

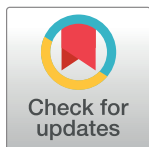
PERSPECTIVE

A call for research to address the threat of paper mills

Jennifer A. Byrne^{1,2*}, Anna Abalkina³, Olufolake Akinduro-Aje⁴, Jana Christopher⁵, Sarah E. Eaton⁶, Nitin Joshi⁷, Ulf Scheffler⁸, Nick H. Wise⁹, Jennifer Wright¹⁰

1 NSW Health Statewide Biobank, NSW Health Pathology, Camperdown, Australia, **2** School of Medical Sciences, Faculty of Medicine and Health, University of Sydney, Camperdown, Australia, **3** Institute for East European Studies, Freie Universität Berlin, Berlin, Germany, **4** Taylor & Francis Group, Milton Park, Milton, United Kingdom, **5** FEBS, Cambridge, United Kingdom, **6** Werklund School of Education, University of Calgary, Calgary, Canada, **7** Adis Journals, Springer Nature, Auckland, New Zealand, **8** Wiley-VCH GmbH, Research Integrity & Case Resolution, Berlin, Germany, **9** Clare College, University of Cambridge, Cambridge, United Kingdom, **10** Cambridge University Press, Cambridge, United Kingdom

* jennifer.byrne@health.nsw.gov.au



Research paper mills are covert organizations that provide low-quality or fabricated manuscripts to paying clients. As members of the United2Act Research Working Group, we propose 5 key research questions on paper mills that require resourcing and support.

OPEN ACCESS

Citation: Byrne JA, Abalkina A, Akinduro-Aje O, Christopher J, Eaton SE, Joshi N, et al. (2024) A call for research to address the threat of paper mills. *PLoS Biol* 22(11): e3002931. <https://doi.org/10.1371/journal.pbio.3002931>

Published: November 22, 2024

Copyright: © 2024 Byrne et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: NHMRC grant APP1184263 <https://www.nhmrc.gov.au/funding/find-funding/ideas-grants> (JAB) NHMRC grant APP2029249 <https://www.nhmrc.gov.au/funding/find-funding/ideas-grants> (JAB) The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: JAB receives NHMRC grant funding to study publications from paper mills. All opinions expressed in this commentary represent the views of the authors and do not represent the views of any listed employer or institution. Authors declare that they have no other competing interests.

Research paper mills are unethical organizations that produce manuscripts at scale using derivative, copied, and/or fabricated text or data sets [1,2]. Manuscripts can be sold to preexisting author teams or individual authorship positions can be sold before and/or after manuscript acceptance [1–5]. Some paper mills may offer other services, including editorial handling and peer review [3,4], post-publication communications [5], and citations to their products [6].

Paper mills are likely to take deliberate steps to conceal their activities and products, while scaling both to maximize profits [1–3,5,7]. It was estimated that between 2 and 46% of manuscripts received by journals between 2019 and 2021 were produced by paper mills [2]. The analysis of a single paper mill identified 450 publications authored by more than 800 scholars from 300 universities [5]. New capabilities to fabricate research manuscripts using generative AI [8] could further scale manuscript production, while also rendering some fabricated manuscripts more challenging to detect [9].

Paper mills have been discussed in the literature since 2013 [1,10], yet their products, operations, and services remain understudied [11]. The lack of empirical research undertaken to date could partly reflect challenges in studying covert activities [12], safety risks for both participants and researchers [5], assumptions that paper mills do not affect mainstream research [11], and limited dedicated funding. However, any lack of paper mill research seems likely to underestimate the problem. For example, paper mills could target many different fields that allow papers to be produced and concealed at scale, where peer review expertise is limited [7], and replication of published results remains challenging [9]. Journal early warning lists [13] and the outputs of single paper mills [5] also encompass research fields beyond those typically linked with paper mill activity [9,11].

Table 1. Five key questions (KQ's) for research to address the threat of paper mills.

