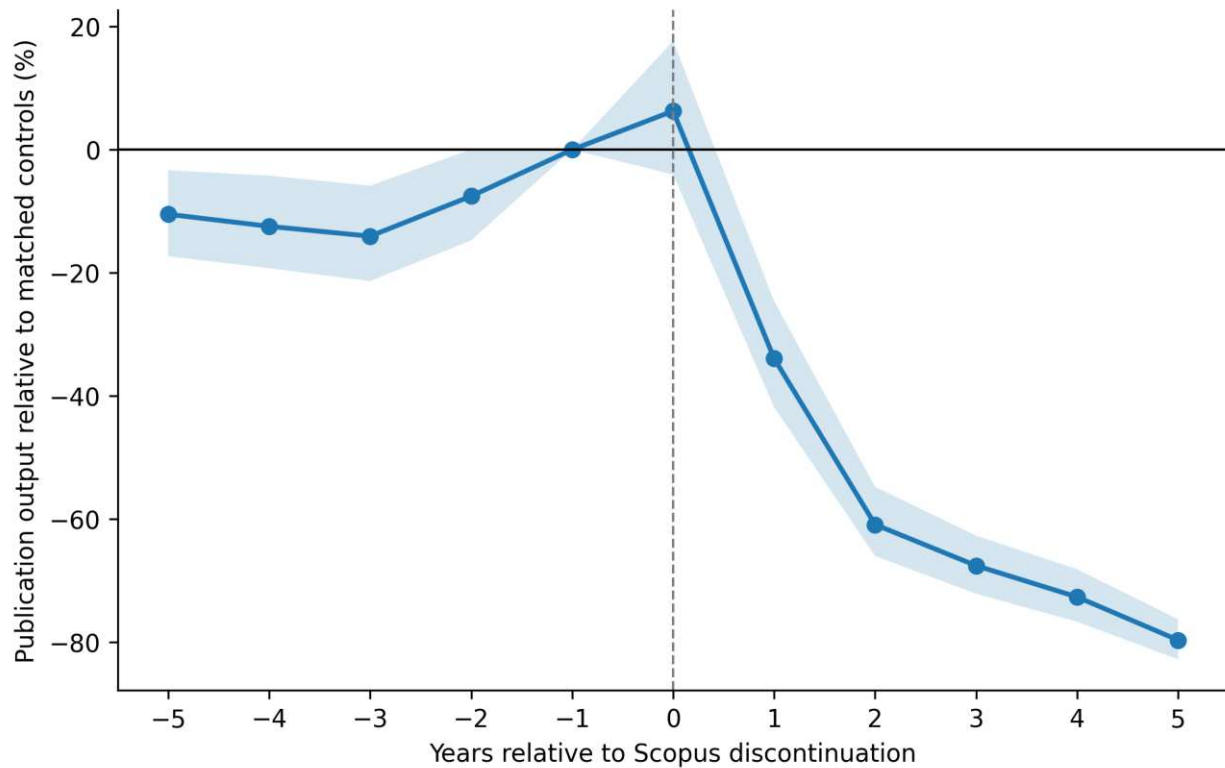
Figure 1 | Construction of the quality-signal cohort and analytical branches. Current Scopus discontinuations were linked to the historical Scopus quality-issue flag, OpenAlex sources, APC price evidence and matched active controls. Counts refer to distinct journals or works at the stage shown.

60 **Publication trajectories diverge after discontinuation**

61 In the balanced 11-year panel of 555 journals, output fell from 406,149 works in years -5 to
62 -1 to 211,443 in years +1 to +5, an aggregate decline of 47.9%. The median journal-level
63 change was -47.7%; 68.8% of journals declined and 31.0% increased. The shift was
64 gradual rather than immediate: the median output index, normalized to each journal's pre-
65 event mean of 100, was 117 at year -1, 75 at +1, 45 at +2 and 12 at +5.

66 To separate journal-specific change from broader publishing trends, we matched affected
67 journals to active Scopus journals with no current discontinuation event or historical quality
68 flag. Matching used only pre-event information: field and open-access status for eligibility,
69 then five-year publication level, pre-event slope, year -1 output, journal age and trajectory
70 similarity. In the primary specification, 467 affected journals were matched to 1,104 unique
71 controls.¹²

72 At the discontinuation year, the treated-control difference was small and imprecise (+6.3%;
73 95% bootstrap confidence interval, -4.0% to +17.7%). The gap then widened to -33.9% at
74 year +1, -60.9% at +2 and -79.7% at +5 (Fig. 2). Averaged across years +1 to +5 relative
75 to years -5 to -1, the matched difference-in-differences contrast was -62.5% (95%
76 bootstrap confidence interval, -66.8% to -58.2%); 78.6% of affected journals had a
77 negative contrast.



