Duplicitous images in Alzheimer's research: systematic evidence, trends, and impacts

M. Tranchero[a,1], **C. Fons-Rosen**[b], **E. Scharfmann**[c], **and L. Fleming**[c,d]

This manuscript was compiled on March 29, 2025

Prior estimates of scientific malpractice have relied upon manual and expert inspection, public retraction, and self-report; these approaches remain difficult to scale and vulnerable to sample and reporting bias. Here we combine AI with manual validation to estimate the prevalence and different types of potentially fraudulent image duplication across the field of Alzheimer's research, both within and between papers, from 1980 to 2020. The 2.1% average incidence of duplication has risen over time and is greater for research 1) done in universities (relative to corporate and clinical settings), 2) employing animal and molecular methods (relative to human subjects methods), and 3) originating from China. Our assignment of malpractice category (sloppy, unclear, and problematic) correlates significantly with retraction and community identification of malpractice. Duplications are less frequent in papers with authors from multiple institutions and published in top journals. Papers with duplicated images are more likely to be cited by future publications and patents, but less likely to be cited in clinical research. Our findings help characterize trends and correlates of scientific malpractice and validate the use of AI as a valuable component of the peer review process.

scientific malpractice| image duplication | Alzheimer's | artificial intelligence

Significance Statement

Fraud in science can delay or prevent solutions to health, economic, or environmental problems and ultimately weaken public trust in science. Prior estimates of fraud have relied upon manual and expert inspection, public retraction, and self-report; these approaches remain difficult to scale and vulnerable to sample and reporting bias. Here we combine AI with manual validation to estimate fraudulent image duplication across the field of Alzheimer's research, from 1980 to 2020. The 2.1% average incidence of duplication has risen over time and is greater for research 1) done in universities (relative to corporate and clinical settings), 2) employing animal and molecular methods (relative to human subjects methods), and 3) originating from China. Our findings validate the use of AI as a valuable component of the peer review process.

The integrity of the scientific enterprise has recently been questioned across multiple fields by high-profile scandals and replication crises (1–3), particularly in the field of Alzheimer's research (4, 5). The societal impacts of research malpractice can be immediate - in reduced or misdirected funding for science - and fundamentally harmful - in delayed or failed solutions for health, economic, or environmental crises, and ultimately in reduced societal trust in science. Although many have estimated the types and prevalence of fraud, these estimates rely upon public retraction (6, 7), self-report (8), and manual expert inspection (9), and hence remain vulnerable to sample bias, reporting bias, and difficulties in scaling. As a result, few studies sample entire fields over time, establish the correlates of misconduct, or identify malpractice that might span multiple papers.

Here we focus on the prevalence of potentially fraudulent image duplications, a readily observed form of data malpractice (9), for Alzheimer's research, where several cases have recently surfaced, and a field which relies on images to communicate research results (4, 5). We combine AI-powered software with manual validation to assess the incidence of duplicate images, both within and across papers, in the corpus of peer-reviewed publications from 1980 to 2020. Over the time period, we find that 2.1% of all papers present duplicated images (1.7% if articles without images are included). The average incidence of duplication has increased over time and is greater in universities (relative to corporate and clinical settings), greater for animal and molecular methods (relative to human subjects methods), and greater for research originating from China.

Our combined AI and manual protocol flags 832 (out of 48,506) papers as presenting image malpractice, i.e., as sloppy, unclear, or problematic; these papers are significantly and monotonically more likely to have been retracted or identified in the PubPeer malpractice forum (particularly if the cause of retraction was image fraud). Additional analyses show that duplications are less frequent in papers with more authors, especially when affiliated with multiple institutions. Though we cannot assess the underlying rate of duplicitous submissions, there is more image malpractice in middle tier journals. Publications with duplicated images are more likely to be cited by future papers and patents, but less likely by future clinical studies.

Materials and Methods

We collected all PubMed papers in English (1980-2020) which were tagged with an Alzheimer's Medical Subject Heading (MeSH) keyword. This identified 302 journals

Author affiliations: [a]The Wharton School, University of Pennsylvania, Philadelphia, PA 19104; [b]Department of Economics, UC Merced, Merced CA 95343; [c]Haas School of Business, UC Berkeley, Berkeley, CA 94720; [d]College of Engineering, UC Berkeley, Berkeley, CA 94720

All authors contributed equally to the paper.

No authors have competing interests.

[1]To whom correspondence should be addressed: mtranc@wharton.upenn.edu.

DRAFT

and 50,604 academic papers, for which we could locate 48,506 PDFs, or 95.9%. In this population, 9,663 papers have no images, 17,457 between 1-3 images, and the remaining 21,386 have 4 or more images. From the papers that include images, the commercial AI tool Imagetwin identified 24,250 as presenting possible duplications. After ruling out obvious false positives (e.g., a duplication of the journal logo on the first page), we manually checked 9,655 papers that presented a possible duplication. Only cases independently confirmed by two of our team were kept (please see SI Figure S1 for an example). Our approach partially automates previously manual and meticulous efforts in which images were scanned "purely by eye" (9). Additional bibliometric data were merged from the OpenAlex and Reliance on Science databases (10, 11)

We find 832 Alzheimer's articles presenting inappropriate duplications, a number 4.4 and 9 times higher than the proportion of Alzheimer's papers flagged in PubPeer (a discussion forum often focused on scientific malpractice) or already retracted, respectively (SI Figure S2). Of the 832 papers, 671 papers present one duplicated image, 127 present two, 25 present three, and 9 present four or more. We validated the articles identified by Imagetwin and our manual inspections by comparing them to the 92 papers in our sample that are listed as retracted on Retraction Watch as of April 2024. The Retraction Watch data characterize the reasons associated with each retracted paper. Regressing our indicator of image duplication on the indicators of 1) retraction, 2) retraction due to image fraud, and 3) retraction due to image duplication, returned monotonically increasing coefficients of 0.267, 0.329, and 0.897, respectively (Table S1). Reassuringly, the subset of papers retracted because of image duplications correlates most highly with our dependent variable of duplicated images. Duplicate figures also appear to be the most important findings of the articles, as evinced by the higher number of mentions they receive in the text (Figure S3 and Table S2).

Building on previous research (9, 12), we manually developed a protocol to attribute the most likely intent of the duplication (see SI for the protocol and examples). We conservatively labeled papers with duplications only representing the same conditions across different experiments as "sloppy" (323 papers), papers with duplications used to represent different conditions across different experiments as showing "unclear" intent (320 papers), and papers with clearly manipulated or externally copied images as "problematic" (189 papers). We validated our protocol by regressing an indicator of retraction on our categorical indicators of a paper presenting sloppy, unclear, or problematic images. Estimates for papers presenting sloppy duplications are very close to zero and not statistically significant, while the converse is true for duplications in the unclear and problematic categories (Table S3). Furthermore, one should expect good-faith mistakes to be more likely as the number of images in papers grows. Indeed, we find that the correlation between sloppiness and the number of images is stronger than the correlation between problematic categorization and the number of images (Table S4).

Results: improper duplications have increased over time

Figure 1 illustrates that the share of published articles with duplicitous images has steadily increased over time. After rising from near zero in the 1980s, the incidence of duplications leveled off between 2.5% and 3% by 2020. Note that this is likely to be a conservative estimate of scientific malpractice since we focus only on the narrow instance of pictorial alteration (8, 13). While the last decade in our data witnessed an increase in duplications classified as sloppiness, the majority of inappropriate duplications remain borderline or outright problematic practices. Notably, the average number of images per paper has not grown as fast as duplications, ruling out a purely mechanical trend. Together, these findings suggest an increase in scientific malpractice in Alzheimer's research over 1980-2020.

The panels of Figure 2 investigate the relationships between duplication and research method, institutional affiliation, and country. We use MeSH keywords to characterize research methods (14), building on the idea that more basic science tends to use less complex model organisms (15) (Table S5). Panel (a) of Figure 2 illustrates a lower incidence of duplication in human-oriented research, relative to research on animals, cells, and molecules. Regressions controlling for journal and year fixed effects estimate that relative to the sample mean, articles employing molecular methods and model organisms are 79% and 183% more likely to present a duplication, respectively (Table S7). Our results are consistent with prior results that more basic and preclinical research is more difficult to replicate, potentially because human subjects research is more closely monitored (16).

Panel (b) of Figure 2 illustrates a higher incidence of image duplication in papers from the academy, while the incidence for papers from healthcare institutions is around the population mean and much less than the mean for company papers. The differences are robust in regressions that control for the methods used (Tables S8 and S9), ruling out that it is just a reflection of corporate researchers carrying out more human research (which as shown above has a lower incidence of duplications). We speculate that the publication pressures of academic institutions might create greater pressures to cut corners and duplicate images, while researchers working in corporations and healthcare institutions face fewer career-related incentives to publish. If firms were more likely to use their own research to develop new products, then market incentives for successful drug development might reduce the propensity to engage in improper research practices. Further research is needed on this topic, given inconsistent past results on whether for-profit organizations are more or less likely to commit scientific malpractice (17, 18).

Panel (c) of Figure 2 illustrates the proportion of papers published with duplication by institutions in the leading countries of Alzheimer's research, sorted by the number of publications from each country. The incidence of duplication in papers originating from China, Spain, and India is above the descriptive mean, while the opposite is true for papers originating from the U.S., Japan, U.K., Germany, and France. These findings echo previous work on duplications in western blot graphs (9). Regressions controlling for journal and year effects only return a significant and positive estimate for papers by authors in Chinese institutions; the coefficient is

DRAFT

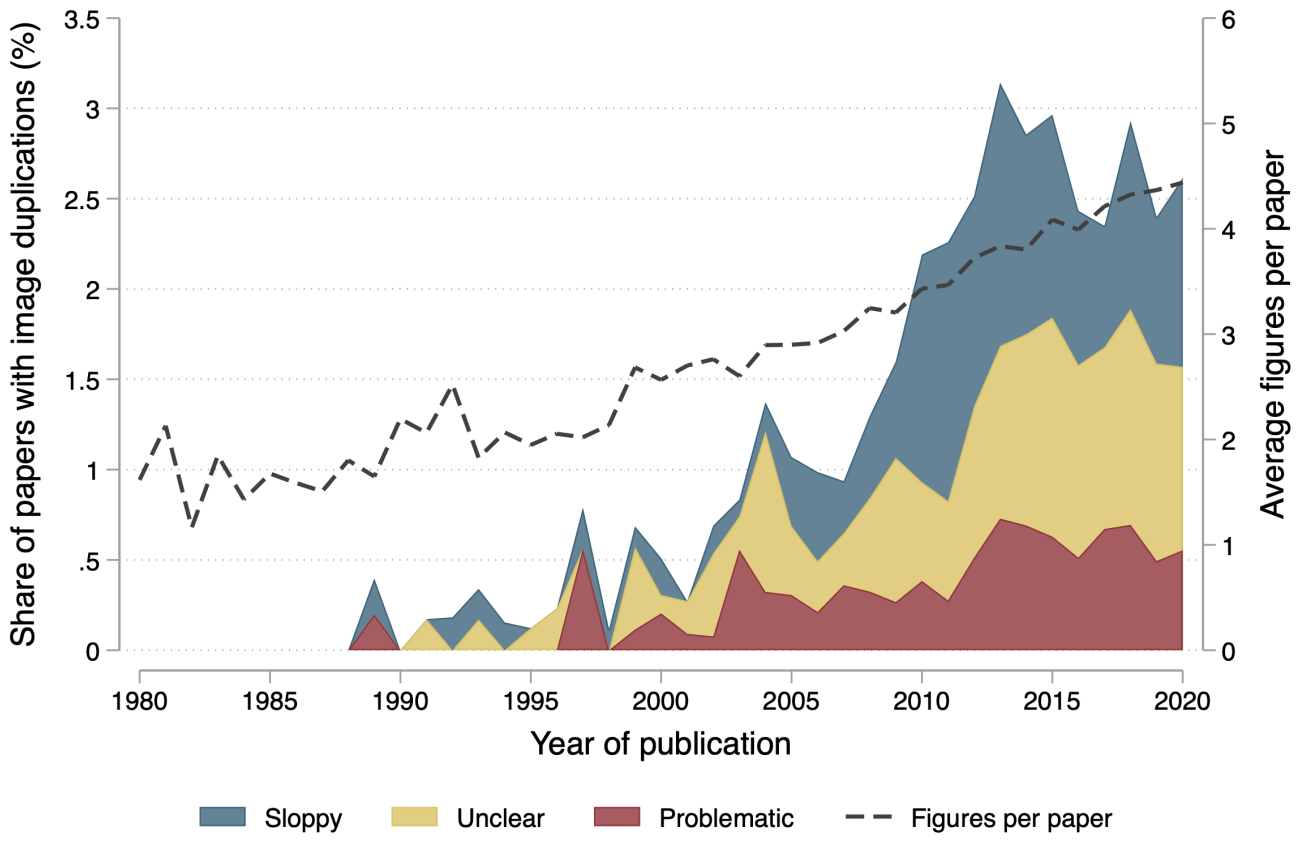


Fig. 1. Share of Alzheimer's research papers with image duplication issues, split out by manually determined categories of sloppy, unclear, and problematic duplication (left axis), and average number of images per publication (right axis), 1980-2020. Data include all papers in the study population of Alzheimer's papers (i.e., they also include papers without any figures).

negative or not statistically significant for all other countries (Table S10). Duplications in papers from China are also less likely to be sloppy duplications and more likely to be unclear or problematic (Figure S14). The share of papers originating from China that report duplicated images has been growing over time (Figure S15), suggesting that the increasing share of Alzheimer's research originating from China has been one of the drivers of the overall increase in duplications

We present additional results in the SI. More prestigious journals are more likely to present image duplications on average, however, the most prestigious journals (top 10% by impact factor) lack a significant association, implying that peer review might be more effective at the very top journals (though we cannot observe the underlying rate of duplicitous submissions to any journal) (Table S11). Papers published by researchers affiliated with higher-prestige institutions are also less likely to present duplications (Table S12), even though high-profile cases in the news tend to highlight elite institutions. Papers with more co-authors and/or more institutional affiliations are less likely to present image duplication, implying benefits from a greater number of eyes on or collective responsibility for a paper (Table S13). The presence of an author with a past history of sloppy, unclear, or problematic duplications increases the chances of the newer article presenting a duplication by 83%, 121%, and 124%, respectively (Table S18).