Key questions	Topics	Expertise	Funding, support, enablers, outcomes
KQ1 What are the features of paper mill products?	<ul style="list-style-type: none"> • Paper mill product features • Targeted fields, topics, methodologies • Screening tool development • Impact of GenAI 	<ul style="list-style-type: none"> • Feature detection/sleuthing • Targeted field, topic, methodological expertise • Text mining • Screening tool development • GenAI/ LLM/ machine learning • Bibliometrics • Publishing 	<ul style="list-style-type: none"> • Funding opportunities could be linked to bonuses for discipline-specific research, to incentivize research on paper mills • (Re)-training opportunities to scale capacity (e.g., screening tool development and application) • Public/private partnerships (e.g., screening tool application) • Scaled post-publication correction capacity required to enable research translation
KQ2 What is the scale of the paper mill problem?	<ul style="list-style-type: none"> • Affected journals, publishers • Duration • Changes over time • Use of GenAI 		
KQ3 How do paper mills operate and evade detection?	<ul style="list-style-type: none"> • Operational models • Clients • Employees, required skills • Concealment • Collusion with brokers, researchers, journals, institutions 	<ul style="list-style-type: none"> • Local knowledge, languages • Working with vulnerable participants • Qualitative research • Contract cheating • Global internet commerce • Criminal networks 	<ul style="list-style-type: none"> • Minimize risks for both researchers and research participants • Consider offering incentives to research participants • Scaled post-publication correction capacity required to enable research translation
KQ4 To what extent are researchers and scholars aware of paper mills?	<ul style="list-style-type: none"> • Researcher awareness (across research fields, settings) • Creation, translation, scaling, delivery, update of awareness and education campaigns 	<ul style="list-style-type: none"> • Discipline expertise • Qualitative research • Student, researcher education • Communication 	<ul style="list-style-type: none"> • Awareness and educational campaigns to be informed by evidence from KQ's 1–3, 5 • Possibility for rapid translation through researcher training/education programmes
KQ5 How are paper mills affecting science and scholarship?	<ul style="list-style-type: none"> • Citations of paper mill products by original publications, reviews, patents, clinical trials, research databases • Impacts on research problem selection, directions, careers, student completions 	<ul style="list-style-type: none"> • Citation analysis • Bibliometrics • Text mining • Targeted field expertise • Qualitative research • Working with vulnerable participants 	<ul style="list-style-type: none"> • Partnerships between citation analysis experts and researchers in targeted fields • Capacity to inform policies on delayed research career progression, student completions

<https://doi.org/10.1371/journal.pbio.3002931.t001>

The lack of research on paper mills, relative to the potential size and significance of the problem, means that we need to describe knowledge gaps, necessary research and expertise, and how the resulting evidence could benefit all stakeholders who rely on publication integrity. United2Act represents an international group of stakeholders who are working together to address the challenge of paper mills. As members of the United2Act Research Working Group, we describe 5 key research questions (KQs) that we believe require resourcing and support, to improve our knowledge of paper mills and their impact on science and researchers (summarized in [Table 1](#)).

KQ1: What are the features of paper mill products?

Knowledge of the features of paper mill products enables their detection and deterrence. The largely incidental discoveries of publications from paper mills, combined with limited coordinated research across stakeholder groups, have likely produced incomplete and fragmented knowledge [9]. We need more complete and contemporary knowledge of both discipline-specific and -agnostic features of paper mill products to improve detection and deterrence.

KQ2. What is the scale of the paper mill problem?

Once features of paper mill support are known, these features must be leveraged to inform understanding of the scale of paper mill-supported publications, across fields, topics, methodologies, and over time. Scaled temporal literature analyses can allow the discovery of new

features, further informing KQ1, and answer questions such as when paper mills likely commenced using generative AI, and how their product features are likely to change [9].

KQ3. How do paper mills operate and evade detection?

While some features of paper mill operations could be predicted by models of operational rationality [12], we currently know very little about paper mill operations. Large, well-resourced paper mills could conceivably provide more sophisticated manuscripts and services [1,5] and could represent arms of other unethical businesses [1,14]. As some information on paper mill operations is now dated [1,3,10], we need research to inform current paper mill operations in different countries and settings, and the demographics of paper mill employees and clients. We also need current information about how paper mills collude with researchers, institutions, journals, and/or publishers to scale operations and evade detection, including the possible use of generative AI and commercially available screening tools. Understanding how paper mills produce manuscripts could also valuably inform product features (KQ1) and problem scale (KQ2).

KQ4. To what extent are researchers and scholars aware of paper mills?

While publishers are increasingly aware of paper mills [2,11], little is known about levels of awareness within other stakeholder groups. As awareness can help researchers to recognize and avoid paper mill products [7], research is required to inform the levels of awareness within different research and scholarly communities. The resulting information must then inform the design and implementation of awareness-raising campaigns and education programs, informed by evidence from KQs 1–3, and KQ5 below, to support all researchers, including those serving as journal editors and peer reviewers [2].

KQ5. How are paper mills affecting science and scholarship?

Just as there is currently limited information about awareness, little is known about how paper mill products are impacting research and scholarship. For example, researchers could unknowingly cite paper mill products, potentially slowing research translation [9] or impacting patient care [4]. We therefore need to understand how paper mill products are cited in different fields and across different publication types and how these citations influence author, publication, and journal metrics. With improved awareness of paper mills (KQ4), researchers could also change their practices, problem choices or research directions in response to publications from paper mills, where individual responses might vary according to research field and/or career stage. At present, we have only scattered anecdotal descriptions (for example, from social media) of how some researchers have been affected or are choosing to respond.

In summary, paper mills represent a threat to genuine research and scholarship that requires expertise, resources, and time [9]. In any call for research on paper mills, it is important to recognize ongoing efforts to address this problem. For example, journals and publishers are leveraging known paper mill features to detect submissions and retract publications, where cross-publisher collaborations such as the STM Integrity Hub provide education, shared tools, and workflows [2,9,11]. At the same time, paper mill product features and operations could rapidly evolve in response to improved detection and new capabilities [2,9,11]. Publishers and researchers alike will therefore benefit from timely research on paper mill products and operations to improve both awareness and responses (Table 1).