78

79 Figure 2 | Publication output diverges after Scopus discontinuation. Effects are percent transformations of
80 treated-control differences in $\log(1 + \text{annual articles and reviews})$ relative to year -1. Shading denotes 95%
81 confidence intervals from 1,000 bootstrap resamples clustered at the treated-journal level. The dashed line
82 marks the discontinuation year.

83 This pattern was stable across design choices. Across nine specifications varying control
84 ratio and matching caliper, mean post-event contrasts ranged from -64.6% to -59.1%;
85 excluding year -1 from the pre-event baseline yielded -61.6% (95% bootstrap confidence
86 interval, -65.9% to -56.1%). Re-basing the event study to year -2 produced modest
87 negative coefficients at years -5 to -3, but the differential linear pre-trend over years -5 to -
88 2 was +0.79% per year (95% bootstrap confidence interval, -2.09% to +3.41%; bootstrap P
89 = 0.585). The data therefore do not establish perfectly parallel pre-trends, but the post-
90 event divergence is not driven by the event-adjacent pre-year.

APC exposure remains substantial under stronger evidence

Observed payments provide the strongest but sparsest financial evidence. In OpenAPC, the strict current-event quality-evidence subset contained 266 article-payment records from 55 journals, totalling EUR 205,312. Only 84 of the 722 journals matched OpenAPC at journal level, so these records document observed spending but cannot support a global total.

Broader estimates used 280,038 OpenAlex works classified as gold or hybrid open access and eligible for APC pricing. Exact historical journal- or source-year price evidence covered 11,730 works and implied US\$17.9 million in calibrated exposure. Adding direct work-level OpenAlex list prices increased coverage to 49,599 works and US\$47.2 million. Extending the hierarchy to publisher-year and year-by-open-access-status medians, while excluding the weakest global-status median, covered 60,866 works and yielded US\$79.0 million, with a calibration-based range of US\$73.6-87.7 million (Fig. 3).

Complete extrapolation to all 280,038 eligible works produced US\$698.3 million (US\$650.8-775.8 million), but 88.7% of that estimate came from the weakest global open-access-status imputation. We therefore treat it as a sensitivity analysis, not as an estimate of money paid. The distinction is also visible in concentration tests: removing the two journals with the largest full-extrapolation values reduced that total to 29.9% of its original value, whereas 96.3% of the contextual A-F estimate remained. The better-supported tiers therefore establish APC exposure on a tens-of-millions scale without relying on the full extrapolation.