Conservative back-of-the-envelope calculations imply that the NIH provided $79.6 million of federal funding for Alzheimer's articles that presented a duplication. This figure ignores, however, how bad science might weaken the foundation for follow-on research, skew funding in the wrong direction, and delay the discovery of clinical solutions. In the case of the field of Alzheimer's, a similar back-of-the-envelope estimation for the larger population of papers that cite duplicitous image papers adds up to $2.85 billion of NIH funding (see SI for details on both calculations). Furthermore, while clinical researchers appear less likely to build upon bad images, scientists and inventors actually increase their attention to papers with such images (Table 1); duplicitous image papers receive, on average, 8 more citations relative to the sample average of 70.4. The increase in scientific citations is not mechanically driven by self-citations (Table S14) or citations with a negative tone (Table S15), suggesting that scientists are not systematically able to recognize or avoid duplicitous papers (19). Such papers are also 11.6% more likely to be cited by patents (Table S16), translating into an average increase of 0.97 patent citations (a 45.9% increase over the sample mean). In contrast, articles presenting a duplication are 7.7% less likely to be cited by clinical

publications (Table S17). It appears that clinical researchers are better able to avoid papers with duplicitous images, even if those papers initially received increased attention and resources from both academics and firms.

Discussion

Improper manipulation and duplication of images is just one type of scientific malpractice, and as such, our approach provides only a partial picture of scientific misconduct. Yet, this type of malpractice has a large impact in fields like Alzheimer's research, where images of experimental conditions and scans of brain activity constitute one of the fundamental ways in which data are presented and results are documented in academic papers. The presence of duplicated images might also be a harbinger of deeper flaws in a paper. Our results illustrate how articles presenting duplications can distort attention, time, and financial resources away from better scientific foundations for clinical applications.

While these results cover just one field over four decades, the systematic nature of our evidence can support tentative policy suggestions. First, given the increase in image duplication, the recent adoption of image-checking software by science journals appears warranted and potentially effective at decreasing one type of data malpractice (20). However, careful manual validation remains exceptionally important to rule out frequent false positives, especially with careers and reputations at stake. Second, if robust and automated metrics of sloppiness or fraud can be established, then institutions, research sponsors, and journals can take those metrics into account in their promotion, funding, and review processes. Third, it might be useful to further characterize and standardize categories of image duplication. Coupled with recent advances in AI, this approach could build upon earlier trailblazing work (9).

Future research is obviously needed on image duplications and their relationships with other types of fraud, why they are increasing, their prevalence in different fields, the effectiveness of efforts to stop it, and which research contexts are more likely to witness fraud. The research agenda should be ambitious and multidisciplinary with the objective of developing a better and causal understanding of the psychological, sociological, and economic drivers of scientific misconduct. On the methodological front, future work should focus on the development of extra-manual methods for scaling the detection of fraud. Besides elucidating the implications for funding, health, and economic outcomes, research on real-time detection of duplicitous images would increase the integrity of the scientific enterprise and public faith in the value of science.

ACKNOWLEDGMENTS. We are grateful to Imagetwin for granting us access to their software for research purposes. Any opinions expressed herein are those of the authors and do not represent the views of Imagetwin. We much appreciate the feedback from Candace Fleming, Joshua Graff Zivin, Dan Gross, Kris Gulati, Joshua Krieger, Saqib Mumtaz, Quentin Nicolas, Felix Poege, Raphael Scharfmann, Olav Sorenson, and John P. Walsh. All errors remain with the authors.

References.

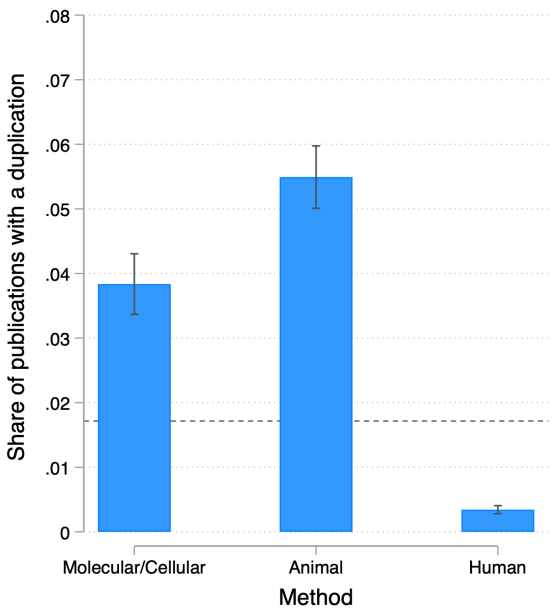
1. CG Begley, LM Ellis, Raise standards for preclinical cancer research. *Nature* **483**, 531–533 (2012).
2. M Herzenstein, S Rosario, S Oblander, O Netzer, The language of (non) replicable social science. *Psychol. Sci.* **35**, 1048–1061 (2024).
3. W Youyou, Y Yang, B Uzzi, A discipline-wide investigation of the replicability of psychology papers over the past two decades. *Proc. Natl. Acad. Sci.* **120**, e2208863120 (2023).
4. C Piller, Blots on a field? *Science* **377** (2022).
5. C Piller, Picture imperfect. *Science* **385** (2024).
6. RG Steen, A Casadevall, FC Fang, Why has the number of scientific retractions increased? *PloS one* **8**, e68397 (2013).
7. R Van Noorden, More than 10,000 research papers were retracted in 2023—a new record. *Nature* **624**, 479–481 (2023).
8. D Fanelli, How many scientists fabricate and falsify research? a systematic review and meta-analysis of survey data. *PloS one* **4**, e5738 (2009).
9. EM Bik, A Casadevall, FC Fang, The prevalence of inappropriate image duplication in biomedical research publications. *MBio* **7**, 10–1128 (2016).
10. M Marx, A Fuegi, Reliance on science by inventors: Hybrid extraction of in-text patent-to-article citations. *J. Econ. & Manag. Strateg.* **31**, 369–392 (2022).
11. J Priem, H Piwowar, R Orr, Openalex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *arXiv preprint arXiv:2205.01833* (2022).
12. P Azoulay, JL Furman, Krieger, F Murray, Retractions. *Rev. Econ. Stat.* **97**, 1118–1136 (2015).
13. G Gopalakrishna, et al., Prevalence of questionable research practices, research misconduct and their potential explanatory factors: A survey among academic researchers in the netherlands. *PloS one* **17**, e0263023 (2022).
14. GM Weber, Identifying translational science within the triangle of biomedicine. *J. Transl. Medicine* **11**, 1–10 (2013).
15. D Li, P Azoulay, BN Sampat, The applied value of public investments in biomedical research. *Science* **356**, 78–81 (2017).
16. LP Freedman, IM Cockburn, TS Simcoe, The economics of reproducibility in preclinical research. *PLoS biology* **13**, e1002165 (2015).
17. E Morreim, Corporations, high-stakes biomedical research, and research misconduct: Yes they can (and sometimes do). *J. Law Biosci.* **8**, lsab014 (2021).
18. P Balasubramanian, L Pierce, T Cummings, Research validity across organizational forms: Evidence from phase 2 oncology clinical trials. *Aalborg Univ. Johns Hopkins Univ. Wash. Univ. St. Louis* (2022).
19. C Catalini, N Lacetera, A Oettl, The incidence and role of negative citations in science. *Proc. Natl. Acad. Sci.* **112**, 13823–13826 (2015).
20. HH Thorp, Genuine images in 2024. *Science* **383** (2024).

Table 1. Papers with improperly duplicated images receive on average a higher number of citations from science papers and patents, but a lower number of citations from clinical articles.

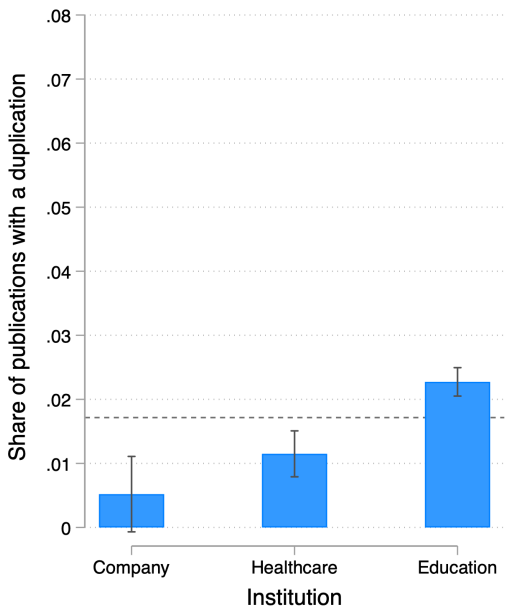
Dependent Variable	Scientific citations (total count)	Patent citations (total count)	Clinical citations (total count)
Model	(1)	(2)	(3)
Paper with duplication (0/1)	8.068**	0.967*	-0.140*
	(2.12)	(1.89)	(2.08)
Constant	70.30***	2.089***	1.831***
	(1090.00)	(241.45)	(1420.29)
Observations	48466	48466	48466

All models report ordinary least squares regressions (Poisson count models return similar substantive results).
Standard-errors clustered by journal. *t* statistics reported in parentheses.
Each column includes year fixed effects and journal fixed effects.
Significance codes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
.

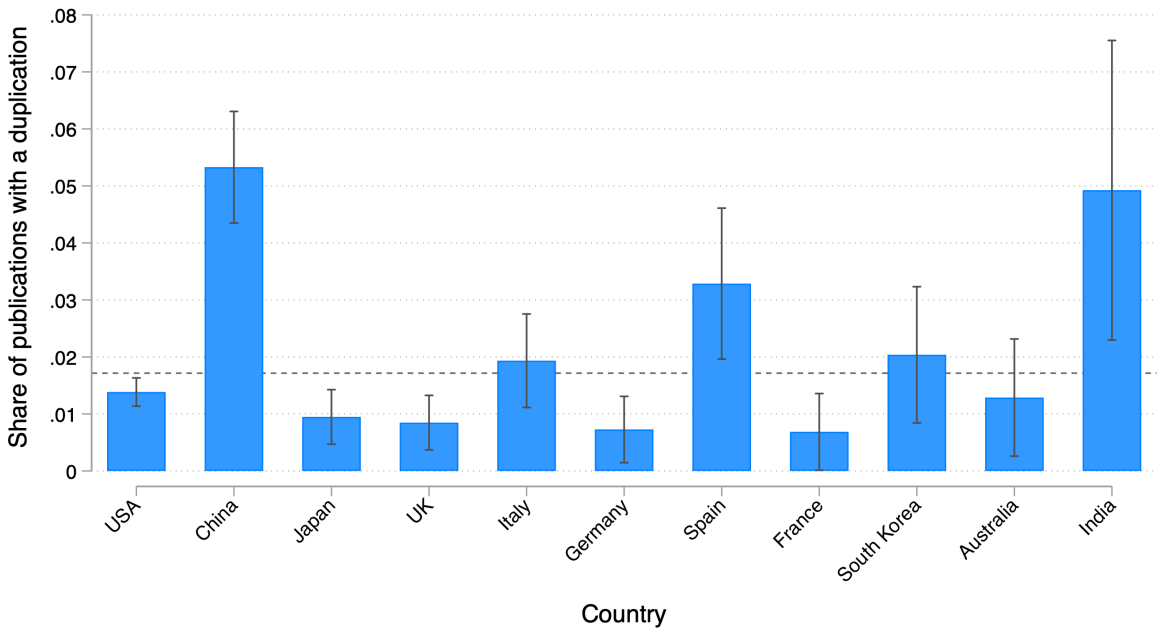
DRAFT



(a) *Incidence of duplications by research methods*



(b) *Incidence of duplications by institutional affiliation*



(c) *Incidence of duplications by country*

Fig. 2. Incidence of image duplications of Alzheimer's research as a function of methods, institutional affiliation, and country of origin, 1980–2020. Panel (a): Proportion of papers categorized by research method that contain at least one image duplication. Panel (b): Proportion of papers all of whose authors are affiliated with the listed type of institution that contain at least one image duplication. Panel (c): Proportion of papers with authors exclusively from the listed country that contain at least one image duplication. Countries are ordered by their total number of Alzheimer's articles. Dotted lines illustrate the whole sample average for comparison. Please see corroborating regression models in SI.

Duplicitous images in Alzheimer's research: systematic evidence, trends, and impacts

Supplementary Information

M. Tranchero,[1*] C. Fons-Rosen,[2] E. Scharfmann,[3] L. Fleming[3,4]
[1]The Wharton School, University of Pennsylvania
[2]Department of Economics, UC Merced
[3]Haas School of Business, UC Berkeley
[4]College of Engineering, UC Berkeley

[*]To whom correspondence should be addressed: mtranc@wharton.upenn.edu.