We recognize that paper mills represent a challenging topic for researchers and funders alike. Nonetheless, research funders and institutions must now take courageous decisions to

provide the necessary resources to transform our understanding of paper mills. Dedicated funding of paper mill research will also signal that paper mills and research fraud represent legitimate, important topics. Research support must enable rapid, ambitious research at scale, matched with systems that fast-track research translation to the literature and its many user communities (Table 1). For example, translation of paper mill research requires faster processes for achieving post-publication corrections at scale [9]. Knowing that papers can be quickly flagged where there is strong suspicion of mill involvement will also be a powerful motivator for researchers and funding agencies, who may likewise see little point in identifying problematic papers if these papers simply remain uncorrected.

Research in all fields relies upon the integrity of the literature. Paper mills and research misinformation undermine trust in research and should therefore be recognized as major global challenges, of similar importance to emerging pandemics or climate change. We must therefore waste no time in discovering the full extent of the paper mill problem and in taking all necessary steps to protect genuine science and scholarship.

Acknowledgments

The authors thank Sabina Alam (Taylor and Francis Group), Deborah Kahn (independent consultant), Hylke Koers (STM Solutions), Elizabeth Moylan (Wiley), Joris van Rossum (STM Solutions), and Alison Avenell (University of Aberdeen) for critical reading and discussions, and gratefully acknowledge support from United2Act, a project jointly funded by COPE and STM. We apologize to authors whose publications were not cited due to reference limits.

References

1. Hvistendahl M. China's publication bazaar. *Science*. 2013; 342:1035–1039. <https://doi.org/10.1126/science.342.6162.1035> PMID: 24288313
2. COPE, STM. Paper Mills—research report from COPE & STM—English. 2022. <https://doi.org/10.24318/jtbG8IHL> (2022).
3. Han J, Li Z. How metrics-based academic evaluation could systematically induce academic misconduct: A case study. *East Asian Sci Tech Soc*. 2018; 12:165–179.
4. Parker L, Boughton S, Lawrence R, Bero L. Experts identified warning signs of fraudulent research: a qualitative study to inform a screening tool. *J Clin Epidemiol*. 2022; 151:1–7. <https://doi.org/10.1016/j.jclinepi.2022.07.006> PMID: 35850426
5. Abalkina A. Publication and collaboration anomalies in academic papers originating from a paper mill: Evidence from a Russia-based paper mill. *Learned Publishing*. 2023; 36:689–702.
6. Candal-Pedreira C, Guerra-Tort C, Ruano-Ravina A, Freijedo-Farinas F, Rey-Brandariz J, Ross JS, et al. Retracted papers originating from paper mills: a cross-sectional analysis of references and citations. *J Clin Epidemiol*. 2024;111397. <https://doi.org/10.1016/j.jclinepi.2024.111397> PMID: 38815634
7. Byrne JA, Grima N, Capes-Davis A, Labbé C. The possibility of systematic research fraud targeting under-studied human genes: causes, consequences, and potential solutions. *Biomarker Insights*. 2019; 14:1177271919829162. <https://doi.org/10.1177/1177271919829162> PMID: 30783377
8. Elali FR, Rachid LN. AI-generated research paper fabrication and plagiarism in the scientific community. *Patterns*. 2023; 4(3). <https://doi.org/10.1016/j.patter.2023.100706> PMID: 36960451
9. Byrne JA, Park Y, Richardson RA, Pathmendra P, Sun M, Stoeger T. Protection of the human gene research literature from contract cheating organizations known as research paper mills. *Nucleic Acids Res*. 2022; 50:12058–12070. <https://doi.org/10.1093/nar/gkac1139> PMID: 36477580
10. Hu ZW, Wu YS. An empirical analysis on number and monetary value of ghostwritten papers in China. *Curr Sci*. 2013; 10:1230–1234.
11. Wittau J, Seifert R. How to fight fake papers: a review on important information sources and steps towards solution of the problem. *Naunyn Schmiedebergs Arch Pharmacol*. 2024; 6:1–4. <https://doi.org/10.1007/s00210-024-03272-8> PMID: 38970685
12. Fielding NG. The shaping of covert social networks: isolating the effects of secrecy. *Trends in Organized Crime*. 2017; 20:16–30.

13. Zhang L, Wei Y, Sivertsen G, Huang Y. The motivations and criteria behind China's list of questionable journals. *Learned Publishing*. 2022; 35:467–480.
14. Eaton SE, Carmichael JJ. Fake degrees and credential fraud, contract cheating, and paper mills: Overview and historical perspectives. In: Eaton SE, Carmichael JJ, Pethrick H, editors. *Fake Degrees and Fraudulent Credentials in Higher Education. Ethics and Integrity in Educational Contexts*. Springer; 2023. p. 1–22.