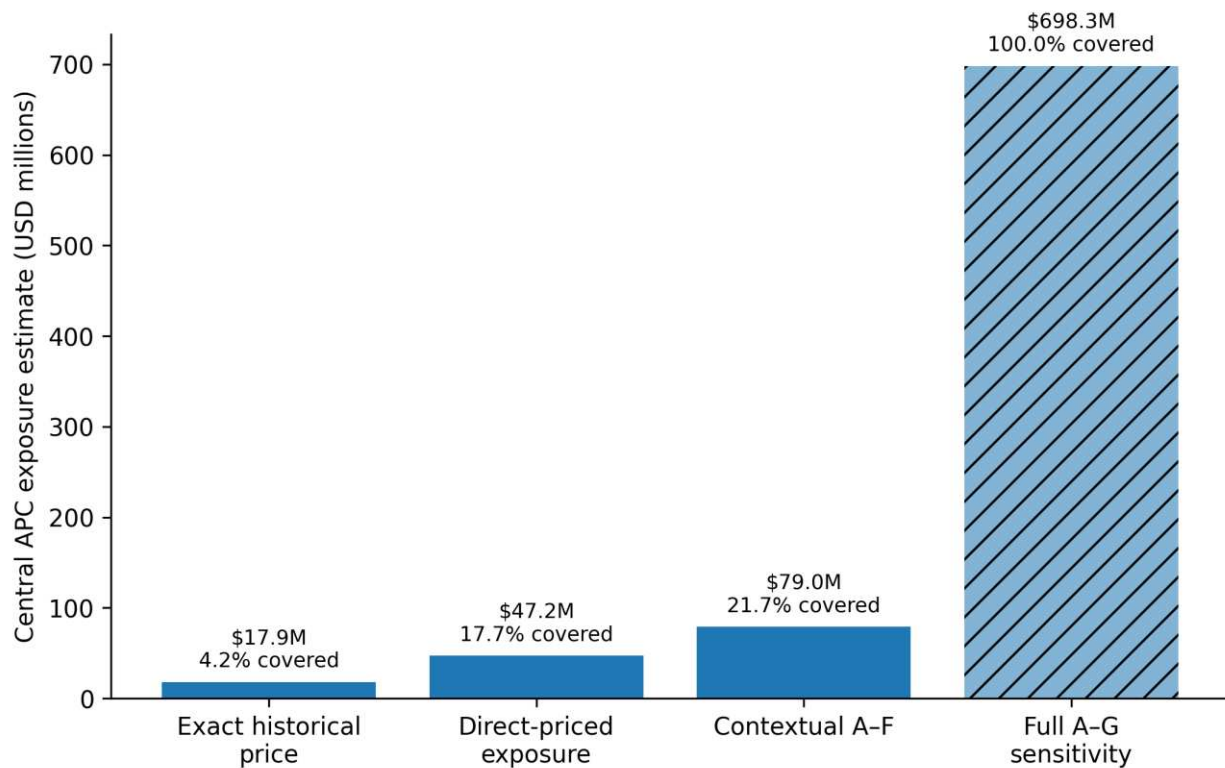


Figure 3 | APC exposure by price-evidence tier. Exact and direct-price tiers use journal/year or work-level list-price evidence. The contextual A-F estimate adds publisher/year and year/open-access-status medians. The full A-G extrapolation additionally imputes a global open-access-status median and is shown as sensitivity only. Values are calibrated central estimates in US dollars, not observed payments.

Late quality signals have a long tail

Together, the analyses reveal a long tail around formal journal-quality signals. More than 1.1 million articles and reviews were linked to journals in the quality-evidence cohort, and nearly one quarter of recovered works appeared after the recorded event year. Publication did not stop at discontinuation; trajectories instead diverged progressively from matched active journals, with little change in the event year and a deficit approaching four-fifths by year +5. Better-supported price evidence places APC exposure in the tens of millions of US dollars. The central pattern is therefore temporal: substantial commitments can accumulate before a formal signal, while adjustment unfolds over years.

126 The financial estimates should be read as a hierarchy of evidence, not as a single global
127 bill. Observed payments are strongest but sparse. Direct list prices broaden coverage but
128 do not capture discounts, waivers, transformative agreements, local taxes or payer
129 identity. Contextual imputation broadens coverage further at the cost of additional
130 assumptions. We therefore interpret US\$47.2 million under direct price evidence and
131 US\$79.0 million under the contextual A-F hierarchy as price-based exposure within
132 covered subsets. The US\$698.3 million full extrapolation is informative mainly as a
133 demonstration of how rapidly totals become assumption-dependent.

134 Several limitations matter. The historical Scopus flag is a journal-level classification, not an
135 adjudication of individual articles. Matching journal identities across Scopus, OpenAlex,
136 OpenAPC and price files is vulnerable to title and ISSN changes and publisher transfers.
137 Event years do not identify the exact date at which indexation changed or when authors
138 became aware of it. Matching cannot eliminate confounding, and the event study does not
139 establish a causal effect of discontinuation. APC estimates rely mainly on list prices rather
140 than invoices. Finally, we measure publication volume and financial exposure, not article
141 validity, citation effects or downstream clinical consequences.

142 These limitations sharpen the main implication. Journal-level signals are retrospective and
143 cannot substitute for article-level assessment. By the time a formal signal appears,
144 substantial scientific and financial commitments may already be embedded in the scholarly
145 record, while adjustment may take years. More transparent status histories, interoperable
146 reason codes and more complete reporting of publication payments would make these
147 temporal lags easier to observe and study without turning journal-level signals into
148 simplistic blacklists.