A. Materials and Methods

Detailed information on materials and methods is provided in this supplement and organized as follows:

1) Data sampling and assemblage
2) Description of Imagetwin software
3) Examples of image duplications detected by Imagetwin
4) Results and validation of duplication data
5) Mention of duplicated images in article text
6) Detailed protocol for classifying duplication types
7) Examples of types of duplication classifications
8) Validation of classification of duplication types
9) Classifying the research methods of articles
10) Descriptive statistics

1) Data sampling and assemblage. We began by using data from Pubtator3 (https://www.ncbi.nlm.nih.gov/research/pubtator3/) to sample from all papers indexed in PubMed up to December 2023 inclusive. In particular, we downloaded the publicly-available FTP file "disease2pubtator3.gz" which provides all the MeSH-assigned keywords for each publication. We then restricted the sample to publications related to Alzheimer's disease by retaining only those tagged with the identifier "D000544 - Alzheimer Disease." This created a sample of 116,476 articles indexed in PubMed. We then merged these articles with bibliometric information from ICite NIH (https://icite.od.nih.gov/). This step left us with 116,034 articles indexed in both PubMed and ICite. Restricting to English-only articles left 82,792 articles and further restricting the focus to journals that frequently host Alzheimer's research, i.e., that during the 41 years considered have published at least 50 papers on the disease, left 62,661 articles.

We focus our analysis on articles published between the years 1980-2020 to avoid truncation bias at the end of the time period, i.e., leaving a window too short to observe the downstream impact of research. The resulting 52,240 articles are merged with the OpenAlex database (https://openalex.org/) via each paper's DOI, which dropped a few articles without the DOI in the ICite data and/or those not present in OpenAlex. The analyses rely upon the OpenAlex database for information about the affiliation of authors (geography and institution type) as well as research impact (count of citations from papers). In sum, we are left with a final list of 50,604 publications capturing the relevant corpus of peer-reviewed Alzheimer's research published in English journals from 1980 to 2020 inclusive.

We then assembled PDFs for each of the Alzheimer's papers in our sample. We obtained them from a variety of sources, including open-access repositories such as PubMed Central (PMC) and online archives (e.g., medRxiv). The process also required thousands of manual downloads that were partially limited by the coverage of our institutions' libraries. Relative to the articles we could download, those missing are comparable in year of publication (average year of publication=2008.7 compared to a sample average of 2008.2). However, they tend to be published in less prestigious journals (average impact factor of journal=3.96 compared to a sample average of 4.51), which may partially explain the incomplete coverage of our institution's libraries. Overall, the result of our efforts is a dataset of 48,506 PDFs (or 95.9%), which, to the best of our knowledge, constitutes the largest available database of full-text and image published work in Alzheimer's. The PDFs were only used to scan images, without any violation of their copyrights.

We added data from a number of additional sources. We collect information about papers retracted up to 2023 inclusive from Retraction Watch as of April 2024 (https://retractionwatch.com/). The Retraction Watch data helpfully include a classification for the main reason that prompted the retraction of each article. We also obtain information from PubPeer (https://pubpeer.com/). PubPeer is the most prominent blog where researchers discuss and review scientific research after publication, i.e., what is known as post-publication peer review. The data that were graciously shared with us by the blog administrators contain the DOI of all Alzheimer's articles discussed on the PubPeer blog as of April 2024. Data on the author's affiliation and geographic location are from OpenAlex and available for 30,263 Alzheimer's publications. Patent-to-paper citations come from the Reliance on Science database (https://zenodo.org/records/7996195).

2) Description of Imagetwin software. We partnered with Imagetwin (https://imagetwin.ai/), a startup that has developed a widely used artificial intelligence tool to assess the integrity of scientific images (5). Imagetwin scans each paper's images looking for repetitions or alterations of the same image within the same paper (or internal duplications), plagiarism of the same image in a different paper (or external duplications), and manipulation in western blots (or splicing). The type of duplications that Imagetwin locates widely vary in their type. Sometimes, duplications include instances where the same image, or a piece of it, appears more than once within the same paper. Other times, more troubling instances emerge when an image has been obviously cropped, rotated, colored, or altered. However, it must be noted that the software picks up all types of duplications, including appropriate ones such as zoomed portions of graphs, conditions exposed to different dies, or images reproduced with permission. Therefore, whether a repeated image truly constitutes a case of inappropriate duplication or outright fraud requires additional and careful human evaluation.

We scanned 48,506 PDFs with Imagetwin's API between March 2024 to June 2024. Of those, 24,250 were flagged as presenting possible duplications. After ruling out obvious false positives (e.g. a duplication of the journal logo on the first page), we manually checked the 9,655 papers that plausibly presented a duplication. Each paper was assessed independently by two of this paper's authors and only cases independently confirmed by both of them were kept. Then, each flagged image was manually inspected in the original PDF, where information in the text and the figure's footnote were used to rule out instances where the duplication was intentional and correctly

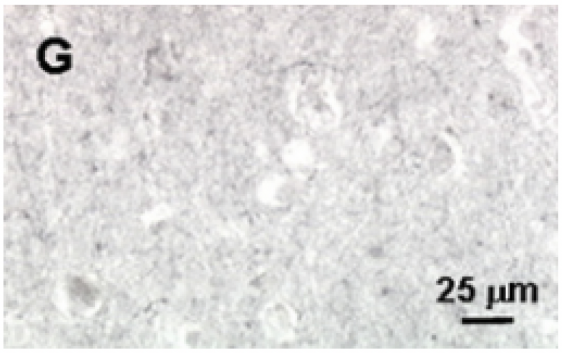
DRAFT

documented. Our approach partially automates previously painstaking efforts in which images were carefully scanned by eye (9), however our approach does not fully rely on machine learning which is known to have several shortcomings, e.g., false positives.*

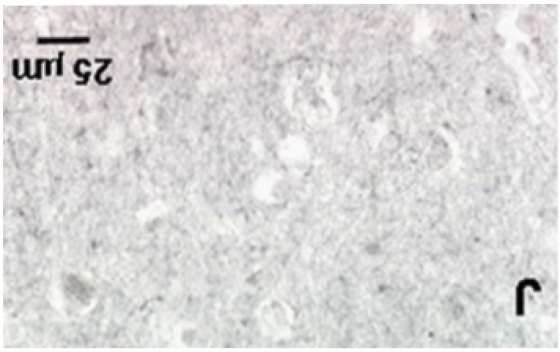
3) Examples of image duplications detected by Imagetwin. To begin we describe two examples of duplications detected by the Imagetwin software. Note that in what follows, we list the corresponding retraction notice every time that we need to refer to a retracted article in order to avoid any citation occurring to them.

Fig. S1. Example of internal (panels a and b) and external (panels c and d) image duplications detected by Imagetwin.

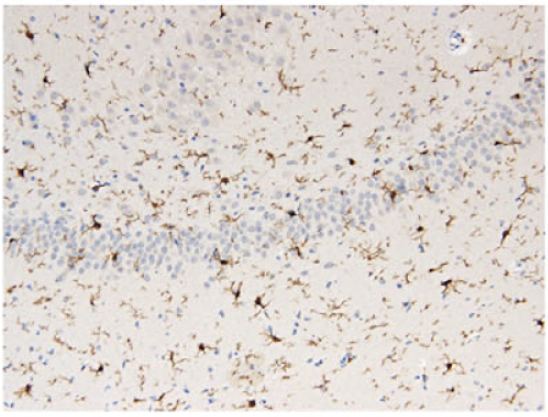
(a) *Panel G in Fig. 7 from Goni et al. (2010)*



(b) *Panel F in Fig. 7 from Goni et al. (2010)*



(c) *Panel A(1) in Fig. 3 from Zhou et al. (2021)*



(d) *Panel E(4) in Fig. 3 from Zhang et al. (2021)*



The first type of duplication is what we label as "internal", i.e., repetitions or alterations of the same image within the same paper. This type of duplication has been previously studied by Elisabeth Bik, Arturo Casadevall, and Ferric Fang in their pioneering work (9), since it can be identified with manual scans by eye. Panels (a) and (b) of Figure S1 shows an example in our data for an article that has not been retracted. As the figure makes clear, the same image is reported in panel G and panel J of Figure 7 of the original paper, with the latter rotated by 180 degrees (panel F is shown rotated for ease of comparison). Indeed, this duplication has not gone unnoticed; an editorial note reported that the figure could not be rectified by the authors because *"none of the raw data underlying the article's results are currently available."* † However, the article has not been retracted because *"the article's overall results and conclusions remain supported by other results reported in the article."*

The second type of duplication is what we label as "external", i.e., plagiarism of the same image from a different paper. This type of duplication is usually much harder to detect manually and has not been studied at scale previously.‡ Panels (c) and (d) of Figure S1 shows two examples from our data. The two papers, from different groups of authors, present identical panels. The Imagetwin software is able to recognize these duplications despite being at different levels of resolution in the original PDFs. The duplication was eventually discovered, alongside several others between the two articles. Both papers were eventually retracted because *"several image elements of the experimental data were published elsewhere in a different scientific context."* §

4) Results and validation of duplication data. Our procedure yields 832 articles presenting inappropriate duplications. Of those, 671 papers present one duplicated image, 127 present two, 25 present three, and 9 present four or more. Our findings imply that 1.7% of Alzheimer's publications report duplicated images, a number growing to 2.1% when excluding articles without figures. Due to our conservative

*More recently, ImageTwin has launched a feature that automatically sorts out irrelevant duplicates (https://imagetwin.ai/posts/confidence). This positive update is outside our period of use but suggests that AI-powered tools to recognize duplicated images will gradually increase precision and thus reduce the need for manual checks.

†The editorial note for this article is available here: https://journals.plos.org/plosone/journal.pone.0297928.

‡Imagetwin is able to find external duplications by comparing each paper against their large proprietary corpus of images. In a few cases, the papers in our dataset were identified as "victims" of duplications from papers not in our data (e.g., on different diseases or outside our sample period). Those cases were excluded and our data report only situations where the duplicator was a paper within our database.

§The retraction notices from the publishers are available online: https://onlinelibrary.wiley.com/doi/10.1002/jcb.30059 and https://onlinelibrary.wiley.com/doi/10.1002/jcb.30062, for the two papers, respectively.

categorization processes (details below), this figure represents a likely conservative and lower-bound estimate of the incidence of this type of scientific malpractice. Yet, the number of papers we find as presenting evidence of malpractice are still 4.4 and 9 times higher than PubPeer and Retractions Watch as of April 2024, respectively (Figure S2).

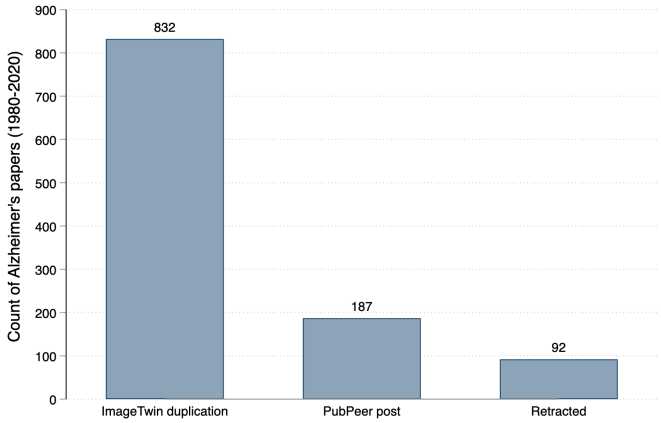


Fig. S2. Comparison of Imagetwin results identified in this paper with the number of papers flagged on the PubPeer blog and retractions from Retractions Watch, 1980-2020.

We validated the articles identified by Imagetwin by comparing them to the 92 papers in our sample that are listed as retracted on Retraction Watch as of April 2024. Since the Retraction Watch data characterize the reasons associated with each retraction, we can explore if the Imagetwin plus manual process is able to isolate the same papers that were caught presenting problems in their images. Regressing our indicator of any image duplication on the Retraction Watch indicators of retraction, retraction due to image fraud, and retraction due to image duplication, returned monotonically increasing coefficients of 0.267, 0.329, and 0.897, respectively (Table S1). Since these coefficients derive from a linear probability model (LPM), they can be interpreted as a percentage increase in the likelihood of presenting a duplication. Column 1 shows that retracted papers are more likely to present duplicated images, which the estimate quantifies as an increase of 15.5 times over the sample mean of 0.01715 (or 1.72%). Increasing our faith in the protocol, the effect grows as the reason for retraction becomes more related to issues with images (column 2) and, more specifically, to image duplication (column 3). Reassuringly, the subset of papers retracted because of image duplications correlates most highly with our dependent variable of duplicated images. Taken together, this analysis supports a combined approach of Imagetwin's software and manual validation, and illustrates that it has the ability to detect issues in images in a sample of papers where their incidence is known to be higher ex ante.

5) Mention of duplicated images in article text. Duplicitous images may have very different implications depending on whether they represent the results that are most central to a given paper's findings. However, it is usually hard to assess the relative importance of each image. To make progress on this issue, we build on the intuition that more important images should be discussed more extensively in the article's text. We employ natural language processing techniques to parse the full text of the articles in our sample and count the number of mentions of each figure. In total, we are able to collect this information for 36,806 of the articles in our sample. On average, each figure receives 3.14 mentions (s.d. 3.019), with a minimum of 1 to a maximum of 78 mentions in the text.

Figure S3 shows the likelihood that any image of a paper presents a duplication depending on how many times the figure is mentioned in the article's text. The pattern is striking: figures with more mentions in the text are associated with a higher likelihood of presenting duplications. Table S2 employs OLS to further explore the relationship between duplicated images and the number of text mentions they receive. The results confirm that figures with duplications tend to be discussed more within an article, and are thus more likely to be central to it. The result holds even when restricting the comparison to the other figures within the same article (column 3 details a model

Table S1. Articles presenting a duplication according to Imagetwin software and additional manual validation are more likely to have been retracted, especially due to duplicated images.

Dependent Variable	Paper with duplication (0/1)		
Model	(1)	(2)	(3)
Retracted for any reason (0/1)	0.267***		
	(4.25)		
Retracted for image fraud (0/1)		0.329**	
		(2.25)	
Retracted for duplicated image (0/1)			0.897***
			(17.40)
Constant	0.0166***	0.0169***	0.0167***
	(246.50)	(698.53)	(1428.34)
Observations	48466	48466	48466

Standard-errors clustered by journal. *t* statistics reported in parentheses.
Each column includes year fixed effects and journal fixed effects.
Significance codes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

NBER DRAFT

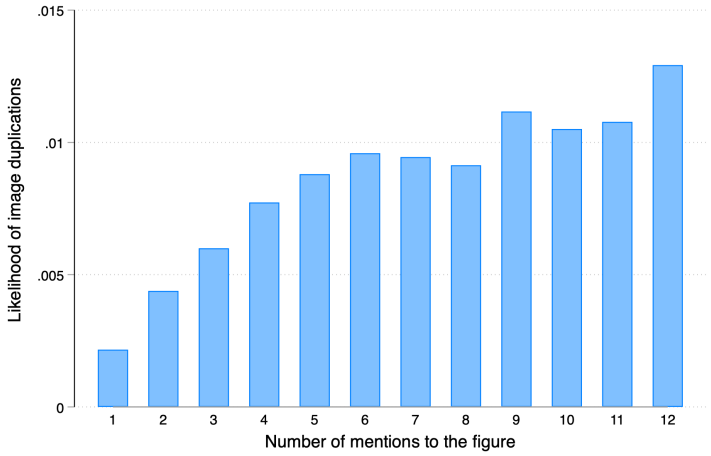


Fig. S3. Figures with more mentions in the articles' text are more likely to present an inappropriate duplication (the average number of mentions for an image is 3.14).

with paper and figure order fixed effects). Taken together, this evidence suggests that duplicated images tend to be those more important to an article's thesis and findings.

Table S2. Duplicated figures receive more mentions in the article text relative to the other figures.

Dependent Variables:	Number of text mentions		
Model	(1)	(2)	(3)
Figure with duplication (0/1)	1.508***	0.701***	0.499***
	(11.39)	(6.76)	(5.06)
Constant	3.128***	3.176***	3.177***
	(239.76)	(5672.66)	(5961.03)
Observations	164770	158932	158932

Standard-errors clustered by journal. *t* statistics reported in parentheses.
Column 1 does not include controls; Column 2 includes article fixed effect;
Column 3 includes article and figure order fixed effects.
Significance codes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

DRAFT

6) Detailed protocol for classifying duplication types. Our conservative approach labels image duplications as inappropriate only if two of this paper's authors confirm the result of the Imagetwin-flagged instance of duplication. However, even among inappropriate duplications, large differences exist in the type of behavior and its correlates and consequences. For instance, previous studies focusing on western blots found that around a third of duplications are simple repetitions that might result from good faith mistakes or sloppiness (9). Other times, manipulations seem more likely the result of intentional fraud, such as when images present unnatural patterns in the background or plagiarized images that present evidence of alteration (5). Distinguishing between these types of duplications is crucial to interpreting the evidence and inferring the likely intent that led to them. Building on previous research (9, 12), we developed a protocol to manually classify the type and severity of image duplication. We established two different processes depending on whether the duplication was within the same paper ("internal") or in two different papers ("external").

Protocol for internal inappropriate duplications

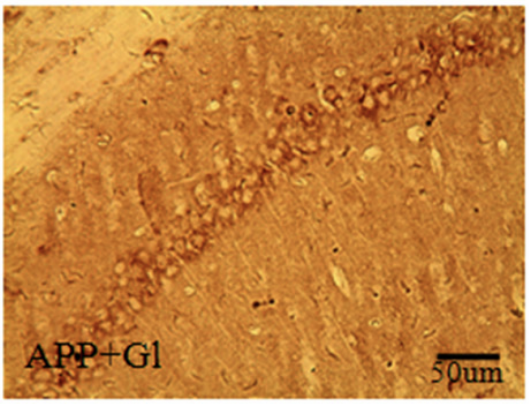
1. Establish the type of duplication:
   (a) Simple duplication for the same condition: if the same image (complete or in part) is used to represent the same condition but in another experiment, without apparent manipulations.
   (b) Simple duplication for different conditions: if the same image (complete or in part) is used to represent a different condition in the same or another experiment, without apparent manipulations.
   (c) Western blot splicing: if a western blot presents clear evidence of vertical splicing that is not documented or noted in the footnote or text.[¶]
   (d) Duplication with repositioning and/or alteration: if the same image (complete or in part) is repositioned, rotated, mirrored, shifted, stretched, overlapped, cropped, modified, or if colors are altered,[‖] labels changed, or when the image presents unnatural repetition patterns in the background or other explicit manipulation in how the image is assembled.

[¶] We followed the methodology of (9) and do not consider as inappropriate instances of potential beautification, such as cases where part of the background of a blot shows signs of patching to remove a smudge or stain. Instead, we focused on vertical splicing across lanes of the western blot that are commonly considered inaccurate data representation (e.g., see https://www.frontiersin.org/news/2021/western-blot-image-cropping).

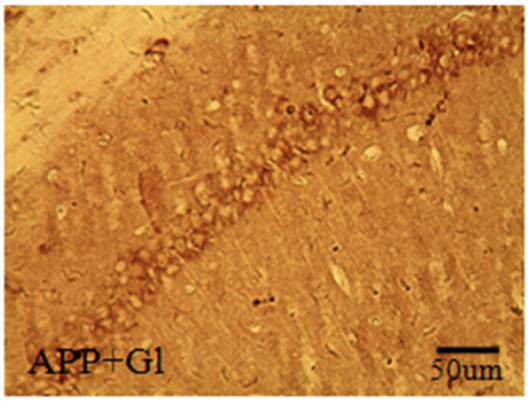
[‖] Note that when the duplication was due to different stains of the same picture done on purpose, we did not consider it as an inappropriate duplication at all. We only flagged duplicated images where the color appeared to be altered to misrepresent the image as the result of a different experiment.

Fig. S4. Example of internal duplication manually classified as "sloppy" according to our protocol.

(a) *Panel 4 in Fig. 4 from Shi et al. (2016)*



(b) *Panel 4 in Fig. 5 from Shi et al. (2016)*



2. Assign the type of likely behavior evinced by the duplication:
   (a) sloppy: if the duplication is a simple duplication for the same condition.
   (b) unclear: if the duplication is a simple duplication for different conditions or the image presents splicing, and residual category for ambiguous cases.
   (c) problematic: if the duplication is with repositioning and/or alteration or the image presents splicing with duplications.

Protocol for external inappropriate duplications

1. Establish if the focal paper is the duplicator or the victim of the duplication (assuming that the first published paper is the victim).
2. Establish if the duplicator paper at least cites the paper from which it copies, even if the image is not properly attributed.
3. Establish the type of duplication:
   (a) Poor acknowledgment: if the image is clearly taken from another paper for illustrative purposes, but it fails to explicitly acknowledge the source of the image in the footnote or text.
   (b) External duplication without alteration: if the image (complete or in part) is taken from another paper and used to represent a new experiment.
   (c) External duplication with repositioning and/or alteration: if the image (complete or in part) is taken from another paper and repositioned, rotated, mirrored, shifted, stretched, overlapped, cropped, modified, or if colors are altered, labels changed, or when the image presents unnatural repetition patterns in the background or other explicit manipulation in how the image is assembled.
4. Assign the type of likely behavior evinced by the duplication:
   (a) sloppy: if the duplication is poorly acknowledged.
   (b) unclear: if the external duplication is without alterations and cites the original paper,** and residual category for ambiguous cases.
   (c) problematic: if the external duplication is without alterations and does not cite the original paper, or if the external duplication is with repositioning and/or alteration.

7) Examples of duplication classification. We present a few examples of cases from our data and the corresponding classification according to our protocol. Examples of duplications are mostly chosen from papers that have been retracted or flagged as presenting issues on the PubPeer forum.

Simple duplication for the same condition Figure S4 shows two panels from Figure 4 and Figure 5 of a retracted paper.†† Both figures are composed of six panels each and present instances of repeated identical images. This instance of duplication found with Imagetwin software confirms the findings by the editors of the journal. The paper has been retracted and the editorial note reported several issues, including that *"Figure 4: the bottom left panel appears to be identical with the bottom left panel of Fig. 5."* Since the image does not present signs of alterations and is used to represent the same experimental condition, our protocol gives the authors the benefit of the doubt and classifies this case of potential sloppiness. However, this illustrates how conservative our approach is, since this type of duplication was enough to impair the credibility of the entire manuscript and result in the retraction of the paper.

Simple duplication for different conditions Figure S5 shows a portion of Figure 6 from another retracted paper.‡‡ Two panels of the figure appear identical despite showing data about different experimental treatments (highlighted by Imagetwin in green in Figure S5). In particular, the top-left panel supposedly reports 5xFAD mice treated with SLCP for two days, while the middle-bottom panel reports 5xFAD mice treated with Cur for two days. However, the two figures are suspiciously similar. Due to this duplication and a few similar ones, the article was eventually retracted with the agreement of all the authors. Since the duplicated figure is used to represent different experimental conditions, our protocol classifies this case as due to unclear reasons.

** Given our conservative approach, we coded this instance as showing unclear intent because it may stem from good faith mistakes in failing to acknowledge the focal paper.

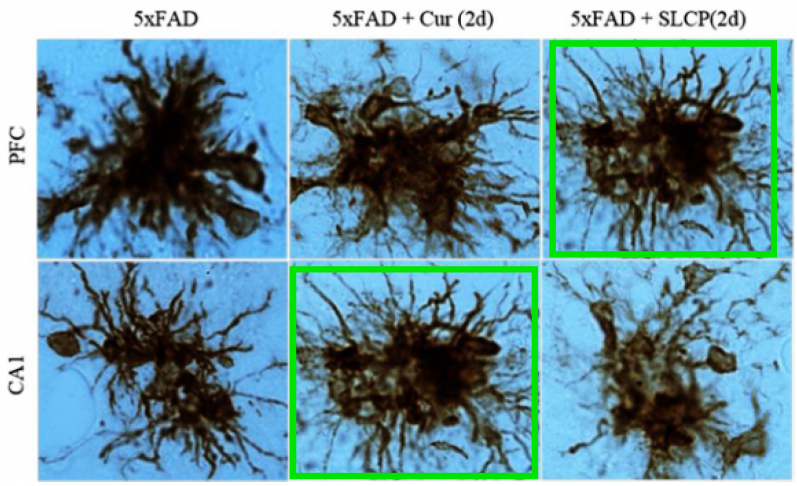
†† The retraction note is available here: https://bmccomplementmedtherapies.biomedcentral.com/articles/10.1186/s12906-021-03331-0.

‡‡ The retraction note is available here: https://bmcneurosci.biomedcentral.com/articles/10.1186/s12868-023-00785-5

DRAFT

Fig. S5. Example of internal duplication manually classified as "unclear" according to our protocol.

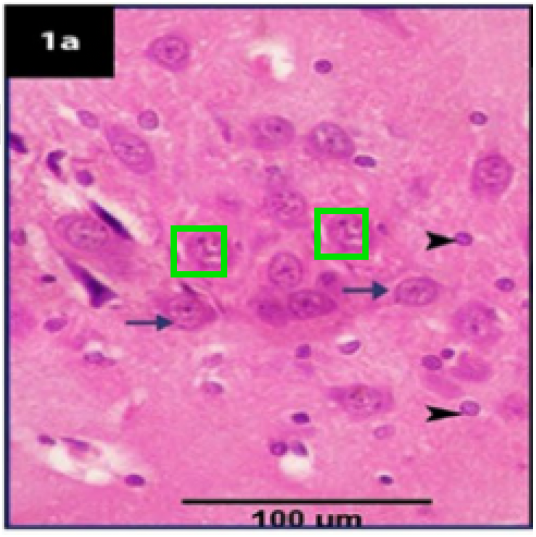
(a) *Portion of Fig. 6 from Maiti et al. (2018); duplicated panels are highlighted in green*



Internal duplication with alteration Figure S6 shows a portion of Figure 7 from a retracted article.[§§] The green square marked by Imagetwin in the figure highlights a worrying pattern: several areas of the figure show identical repetitive features, which are highly likely to be the result of manipulations. In this specific case, the article was retracted, and the accompanying notice is quite telling: *"After publication, concerns were raised regarding repetitive elements in the images presented in Figs. 6, 7, and 8. The authors have stated that the images were edited to remove artifacts. Due to the number and extent of image modifications, the Editor-in-Chief no longer has confidence in the presented data."* Owing to the presence of alteration of the background of the figure, which is highly unlikely to result from a good faith mistake, we classified this instance of duplication as problematic.

Fig. S6. Example of internal duplication manually classified as "problematic" according to our protocol.

(a) *Panel 1a of Fig. 7 from Rezk et al. (2022); duplicated areas are highlighted in green*



Splicing Figure S7 shows panel A of Figure 3 from a paper that has not been retracted but has received substantial scrutiny on the PubPeer blog.[¶¶] The figure presents a western blot graph, a common laboratory technique used to detect a specific protein in a blood or tissue sample. However, this figure shows potential evidence of inappropriate splicing in the rightmost lane of the photo (which we marked with a red arrow for illustrative purposes). While cropping in western blots is sometimes warranted, best practices require full disclosure and explanation of the rationale behind it. In this case, the text and figure footnote do not present any justification for the manipulation of the image. Furthermore, anonymous users on the PubPeer blog have noticed that the cropped lane is an exact duplication

[§§] The retraction note is available here: https://link.springer.com/article/10.1007/s11011-023-01282-8

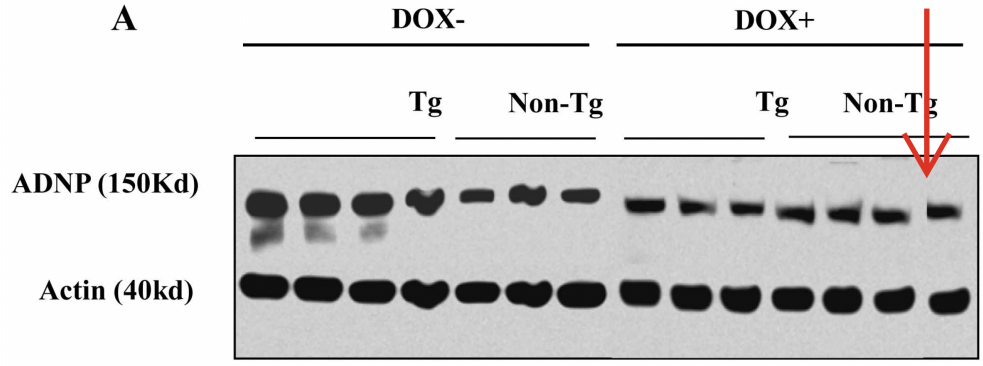
[¶¶] The PubPeer blog discussion with detailed evidence on the figure is publicly available here: https://pubpeer.com/publications/62CD938FC80C350418F2FC5CEF3481

DRAFT

of the neighboring one, further compounding doubts about the integrity of this data. According to our protocol, we classified this instance of splicing as problematic.