149 References

- 150 1. Baas, J., Schotten, M., Plume, A., Côté, G. & Karimi, R. Scopus as a curated, high-quality bibliometric
151 data source for academic research in quantitative science studies. *Quant. Sci. Stud.* **1**, 377–386 (2020).
- 152 2. Elsevier. Scopus content policy and selection | Elsevier. [www.elsevier.com](https://www.elsevier.com/products/scopus/content/content-policy-and-selection)
153 <https://www.elsevier.com/products/scopus/content/content-policy-and-selection> (2026).
- 154 3. Cortegiani, A. *et al.* Citations and metrics of journals discontinued from Scopus for publication
155 concerns: the GhoS(t)copus Project. Preprint at <https://doi.org/10.12688/f1000research.23847.2> (2020).
- 156 4. Grudniewicz, A. *et al.* Predatory journals: no definition, no defence. *Nature* **576**, 210–212 (2019).
- 157 5. Shen, C. & Björk, B.-C. ‘Predatory’ open access: a longitudinal study of article volumes and market
158 characteristics. *BMC Med.* **13**, 230 (2015).
- 159 6. Jahn, N. & Tullney, M. A study of institutional spending on open access publication fees in Germany.
160 *PeerJ* **4**, e2323 (2016).
- 161 7. Butler, L.-A., Matthias, L., Simard, M.-A., Mongeon, P. & Haustein, S. The oligopoly’s shift to open
162 access: How the big five academic publishers profit from article processing charges. *Quant. Sci. Stud.* **4**,
163 778–799 (2023).
- 164 8. Haustein, S. *et al.* Estimating global article processing charges paid to six publishers for open access
165 between 2019 and 2023. Preprint at <https://doi.org/10.48550/arXiv.2407.16551> (2024).
- 166 9. Matthias, L. *et al.* An open dataset of article processing charges from 14 scholarly publishers, 2019–
167 2025. Harvard Dataverse <https://doi.org/10.7910/DVN/AZ985C> (2026).
- 168 10. Matthias, L. *et al.* A dataset of article processing charges from 14 scholarly publishers, 2019–2025.
169 Preprint at <https://doi.org/10.48550/arXiv.2608.14116> (2026).
- 170 11. Priem, J., Piwowar, H. & Orr, R. OpenAlex: A fully-open index of scholarly works, authors, venues,
171 institutions, and concepts. Preprint at <https://doi.org/10.48550/arXiv.2205.01833> (2022).
- 172 12. Stuart, E. A. Matching methods for causal inference: A review and a look forward. *Stat. Sci. Rev. J. Inst.*
173 *Math. Stat.* **25**, 1–21 (2010).
- 174 13. OpenAPC contributors. Datasets on fee-based Open Access Publishing. *Datasets on fee-based Open*
175 *Access Publishing* <https://openapc.github.io/> (2026).

177 Figure legends

178 Figure 1 | Construction of the quality-signal cohort and analytical branches. Current Scopus discontinuations
179 were linked to the historical Scopus quality-issue flag, OpenAlex sources, APC price evidence and matched
180 active controls. Counts refer to distinct journals or works at the stage shown.

181 Figure 2 | Publication output diverges after Scopus discontinuation. Effects are percent transformations of
182 treated-control differences in $\log(1 + \text{annual articles and reviews})$ relative to year -1. Shading denotes 95%
183 confidence intervals from 1,000 bootstrap resamples clustered at the treated-journal level; the dashed line
184 marks the discontinuation year.

185 Figure 3 | APC exposure by price-evidence tier. Exact and direct-price tiers use journal/year or work-level
186 list-price evidence. The contextual A-F estimate adds publisher/year and year/open-access-status medians.
187 The full A-G extrapolation additionally imputes a global open-access-status median and is shown as
188 sensitivity only. Values are calibrated central estimates in US dollars, not observed payments.

Methods

Journal cohort and provenance

We combined current and historical Scopus title lists with OpenAlex, OpenAPC, the GhoS(t)copus journal dataset and annual APC list prices. The current Scopus source list was the August 2026 snapshot. We retained records whose current indexation change was 'Discontinuation' and excluded 'Journal change policy' events. A historical Scopus file supplied by the authors (March 2023 file; source sheet labelled October 2022) contains the field 'Titles discontinued by Scopus due to quality issues'. A positive value in this field defined the primary quality-evidence cohort. The GhoS(t)copus dataset³ was used as a stricter secondary publication-concerns flag, not as the sole cohort definition. Absence of either historical flag was coded as reason unavailable, not as evidence of absence of quality concerns. The study analysed bibliographic and institutional payment metadata and did not involve human participants.

OpenAlex publication universe

Scopus titles were matched to OpenAlex sources using normalized ISSNs and EISSNs and title evidence. We retrieved article and review records for quality-evidence sources through the OpenAlex API in September 2026. Source-level preflight counts were used to audit completeness; under-retrieved sources were re-fetched source by source. OpenAlex work identifiers were deduplicated after assigning a deterministic canonical Scopus source/event mapping. The final dataset contained 1,127,977 unique works, 99.98% of the OpenAlex preflight expectation.¹¹