Fig. S7. Example of splicing manually classified as "problematic" according to our protocol. Vertical red arrow is added here and not from original figure.

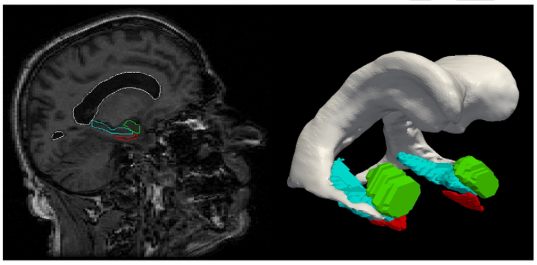
(a) *Panel A of Fig. 3 from Schirer et al. (2014); location of the splice is marked by a red arrow*



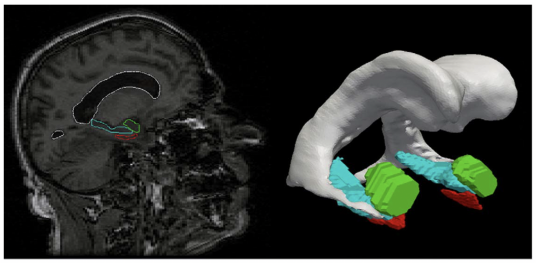
Poor acknowledgment of external duplication Panel (a) of Figure S8 shows Figure 1 from an article published in 2013,*** while panel (b) of the same figure shows Figure 2 from an article published in 2015.†††. The figures are clearly identical, despite the text or footnotes of the later article failing to acknowledge so. However, several reasons lead us to believe that this duplication is more likely to be a case of poor acknowledgment rather than more problematic malpractice. First, the image has clear illustrative purpose and intent, and as such does not purport to represent the result of a novel study. Second, the two articles share authors, and the second paper cites the first extensively. Both articles explicitly present different sets of results from the same larger collaborative study called BIOCARD. Therefore, our protocol leads us to classify this instance of external duplication as just due to sloppiness in acknowledgments. This example is helpful in showing how our protocol is designed to flexibly account for duplications that are unlikely to be due to bad faith behavior.

Fig. S8. Example of external duplication manually classified as "sloppy" according to our protocol.

(a) *Fig. 1 from Miller et al. (2013)*



(b) *Fig. 2 from Soldan et al. (2015)*



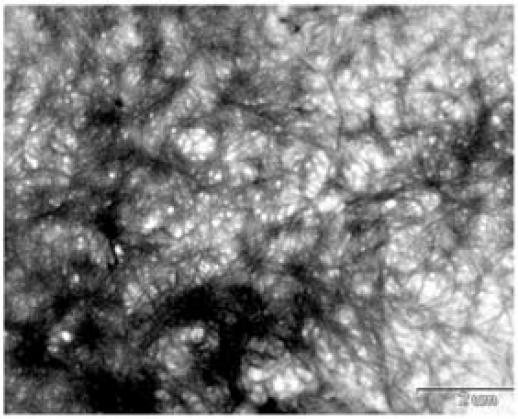
External duplication without alteration Panel (a) of Figure S9 shows Figure 2(C) from an article published in 2010‡‡‡, while panel (b) of the same figure shows Figure 1(D) from a subsequent publication.§§§ Also in this case, the subsequent article shares the same authorship of the duplicated one, as well as citing it. However, the article does not acknowledge the use of the same image. This case is more ambiguous because the same image is used to represent the control condition of experiments in different papers. In the first paper, the figure represents the A_{1-42} alone control for a treatment of $A_{1-42} + NQTrp$. In the second paper, the figure represents the A_{1-42} alone control for a treatment of $A_{1-42} + Cl - NQTrp$. Therefore, our protocol leads us to classify this case of duplication as unclear in nature.

External duplication with alteration Panel (a) of Figure S10 shows Figure 6(a) from an article published in 2018,¶¶¶ while panel (b) of the same figure shows Figure 1(b) from a paper published two years later.[17] This case presents the same image being cropped and reproduced in another manuscript, without any attribution to the original paper. Not only did the second article fail to report the source of the specific figure, it also does not cite the first article at all. The papers do not share any author. Such external duplications would appear less likely to be the result of honest mistakes in assembling the manuscript. This type of issue is frequent among papers that are eventually retracted, and those cases have often belied a fraudulent intent; indeed, the paper in question was eventually retracted (see prior footnote) because *"The results are flawed and the final conclusions could not be confirmed."*

*** The article is available here: https://www.sciencedirect.com/science/article/pii/S2213158213001186
††† The article is available here: https://onlinelibrary.wiley.com/doi/ftr/10.1002/hbm.22810
‡‡‡ The article is available here: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0011101
§§§ The article is available here: https://www.sciencedirect.com/science/article/abs/pii/S0969996112000769
¶¶¶ The article is available here: https://www.nature.com/articles/s41598-018-19732-0
[17] The retraction note is available here: https://www.degruyter.com/document/doi/10.1515/tnsci-2022-0262/html

Fig. S9. Example of external duplication manually classified as "unclear" according to our protocol.

(a) *Panel C in Fig. 2 from Scherzer-Attali et al. (2010)*



(b) *Panel D in Fig. 1 from Scherzer-Attali et al. (2012)*

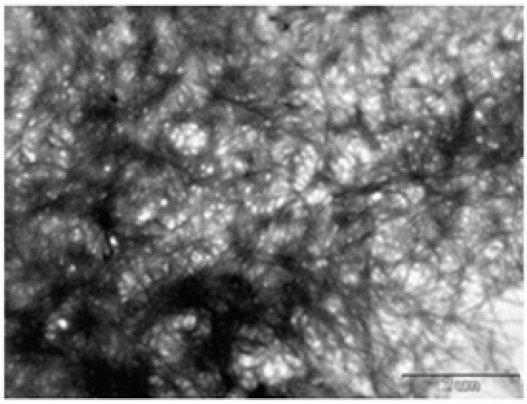
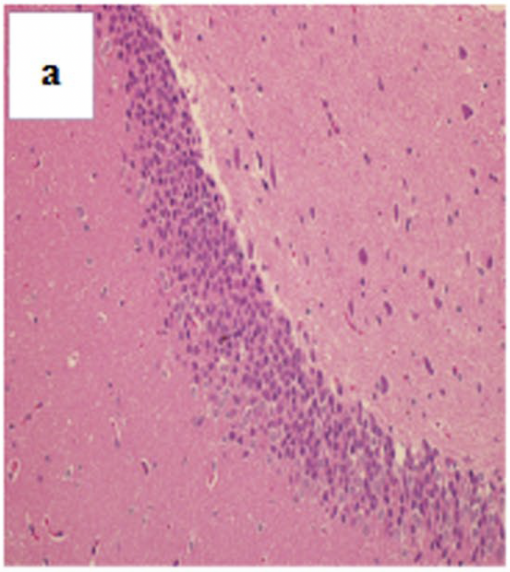
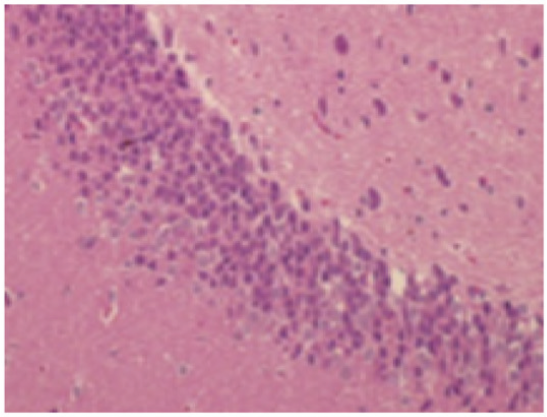


Fig. S10. Example of external duplication manually classified as "problematic" according to our protocol.

(a) *Panel a in Fig. 6 of Chen et al. (2018)*



(b) *Panel b in Fig. 1 of Cheng et al. (2020)*



8) Validation of classification of duplication types. We followed the protocols described above and classified 358 papers as having duplications likely stemming from sloppiness, 338 papers as presenting duplications of unclear intent, and 189 papers as presenting duplications with highly problematic features. Note that the total is larger than 832 because some papers present multiple duplications of different types. To simplify the interpretation of our analyses, we assigned each the cases to their most serious category, leaving us with 323, 320, and 189 papers in the sloppy, unclear, and problematic categories, respectively. It is worth reiterating that because it is difficult to infer the intent behind duplicated images, our classification follows a conservative approach. In particular, when we could not definitively categorize a duplicate as just a good faith result of sloppiness, we very carefully examined the image. Unless there was blatant evidence of problematic behavior, we defaulted on giving the benefit of the doubt and classified the image as unclear.

We provide evidence for the face validity of our classification in two ways. First, we would expect papers that have been retracted to present more serious image issues. We use OLS to regress an indicator for papers that have been retracted on our categorical dummies indicating sloppy, unclear, and problematic images (we get substantively the same results using non-linear Logit or Poisson models). Data on retractions come from Retraction Watch, a comprehensive crowd-sourced database that includes the motivation for a paper's retraction or notices, downloaded in April of 2024. Table S3 shows that estimates for papers presenting sloppy duplications are very close to zero and not statistically significant, confirming that duplications appearing as good faith mistakes are not likely to prompt retractions. Tellingly, the converse is true for duplications in the unclear and problematic categories. The association of problematic duplications with retracted papers is almost 47% larger than for the unclear category, providing face validity for the ordering of our protocol.

Second, we investigate the correlation between types of duplication and the number of images in a paper. The rationale of this validation is as follows. On average, we expect a positive correlation between the count of images and the presence of a duplication – as a limit case, no image duplication is possible in articles without images. However, the total count of duplications should be more positively

Table S3. Retracted papers are more likely to present unclear and problematic image duplications (data from Retraction Watch).

Dependent Variable	Paper retracted (0/1)		
Model	(1)	(2)	(3)
sloppy duplication	0.00657		
	(1.43)		
unclear duplication		0.0219***	
		(2.87)	
problematic duplication			0.0321**
			(2.07)
Constant	0.00102***	0.000922***	0.000954***
	(30.51)	(17.56)	(16.61)
Observations	48466	48466	48466

Standard-errors clustered by journal. *t* statistics reported in parentheses.
Each column includes year fixed effects and journal fixed effects.
Significance codes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

associated with mistakes in good faith, while fraudulent behavior is less likely to just be a mechanical reflection of the number of figures. Table S4 shows results confirming this intuition. We find that the correlation between sloppiness and the number of images is stronger than the correlation for problematic images. Overall, both of these analyses support the validity of our manually assessed category assignments.

Table S4. The number of figures in a paper correlates more strongly with sloppy than problematic duplications.

Dependent Variable	sloppy duplication (0/1)	unclear duplication (0/1)	problematic duplication (0/1)
Model	(1)	(2)	(3)
Number of figures	0.00294***	0.00264***	0.00130***
	(7.87)	(8.51)	(7.33)
Constant	-0.00250**	-0.00192*	-0.000701
	(-1.98)	(-1.83)	(-1.17)
Observations	47778	47778	47778

Standard-errors clustered by journal. *t* statistics reported in parentheses.
Each column includes year fixed effects and journal fixed effects.
Significance codes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

9) Classifying the research methods of articles. We use the Translation module of NIH iCite to characterize the research method of the articles. This helps characterize the measures and validate our result shown in Panel (a) of Figure 2 of the main paper (on human methods having a lower likelihood of duplication, compared to cell and molecular, and animal methods). In particular, Medical Subject Heading (MeSH) keywords assigned to each article permit us to know the type of model organism used in each paper's analyses (15). The following MeSH terms are employed:

- Molecular/cellular biology: A11 (Cells), B02 (Archaea), B03 (Bacteria), B04 (Viruses), G02 (Chemical Phenomena), G04 (Cell Physiological Phenomena)
- Animal: B01 (Eukaryota) but not B01.050.150.900.649.313.988.400.112.400.400 (Human)
- Human: M01 (Persons) and B01.050.150.900.649.313.988.400.112.400.400 (Human)

iCite computes for each article the fraction of MeSH terms that are in the molecular/cellular, animal, and human categories, conditional on the sum of the article's MeSH terms in those categories. Additional details are available at: https://icite.od.nih.gov/user_guide_trans. We assign each article to its most prevalent method based on the score of these categories: 6,439 articles (or 13.27%) are classified as employing molecular/cellular methods, 8,558 articles (or 17.64%) are classified as employing animal methods, and 33,509 articles (or 69.08%) are classified as employing human methods.

We compare this article-level classification with the Approximate Potential to Translate (APT) score compiled and made available by NIH iCite. The APT score is computed with a machine learning system that estimates the likelihood that knowledge from a paper will be employed and cited in subsequent clinical articles. APT is a probability that ranges from 0.05 (no detectable predictors of translation) to 0.95 (strong predictors of translation). Table S5 uses OLS to compare the APT score of articles that we classified as dominated by molecular/cellular, animal, and human methods. The coefficients show a monotonically growing likelihood of clinical translation for findings based on organisms of growing complexity. This supports the intuition that simpler organisms are used in more basic research aimed at elucidating fundamental biology, while evidence based on humans is closer to applied and therapeutic uses (15).

10) Descriptive statistics. Table S6 details descriptive statistics for the population of Alzheimer's papers separately for articles with and without duplicated images. The table reports the mean of each variable, the standard errors in parenthesis, and the p-value of a t-test between the two groups of articles.

Figure S11 provides histograms of the number of figures (panel a) and tables (panel b) per paper. Figure S12 provides a histogram of the number of Alzheimer's papers all of whose authors are affiliated with institutions from the given country. The countries are sorted by the total number of articles.

DRAFT

Table S5. Approximate Potential to Translate (APT) for articles employing different methods (data from NIH iCite).

Dependent Variable	Approximate Potential to Translate (APT)		
Model	(1)	(2)	(3)
Molecular/cellular methods (0/1)	-0.146***		
	(-13.80)		
Animal methods (0/1)		-0.115***	
		(-6.83)	
Human methods (0/1)			0.209***
			(19.36)
Constant	0.507***	0.508***	0.343***
	(362.54)	(171.96)	(45.99)
Observations	48466	48466	48466

Standard-errors clustered by journal. *t* statistics reported in parentheses.
Each column includes year fixed effects and journal fixed effects.
Significance codes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table S6. Descriptive statistics

Sample of articles	Without duplication N=47,674	With duplication N=832	Difference (p-value)
Molecular/cellular methods (0/1)	0.13 (0.34)	0.30 (0.46)	<0.001
Animal methods (0/1)	0.17 (0.38)	0.56 (0.50)	<0.001
Human methods (0/1)	0.70 (0.46)	0.14 (0.35)	<0.001
Researchers in firms (0/1)	0.01 (0.11)	0.00 (0.06)	0.026
Researchers in health instit. (0/1)	0.07 (0.26)	0.05 (0.21)	0.008
Researchers in universities (0/1)	0.36 (0.48)	0.47 (0.50)	<0.001
Figures per paper	3.32 (2.88)	6.83 (2.44)	<0.001
Tables per paper	1.95 (2.13)	0.96 (1.75)	<0.001
Journal h-index	267.06 (206.80)	283.27 (211.69)	0.025
Journal impact-factor	4.52 (4.24)	5.42 (4.31)	<0.001
Scientific citations	70.42 (176.82)	69.64 (158.08)	0.899
Scientific citations (exclude self-cites)	63.30 (168.38)	61.82 (147.21)	0.801
Patent citations	2.09 (13.22)	2.89 (14.73)	0.085
Patent citations (exclude self-cites)	1.96 (12.60)	2.72 (14.44)	0.082
Any clinical cite (0/1)	0.41 (0.49)	0.21 (0.41)	<0.001
Clinical citations	1.85 (11.61)	0.42 (1.23)	<0.001
Number of authors	7.28 (8.50)	7.94 (4.01)	0.025
Number of institutions	3.08 (3.04)	3.00 (2.20)	0.467
Number of non-university institutions	1.60 (1.84)	1.35 (1.44)	<0.001
Year	2008.16 (9.42)	2013.42 (5.56)	

DRAFT

B. Additional Results

Detailed information on additional results is provided here and organized as follows:

1) Robustness regressions for Figure 2
2) Category of duplication breakouts for Figure 2
3) Time trends in USA and China
4) Image duplication by journal prestige
5) Image duplication by prestige of institutional affiliation
6) Correlation of number of collaborators with duplication
7) Correlation between duplication and future impact
8) Scientists with multiple articles presenting duplications
9) NIH-funded articles with duplication

1) Robustness regressions for Figure 2. Table S7 details the analogous regression to the descriptive Panel (a) of Figure 2 from the main body of the paper. Analyses are done by assigning each paper exclusively to its main method, but unreported results are consistent if we consider multiple methods for each paper. Table S8 details the analogous regression to descriptive Panel (b) of Figure 2. Analyses are done considering only papers with all authors affiliated with the same type of institution, but unreported results are consistent if we also consider papers with at least one author affiliated with a given type of institution. Table S10 details the analogous regression to descriptive Panel (c) of Figure 2. Analyses are done considering only papers with all authors affiliated with institutions from the same country, but unreported results are consistent if we also consider papers with at least one author affiliated with an institution from a

Table S7. Articles employing exclusively molecular/cellular and animal methods are more likely to present duplicated images than articles employing exclusively human methods.

Dependent Variable	Paper with duplication (0/1)			
Model	(1)	(2)	(3)	(4)
Molecular/cellular methods (0/1)	0.0136***			0.0324***
	(3.52)			(8.11)
Animal methods (0/1)		0.0315***		0.0427***
		(9.66)		(11.58)
Human methods (0/1)			-0.0385***	0
			(-11.12)	(.)
Constant	0.0151***	0.0114***	0.0435***	0.00510***
	(29.46)	(19.78)	(18.19)	(4.85)
Observations	48466	48466	48466	48466

Standard-errors clustered by journal. *t* statistics reported in parentheses.
Each column includes year fixed effects and journal fixed effects.
Note that the coefficient for human methods gets absorbed into the constant in model 4.
Significance codes: $^{*}\ p < 0.10, ^{**}\ p < 0.05, ^{***}\ p < 0.01$

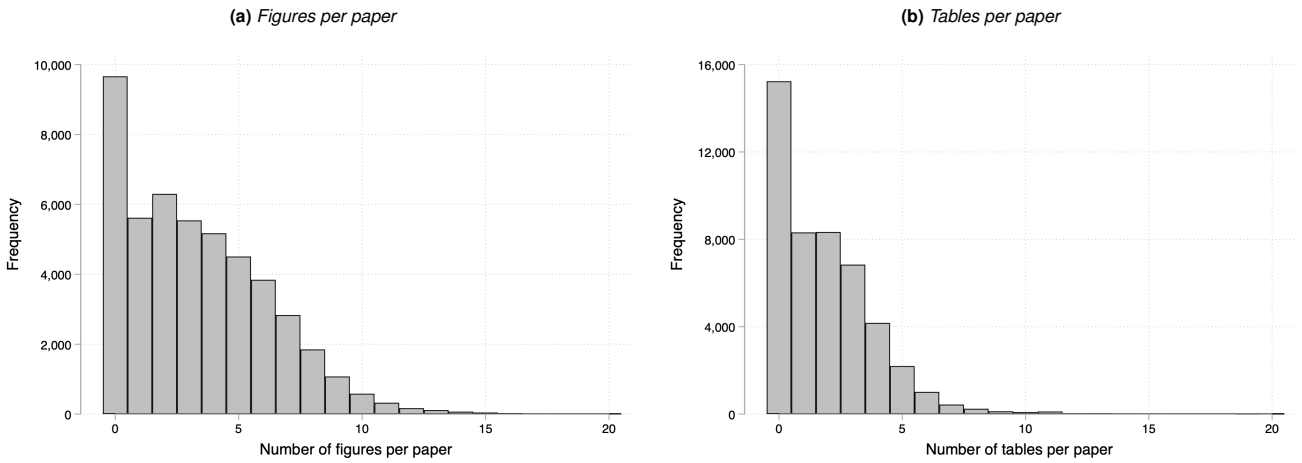
DRAFT

Table S8. Articles written by researchers exclusively affiliated with firms are less likely to present duplicated images than articles by researchers exclusively affiliated with universities.

Dependent Variable	Paper with duplication (0/1)			
Model	(1)	(2)	(3)	(4)
Researchers in firms (0/1)	-0.0117***			-0.00934**
	(-3.29)			(-2.51)
Researchers in health institutions (0/1)		-0.00150		0.000558
		(-0.80)		(0.30)
Researchers in universities (0/1)			0.00600***	0.00586***
			(3.30)	(3.15)
Constant	0.0171***	0.0170***	0.0148***	0.0149***
	(402.31)	(129.89)	(22.73)	(21.02)
Observations	48466	48466	48466	48466

Standard-errors clustered by journal. *t* statistics reported in parentheses.
Each column includes year fixed effects and journal fixed effects.
Significance codes: $^{*}\ p < 0.10, ^{**}\ p < 0.05, ^{***}\ p < 0.01$

Fig. S11. Histogram of number of figures and tables per paper in Alzheimer's research, 1980-2020.



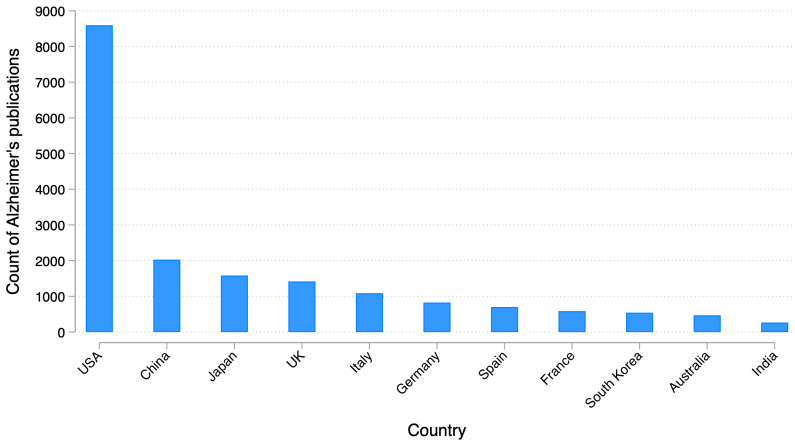


Fig. S12. Count of publications in Alzheimer's research by country, 1980-2020. Determined by geographical location of all of the authors' affiliations (i.e., we consider only papers with all authors from the same country).

given country. Regressions support the descriptive proportions illustrated in Figure 2 of the main paper, with the partial exception of the insignificant coefficients for Spain and India in table S10.

One may question the result illustrating that academic papers are more likely to present duplications, if the results simply reflect different research foci or methods of corporate scientists. Specifically, given the more applied nature of corporate relative to academic research, we would expect corporate scientists to perform more studies in humans, which were documented above as having a lower incidence of duplications. Table S9 directly explores this possibility by repeating the analysis of Table S8 while controlling for the method used in the paper. Despite the coefficient for university researchers in column (3) decreasing by almost half, it remains significant statistically (and still large). The coefficient suggests that around 45% of the difference can indeed be ascribed to corporate researchers being more likely to work on human subject research. Nonetheless, the overall pattern of results remains unchanged, supporting an argument that the different incentive structures of corporate researchers may lead to a lower incidence of image duplications.

2) Category of duplication breakouts for Figure 2. Refining the results of Figure 2 in the main paper, Figure S13 illustrates estimated means and confidence intervals for each type of duplication as a function of method, institution, and country of researchers, from OLS regressions.

Figure S13a breaks out the three methods (cell and molecule, animal, and human) against the more refined classification of the duplication as sloppy, unclear, or problematic. Consistent with the aggregate result, papers using cell and molecule methods demonstrate little variance; all categories associate positively and mostly significantly with duplication. Also consistent, papers using animal models demonstrate the greatest correlation with duplication; furthermore, the effect for sloppy papers is significantly greater than for problematic papers. Finally, all human methods correlate negatively with duplication, with the effect for sloppy papers significantly less than the effect for problematic papers.

Recall from the main paper that university-affiliated papers were more likely to present duplication and corporate papers less likely, with healthcare-affiliated papers intermediate between the two. Figure S13b provides additional nuance to this result. Consistent with an argument that publication (and assumedly time) pressure makes academics more likely to duplicate, the greater association of the academy with duplication is strongest for sloppy papers. Consistent with an argument that healthcare settings are more likely to provide protocols that minimize error, the intermediate result for healthcare affiliates breaks down into a positive association with problematic

Table S9. Articles written by researchers exclusively affiliated with firms are less likely to present duplicated images than articles by researchers exclusively affiliated with universities, even after controlling for the methods used in the articles.

Dependent Variable	Paper with duplication (0/1)			
Model	(1)	(2)	(3)	(4)
Researchers in firms (0/1)	-0.0166***			-0.0153***
	(-4.57)			(-4.05)
Researchers in health institutions (0/1)		0.000412		0.00137
		(0.22)		(0.74)
Researchers in universities (0/1)			0.00330**	0.00310**
			(2.39)	(2.17)
Constant	0.0171***	0.0169***	0.0157***	0.0159***
	(394.49)	(131.89)	(31.92)	(28.14)
Observations	48466	48466	48466	48466

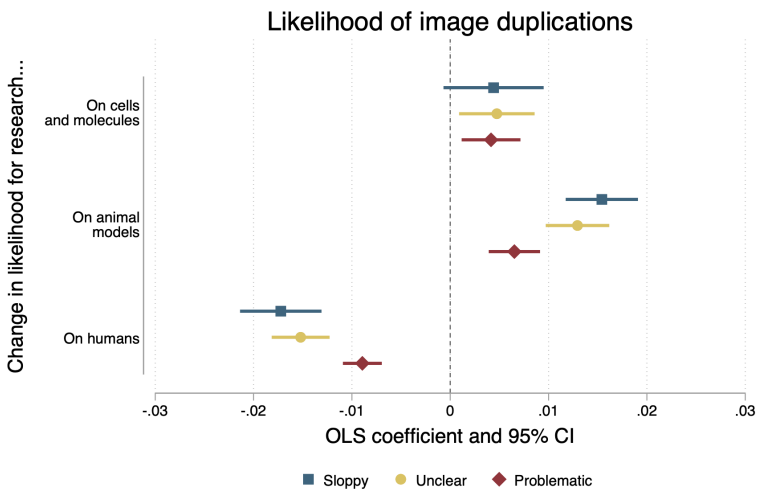
Standard-errors clustered by journal. *t* statistics reported in parentheses.
Each column includes fixed effects for the method used, as well as year fixed effects and journal fixed effects.
Significance codes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table S10. Articles written by researchers based exclusively within listed country.