Observed and estimated APC exposure

OpenAPC records¹³ were used only as observed payment evidence and for price calibration. Corresponding-author locations from OpenAlex were treated as scholarly attribution rather than payer identity. Gold and hybrid open-access works were eligible for model-based APC pricing. We applied a prespecified hierarchy: A, exact journal/source-year price, preferring the Matthias et al. annual price dataset^{9,10}; B, work-level OpenAlex list price; C, nearest same-journal/source price within plus or minus two years when available; D, publisher-year plus open-access-status median; E, publisher-year median; F, year plus open-access-status median; and G, global open-access-status median. Levels D-G are explicit imputations. The primary contextual estimate includes A-F; level G is reserved for the complete-extrapolation sensitivity analysis. List-price values were calibrated against OpenAPC payments with exact Matthias et al. EUR prices. Across 102 matched cases, the paid-to-list ratio had a 25th percentile of 0.839 and a median of 0.900; these defined low and central multipliers, while the high scenario used 1.0.

Balanced panels and matched controls

The descriptive publication panel included journals with a complete plus-or-minus-five-year observation window and nonzero pre-event output; the event year was omitted from pre/post totals. For matched controls, candidate journals had to be active in 2026, absent from current Scopus discontinuation or change events, without a historical quality flag and matchable to OpenAlex. Only complete calendar years through 2025 were used in the primary plus-or-minus-five-year event window. Static matching tiers prioritized broad field and Scopus open-access status. Final distance used only pre-event features over years -5 to -1: log mean publication level, log publication slope, log year -1 output, journal age and the root-mean-square error of the log publication trajectory, scaled by global standard

deviations and combined with prespecified weights. The primary specification required three controls within caliper 4.0; controls could be reused across treated journals.¹²

Event-study statistics and robustness

For each matched set and relative year, the outcome was $\log(1 + \text{annual article/review count})$. The control outcome was the equal-weight mean across matched controls. The event-study effect was the treated-control log gap minus its value at year -1, transformed as $100[\exp(\beta)-1]$. The primary pre/post contrast compared the mean log outcome over years -5 to -1 with years +1 to +5; year 0 was excluded. Confidence intervals were obtained by 1,000 bootstrap resamples clustered at the treated-journal level using a fixed seed (20260926). Sensitivity analyses varied controls per treated journal (1, 3 or 5) and calipers (3, 4 or 5). A robustness analysis re-based the event study to year -2, recomputed the pre/post contrast using years -5 to -2 and estimated the linear differential pre-trend over years -5 to -2 with treated-journal bootstrap inference. All matched estimates are associations conditional on the design and are not interpreted as causal effects.

Software and generative AI

The analytical workflow was implemented in R with cached OpenAlex retrieval, deterministic journal/source matching, audit tables and fixed bootstrap seeds. OpenAI ChatGPT (GPT-5.6 Sol; accessed September 2026) was used to assist with analysis-code development and debugging. All analytical decisions, code, numerical outputs, references and manuscript are made by the author. No generative image model was used to synthesize or alter scientific image content; figures are deterministic plots of study data generated with standard plotting software (R, version 4.4.1).

256 **Data availability**

257 Code and derived data required to reproduce the figures and tables will be deposited at
258 Zenodo before publication. Scopus source-title lists are maintained by Elsevier and openly
259 available. OpenAlex metadata are openly available through OpenAlex. OpenAPC data are
260 available from the OpenAPC release archive. The annual APC price dataset is available
261 under CC0 from Harvard Dataverse.^{9,11,13}

262 **Code availability**

263 The R code used for source matching, OpenAlex retrieval, APC estimation, matched-
264 control construction, event-study estimation, robustness analyses and figure generation
265 will be archived at Github. API credentials and other secrets are excluded from the archive.

266 **Funding**

267 This study was supported by the ÉCOS (Éclairer les Circuits de l'Open Science) project,
268 funded by URFIST-Rennes (France).

269 **Competing interests**

270 The author has no competing interest to declare.

271 **Additional information**

272 Supplementary Information is available for this paper. Correspondence and requests for
273 materials should be addressed to Abdelghani Maddi (abdelghani.maddi@cnrs.fr).

274 **Extended Data legends**

275 Extended Data Figure 1 | Event study re-based to year -2. Effects are percent
276 transformations of treated-control log gaps relative to year -2, with 95% treated-journal

277 bootstrap confidence intervals. This analysis isolates the event-adjacent pre-year (-1) from
278 the reference period.

279 Extended Data Figure 2 | Matching sensitivity. Mean post-event contrasts across nine
280 specifications varying the number of controls per treated journal (1, 3 or 5) and matching
281 caliper (3, 4 or 5). The dashed line indicates the primary 3:1, caliper-4 estimate.