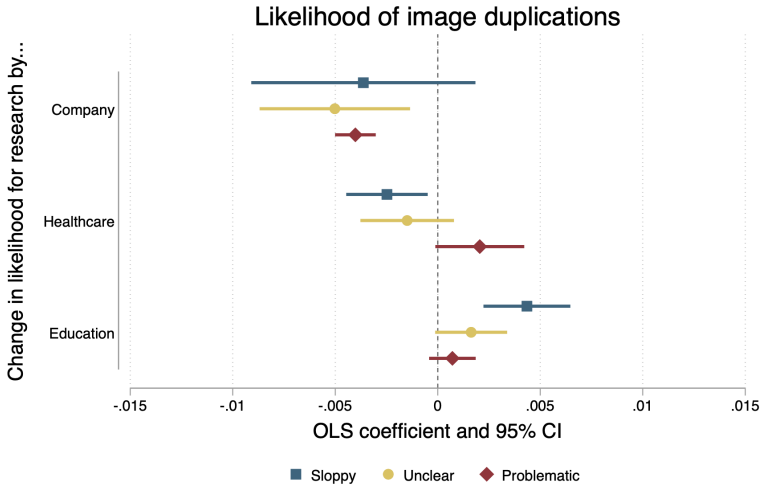
	Paper with duplication (0/1)										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
USA (0/1)	-0.00179										
	(-1.24)										
China (0/1)		0.0180***									
		(3.03)									
Japan (0/1)			-0.00722***								
			(-2.81)								
UK (0/1)				-0.00220							
				(-0.82)							
Italy (0/1)					0.00306						
					(0.57)						
Germany (0/1)						-0.0112***					
						(-3.19)					
Spain (0/1)							0.0101				
							(1.62)				
France (0/1)								-0.00807**			
								(-2.17)			
S. Korea (0/1)									-0.00709		
									(-1.37)		
Australia (0/1)										-0.00403	
										(-0.58)	
India (0/1)											0.0149
											(0.97)
Constant	0.0196***	0.0179***	0.0195***	0.0192***	0.0190***	0.0194***	0.0189***	0.0193***	0.0192***	0.0192***	0.0190***
	(47.57)	(45.17)	(145.45)	(152.85)	(98.41)	(202.28)	(130.86)	(269.26)	(208.79)	(179.51)	(141.92)
Observations	30236	30236	30236	30236	30236	30236	30236	30236	30236	30236	30236

Standard-errors clustered by journal. *t* statistics reported in parentheses.
Each column includes year fixed effects and journal fixed effects.
Significance codes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The smaller number of observations reflects missing data where country affiliation could not be determined from OpenAlex records.

DRAFT



(a) *Type of duplications by research methods*



(b) *Type of duplications by institutional affiliation*

Fig. S13. Association between categories of image duplications (sloppy, unclear, and problematic) and different characteristics of Alzheimer's research papers, 1980-2020. Each dot corresponds to the $\hat{\beta_{OLS}}$ from a separate regression, together with its 95% confidence interval. All regressions control for journal and year fixed effects and restrict data to papers with only the reported characteristics (i.e., authors are only from companies or educational institutions and not both on the same paper). Panel (a): Associations between type of image duplication and research methods. Panel (b): Associations between type of image duplication and type of institutions where authors of Alzheimer's papers are affiliated..

papers and a negative association with sloppy papers. Finally, there is no significant association between corporate papers and sloppy duplication and negative associations between corporate papers and unclear and problematic duplications.

Figure S14 breaks out categories of duplication against the country of institutional affiliations of the researchers, for all the countries of Figure 2 (except India, where the effects were imprecisely estimated). Significant negative associations include: USA with unclear and problematic, Japan with problematic, UK with problematic, Germany with sloppy, France with unclear, S. Korea with problematic, and Australia with unclear. Results only show positive and significant associations with a China affiliation and unclear and problematic duplications.

3) Time trends in USA and China. To better understand the sources of the time trends in Figure 1 of the main paper, we focus on the two main producers of Alzheimer's research, the United States and China. In Figure S15 we show papers from the U.S. and China with image duplication issues as a share of the total number of Alzheimer's papers from those two countries, respectively. While the incidence of duplications per year is relatively constant in the U.S., stabilizing between 2% to 3% every year, the graph shows a recent and marked increase in duplicated images among publications with authors wholey based in China. While more noisy, the numbers suggest that recently between 5% and 8% of Alzheimer's publications from authors based in China present at least one instance of duplication. This graph suggests that a substantial fraction of the increase in the Main Text Figure 1 is due to the growing share of Alzheimer's research carried out in China.

4) Image duplications by journal prestige. We obtain two commonly used metrics of academic journal prestige from the OpenAlex database (https://docs.openalex.org/). The first is the journal's h-index, a synthetic metric to capture the citation impact of the publications

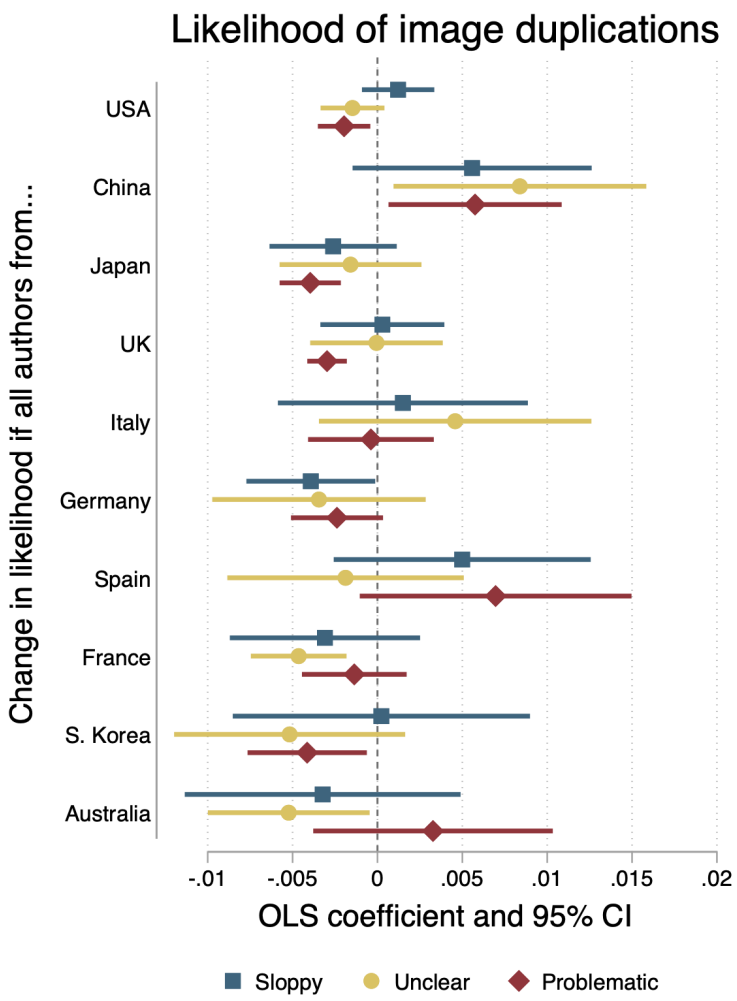


Fig. S14. Association between categories of image duplications (sloppy, unclear, and problematic) and countries of institutions where authors of Alzheimer's papers are affiliated, 1980-2020. Each dot corresponds to the $\hat{\beta_{OLS}}$ from a separate regression, together with its 95% confidence interval. All regressions control for journal and year fixed effects and restrict data to papers with every author affiliated with institutions in a given country. The estimates for India have been excluded because they are too imprecise.

featured by a journal. Mimicking the popular metric at the individual level, the journal h-index is defined as the maximum value of h such that the given journal has published at least h papers that have each been cited at least h times. The second metric is the journal impact factor, defined as the yearly mean number of citations of articles published in the last two years in a given journal. Both metrics are commonly employed as proxies for the relative importance of a journal within its field.

Figure S16 shows the distribution of journal h-index and journal impact factor for papers with and without duplicated images, respectively. Both figures show that the distributions are shifted to the right for papers presenting improper duplications, suggesting that they appear in journals more prestigious on average. Table S11 regresses duplication on the continuous h-index and impact factor variables, an indicator for a top 75%, and top 90% journal, respectively. The effect weakens for the indicator variables and becomes insignificant for the top 90% indicator.

In sum, articles with duplications tend to be published in journals with higher than average impact factor or h-index, as one would expect if authors duplicate images to increase the appeal of their findings. However, the most prestigious journals are not more likely to present image duplications, implying that peer review at better journals might be partially effective in weeding out papers with duplications, though it is not possible with this simple correlation to rule out selection bias.

5) Image duplications by prestige of institutional affiliation. We explore the incidence of image duplication among scientists affiliated with universities of differing prestige. We collect data from the Times Higher Education World University Rankings 2024, a comprehensive ranking including 1,907 universities across 108 countries and regions. The ranking is a composite metric based on a variety of factors, most notably teaching quality, research impact, and international outlook. A manual examination of the ranking suggests that it has a high face validity and limited geographical bias. More information and the full rankings are available here: https://www.timeshighereducation.com/world-university-rankings/2024/world-ranking.

For expositional convenience, we separate articles into four bins according to the prestige of their university of affiliation: top 20 institutions (2,449 articles), top 20 to 100 institutions (7,601 articles), top 100 to 200 institutions (6,775 articles), and remaining institutions (3,289 articles). Table S12 estimates indicators of top 20, 20-100, 100-200, and all other universities. It returns negative coefficients for the better universities. This finding is interesting in light of the fact that high-profile scientific malpractice cases in the news tend to involve

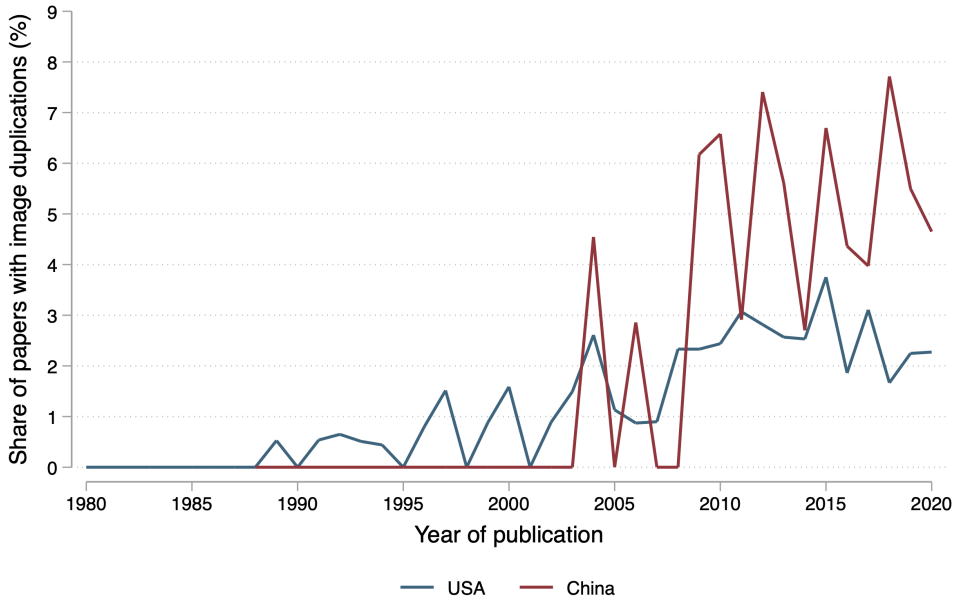
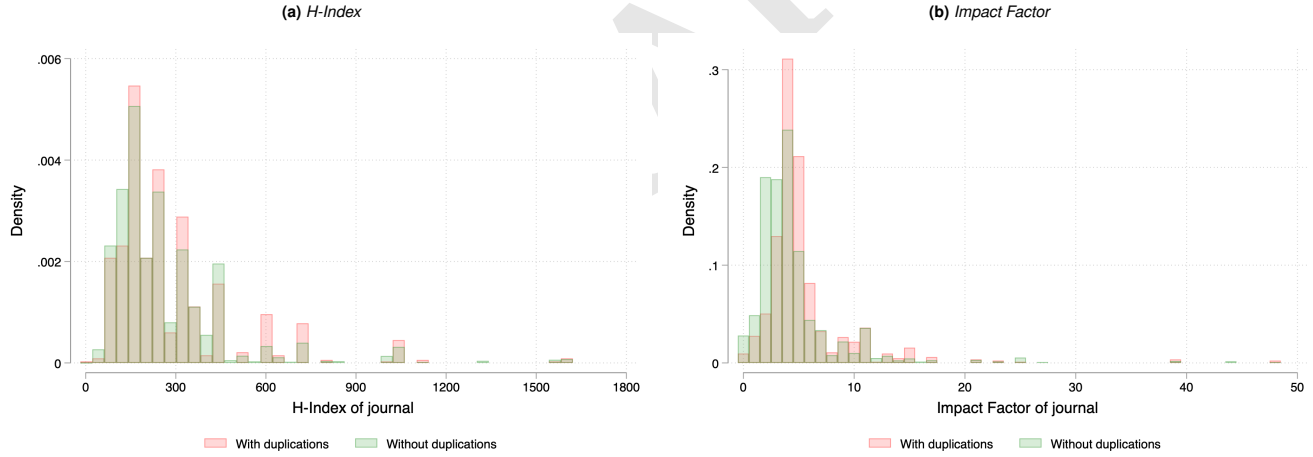


Fig. S15. Share of Alzheimer's research papers from the US and China with image duplication issues as a share of the total number of papers from those countries, 1980-2020. Data include all papers in the study population of Alzheimer's papers (i.e., they also include papers without any figures).

Fig. S16. Histogram of H-Index and Impact Factor of journals where Alzheimer's research is published, 1980-2020.



elite institutions (4). Our evidence suggests that these cases are more likely the exception than the rule and that scholars in top research universities are overall less likely to duplicate their images.

6) Correlation of number of collaborators with duplication. We explore the relationship between the size of the authorship team and the likelihood of duplicated images. The relationship is ambiguous ex ante. On the one hand, it could be that larger authorship teams increase coordination costs and the chances of mistakes. On the other hand, more authors may entail more cross-monitoring, helping reduce the instances of scientific malpractice. Table S13 returns negative and significant coefficients for the number of authors, academic institutions, and non-academic institutions. It appears that papers with more co-authors and/or more institutional affiliations are less likely to present image duplications, suggesting fewer duplications when there are a greater number of eyes on a paper.

7) Correlation between duplication and future impact. We explore the impact of papers with and without duplicated images on downstream research. Table S14 illustrates increased chances of any citation, total citations, and total citations without self-citations. Albeit just presenting means, the patterns are quite striking: the impact of papers including duplicated images is, on average, higher than papers without evidence of malpractice. This finding is interesting when compared with previous research showing that retracted papers see a heavy penalty in terms of citations received (12).

It could be that the increase in citations we document is driven by criticism of the paper (19), thus reflecting scientists' ability to recognize duplicitous images. We explore this possibility using proprietary data from Scite (https://scite.ai/). Scite uses a machine learning algorithm to categorize citations based on the text context. In particular, Scite locates text mentions of articles and uses the

Table S11. Papers with duplicated images are published in journals better than average but are not more likely to appear in the most prestigious journals.

Dependent Variable	Journal h-index	Top 75% journal h-index (0/1)	Top 90% journal h-index (0/1)	Journal impact factor	Top 75% journal impact factor (0/1)	Top 90% journal impact factor (0/1)
Model	(1)	(2)	(3)	(4)	(5)	(6)
Figure with duplication (0/1)	36.81**	0.0773*	0.0287	0.864***	0.0626*	0.0214
	(2.43)	(1.94)	(0.85)	(3.51)	(1.81)	(1.19)
Constant	265.8***	0.348***	0.155***	4.534***	0.188***	0.0753***
	(14.46)	(6.50)	(3.67)	(16.03)	(5.32)	(3.55)
Observations	48506	48506	48506	48506	48506	48506

Standard-errors clustered by journal. *t* statistics reported in parentheses.
All columns includes year fixed effects.
Significance codes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table S12. Researchers affiliated with the most prestigious universities are less likely to publish articles presenting duplicated images.

Dependent Variable	Paper with duplication (0/1)			
Model	(1)	(2)	(3)	(4)
Top 20 university	-0.00364**			
	(-2.00)			
Top 20-100 university		-0.00480***		
		(-3.28)		
Top 100-200 university			-0.00447**	
			(-2.12)	
Other universities				0.00259
				(1.19)
Constant	0.0184***	0.0193***	0.0188***	0.0155***
	(71.60)	(42.68)	(41.63)	(7.81)
Observations	35311	35311	35311	35311

Standard-errors clustered by journal. *t* statistics reported in parentheses.
All columns includes year fixed effects.
Significance codes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table S13. Papers with more authors and authors affiliated with different institutions are less likely to present duplicated images.

Dependent Variable	Paper with duplication (0/1)		
Model	(1)	(2)	(3)
Number of authors	-0.000121**		
	(-2.18)		
Number of institutions		-0.000808***	
		(-3.41)	
Number of non-university institutions			-0.00166***
			(-4.14)
Constant	0.0178***	0.0194***	0.0196***
	(43.79)	(26.64)	(30.58)
Observations	48466	48466	48466

Standard-errors clustered by journal. *t* statistics reported in parentheses.
All columns includes year fixed effects.
Significance codes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

DRAFT

Table S14. Papers with improperly duplicated images receive on average a higher number of scientific citations.

Dependent Variable	Any scientific citation (0/1)	Scientific citations (total count)	Scientific citations (exclude self-cites)
Model	(1)	(2)	(3)
Paper with duplication (0/1)	0.00380***	8.068**	7.395**
	(2.91)	(2.12)	(2.08)
Constant	0.991***	70.30***	63.18***
	(44877.83)	(1090.00)	(1049.40)
Observations	48466	48466	48466

Standard-errors clustered by journal. *t* statistics reported in parentheses.
Each column includes year fixed effects and journal fixed effects.
Significance codes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

surrounding textual context from the citing paper to predict whether the statement provides supporting or contrasting evidence for a referenced work, or simply mentions it. We use this information (which was graciously shared with us by the Scite team) in Table S15. Results are noisy but help rule out that the increase in citations we documented is driven by scientists criticizing duplicitous papers. This suggests that, on average, scientists do not check carefully enough or are not able to recognize and ignore articles presenting duplicated images.

Table S15. Papers with improperly duplicated images are not more likely to receive text mentions with a negative or critical tone by articles citing them.

Dependent Variable	Total text mentions	Total negative text mentions	Total positive text mentions
Model	(1)	(2)	(3)
Paper with duplication (0/1)	3.274	0.225	-0.0230
	(0.71)	(0.74)	(-0.83)
Constant	55.70***	3.136***	0.366***
	(712.72)	(611.56)	(781.07)
Observations	48466	48466	48466

Standard-errors clustered by journal. *t* statistics reported in parentheses.
Each column includes year fixed effects and journal fixed effects.
Significance codes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Inventors (or scientist-inventors) also build technology based on papers with duplicate images. We leverage data on the citations to Alzheimer's papers by U.S. patents taken from the Reliance on Science database (many U.S. patents are taken out by non-U.S. inventors, thus the U.S. data provide some international coverage). Table S16 shows, consistent with the paper citations, a greater likelihood of any patent citation, total citations, and total citations excluding self-citations. It appears that papers with duplications receive increased attention from both academics and firms, unfortunately obfuscating and possibly misdirecting future research and investment choices.

Table S16. Papers with improperly duplicated images receive on average a higher number of patent citations.

Dependent Variable	Any patent citation (0/1)	Patent citations (total count)	Patent citations (exclude self-cites)
Model	(1)	(2)	(3)
Paper with duplication (0/1)	0.0288**	0.967*	0.937*
	(2.32)	(1.89)	(1.83)
Constant	0.248***	2.089***	1.956***
	(1180.06)	(241.45)	(226.29)
Observations	48466	48466	48466

Standard-errors clustered by journal. *t* statistics reported in parentheses.
Each column includes year fixed effects and journal fixed effects.
Significance codes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Finally, we investigate the impact of articles with duplicated images on clinical translation. We use data from iCite about the number of citations each Alzheimer's paper has received from clinical articles, a direct indicator that the knowledge generated by that study has transitioned into clinical research. iCite considers "clinical" all articles describing clinical studies, clinical trials, clinical observational studies, or guidelines. In contrast to science papers and patents, Table S17 returns estimates that papers with duplicated images are less likely to be cited by any clinical paper and to receive fewer clinical citations in total.

8) Scientists with multiple articles presenting duplications. In our full sample, we have 130,702 unique authors. Of them, 5,567 are authors with at least one paper with a duplication. When decomposed by categories, we find that 2,282 authors had previously written articles with sloppy duplications, 2,379 with unclear duplications, and 1,403 with problematic duplications. Note that the same scientists could have written papers containing different duplication types, hence a total greater than 5,567.

DRAFT

Table S17. Papers with improperly duplicated images receive on average a lower number of citations from clinical studies.

Dependent Variable	Any clinical citation (0/1)	Clinical citations (total count)
Model	(1)	(2)
Paper with duplication (0/1)	-0.0518**	-0.140*
	(-2.48)	(-1.84)
Constant	0.412***	1.831***
	(1163.67)	(1420.29)
Observations	48466	48466

Standard-errors clustered by journal. *t* statistics reported in parentheses.
Each column includes year fixed effects and journal fixed effects.
Significance codes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Highlighting the authors with duplicated images in their previous work, Figure S17 shows the histogram of the number of duplicated articles per scientist. Table S18 presents results at the article level documenting whether the likelihood of a duplicated image is higher for researchers with duplications in their previous work. The smaller number of observations reflects the fact that the regressions exclude each author's first article. Regression results show that when any of an article's authors have a history of image duplications in their previous work, their papers are 1.57 percentage points more likely to present duplicated images. Compared to the sample mean of 1.71%, this coefficient implies a percentage increase of 92% in the likelihood of presenting a duplication. Columns 2, 3, and 4 show a monotonically increasing effect for authors who have previously published what we classified as sloppy, unclear, or problematic images. These results confirm the intuition that authors who engaged in sloppy or worse misconduct in the past are more likely to repeat it in the future.

Table S18. Papers with authors that previously published articles with duplications are more likely to publish new articles with duplicated images.

Dependent Variable	Future paper with duplication (0/1)			
Model	(1)	(2)	(3)	(4)
Authors' previous duplication (0/1)	0.0157***			
	(6.57)			
Authors' previous sloppy duplication (0/1)		0.0144***		
		(4.65)		
Authors' previous unclear duplication (0/1)			0.0208***	
			(6.25)	
Authors' previous problematic duplication (0/1)				0.0214***
				(6.09)
Constant	0.0133***	0.0151***	0.0144***	0.0150***
	(23.18)	(36.27)	(34.51)	(46.42)
Observations	43701	43701	43701	43701

Standard-errors clustered by journal. *t* statistics reported in parentheses.
Each column includes year and journal fixed effects.
Significance codes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

9) NIH-funded articles with duplications. We provide a "back-of-the-envelope" calculation of the amount of federal grant funding that has been allocated to academic articles presenting duplicated images. To assign the corresponding funding amount to each NIH-funded publication, we downloaded the details of all NIH projects acknowledged by PubMed-indexed articles from NIH RePORTER (https://reporter.nih.gov/). The resulting dataset is uniquely identified at the *article* × *grant* level, since two scenarios can emerge: 1) a publication acknowledges and thus is linked to multiple NIH grants; and 2) an NIH grant generates multiple publications. On average, each NIH-funded article acknowledges 2.22 NIH grants, and each NIH grant has an average of 17.44 academic papers assigned to it. In total, the dataset covers 2,585,075 academic articles and 329,590 NIH grants between 1908 and 2020.

The NIH RePORTER data reports only the full value of each NIH grant, regardless of how many publications acknowledge it. As such, there is no precise information on what proportion of the grant went to fund each specific article. To construct an approximate measure of funding per article, we divide the total value of each NIH grant by the number of academic publications related to it (including non Alzheimer's publications). This simple procedure leaves us with an average NIH funding of $85,674.90 per publication, which, however, does not yet take into account that each publication may acknowledge multiple grants. Therefore, we collapse the dataset at the article level and add the dollar amounts across all the relevant NIH grants for a given article. For example, consider the article with PubMed ID 1281980.[18] It receives money from three NIH grants with the total cost per paper for each of these grants being $54,768.69, $23,880.71, and $22,512.91. Therefore, we conclude that this article has received about $101,162.31 in NIH funding. Following the same procedure for every NIH-funded article, we find that each of them receives on average $190,574 from the NIH.

We merge this article-level information with our dataset of Alzheimer's research. Out of the 48,506 articles for which we have the PDF, 14,046 articles have received funding from the NIH according to the RePORTER data. Of those, 235 articles present instances of inappropriate image duplications. We find that 1.74% of NIH funding goes to papers that have some type of image duplication. This

[18] The article can be found here: https://pubmed.ncbi.nlm.nih.gov/1281980/

DRAFT

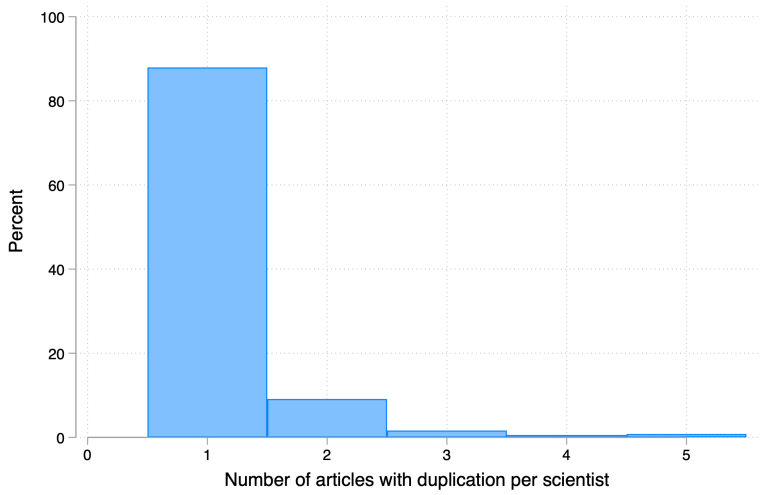


Fig. S17. Distribution of Alzheimer's scientists by the number of their articles presenting evidence of inappropriate image duplications.

number is very similar to the overall incidence of duplication in our sample. Given that each of these articles has been granted on average $338,539.90, this leads to an estimate of at least 235 × $338,539,90 = $79,556,886 of federal funding absorbed by articles with duplications. We convert all monetary values to constant U.S. dollars in 2020 to facilitate comparison.

In order to understand the trickle-down effect on follow-up research undertaken by the academic community, we go one step further and ask ourselves how much NIH funding has gone to academic articles that cite at least one article with instances of inappropriate image duplications. We find 45,758 academic articles that cite any of our 832 academic articles with image duplications. Out of those, 8,288 have been funded by the NIH (18.1%). Given that each of these articles has been granted on average $344,234.10, this leads to an estimate of at least 8288 × $344,234.10 = $2,853,012,224 of federal funding absorbed by articles that cite previous work with image duplications.

DRAFT