Retracted, still cited as support: a census of post-retraction citations to 22,483 retracted biomedical papers

Stephan Joseph (ORCID 0009-0001-0250-5446)
SaneApps LLC (SaneCite, an independent public-interest research project); Georgia, USA
Correspondence: letters@sanecite.com

Keywords: retraction; citation analysis; citation context; research integrity; evidence synthesis; large language models

Abstract

Background. Retracted papers keep being cited. Studies that read the citing text have covered at most about a thousand citations, so how often citing text uses a retracted paper as support has not been measured at scale. **Methods.** A pre-registered census of every Retraction Watch retraction with a PubMed ID, retracted 1990–2023 (22,483 papers). Europe PMC supplied citations from later calendar years. In each open-access citing paper we located every in-text citation of the retracted paper and classified it with a panel of three language models and a fixed rule. The panel was tested against a separate language model's blind labels on two random samples and, for acknowledgement of the retraction, against published hand-confirmed labels. **Results.** 15,992 retracted papers (71.1%) were cited in a later year (130,229 citations); we read 55,676 (42.8%). Excluding administrative retractions, 40.7% of 51,614 judged citations were cited as support (95% CI 39.9–41.5; 38.1% after correcting for the error measured on the blind sample) and 3.3% (2.7–4.2) were flagged as retracted or unreliable. The rate fell from 43.5% one to two years after retraction to 34.9% after more than ten. Where the citing paper's own reference list marked the work as retracted, 40.4% were still cited as support. An exploratory check found 472 meta-analyses, systematic reviews and guidelines that counted a retracted study as data. A registered extension beyond open access raised the share read to 48%, with a combined rate of 40.1%. **Conclusions.** In open-access biomedicine, text that cites a retracted paper rarely mentions the retraction, and about two in five such citations cite the retracted work as support. The retraction reaches reference lists more often than it reaches the citing text.

Introduction

A retraction is a journal stating in public that a paper should not be trusted. It is the strongest correction the literature has, and it is machine-readable: the Retraction Watch database, which Crossref has made openly available since 2023, listed 67,706 retractions, with dates and reasons, when we downloaded it on 13 October 2026. If citing authors heard that message, citations would stop or would mention the retraction.

They do not. In the PubMed Central open-access subset, only 5.4% of citation contexts citing a paper after its retraction acknowledged the retraction: 722 of 13,252 contexts, in a study covering

7,813 retracted papers (Hsiao and Schneider 2021). Smaller studies that read the citing text found that most post-retraction citations treat the work as valid (Budd et al. 1998; Bar-Ilan and Halevi 2017), and in anesthesiology about one in ten citations from original research appeared to build on the retracted paper (Frasco et al. 2022). Retracted trials also reach systematic reviews and guidelines, often after the retraction (Kataoka et al. 2022; Fanelli et al. 2022; Xu et al. 2025).

A count of citations cannot tell a warning from a use. A paper may cite a retracted study to criticise it, to list it among many, or to build on its result. Only the last carries the retracted claim forward. Reading the citing text answers the question, and it has been done only on samples of a few hundred to about a thousand citations, because it means reading every citing sentence. Open full text for a large share of biomedicine (Europe PMC), the largest open list of retractions (Retraction Watch via Crossref), and language models cheap enough to read every sentence now make a census possible, provided the models' reading is checked.

This is, to our knowledge, the largest reading of post-retraction citations to date: 55,676 citations (citing pairs) from 51,957 citing papers to 12,857 retracted papers, 5.8 times the citing pairs and 4.7 times the retracted papers of the largest earlier citation-context study (9,548 pairs to 2,763 papers; Hsiao and Schneider 2021), from a frame 2.9 times larger. It asks whether each citation uses the retracted work as support, not only whether the retraction is mentioned. The error of its model labels is measured on blind random samples and against published labels, and the method, code, labels and every citing passage are public.

Methods

Design and registration

This is a census of post-retraction citations, an observational cross-sectional study of published papers, reported following STROBE where it applies (von Elm et al. 2007; checklist in the supplement). The analysis protocol was posted before the final analysis (github.com/MrSaneApps/sanecite-retraction-census, 13 October 2026, archived at web.archive.org); its dated amendments are listed under "Deviations from the registered protocol" below.

Frame

We took every record in the Retraction Watch database (via Crossref, downloaded 13 October 2026) whose nature is "Retraction", that gives a PubMed ID for the original paper, and that was retracted between 1990 and 2023. Retractions from 2024 on were left out so that every paper has at least two full calendar years of exposure after retraction. Records were deduplicated by PubMed ID: 22,553 records. We then checked each record's PubMed ID against Europe PMC. A record whose PubMed title shares under half the words of the Retraction Watch title points to a different work, unless the difference is explained by a translated or re-titled record with the same DOI, by a PubMed record replaced by its own retraction notice, or by a generic Retraction Watch title with the same DOI; whole meeting-abstract supplements count as different works because their citations cannot be attributed to one abstract (protocol amendment 11). This removed 70

records carrying 86 later citations, leaving 22,483 retracted papers. This is a census of that frame, not a sample.

Finding later citations

For each retracted paper, Europe PMC was queried for papers that cite it and were published in a calendar year after the retraction year (`CITES:<pmid>_MED AND PUB_YEAR:[ry+1 TO 2026]`). Citations in the retraction year itself were not counted, so every citation count is a lower bound. Citing papers with open-access full text in Europe PMC (`OPEN_ACCESS:y AND HAS_FT:y`) were read; the rest were counted only.

Locating the citation

For each (retracted paper, citing paper) pair we fetched the citing paper's JATS XML and found the retracted paper in its reference list: by PubMed ID, then DOI, then title. A title match had to cover at least 85% of the title's words, the entry could not carry another paper's ID, the entry had to name the original publication year (±1), and only the single best entry was taken. In the body text we found every in-text citation pointing at that entry, including citation ranges ("[16–19]", "[51]–[53]"), unlinked numbered citations in numbered reference lists, and table cells. Each citing sentence was kept with one sentence either side, up to three mentions per paper. Formula markup was stripped; numbered lists that only look like citations ("[1] aged 18–40; [2] …") were dropped. Retraction notices, corrections and expressions of concern that cite the paper were set aside, by JATS article type or by a title beginning "Retraction", "Retracted", "Correction", "Erratum", "Corrigendum", "Expression of concern" or "Notice of".

Reading the citation (codebook v3.3)

Whether the retracted paper shares its citation group with other references is counted by code (for example "[31, →72]" is two references), not judged by a model. If the paper is cited more than once in a sentence, its smallest group counts; numbers that belong to names ("BMP-12") are not read as citations. Code also notes when the paper backs one item of a list on its own. Each mention is then read by three models from different families, run on Cloudflare: Jev (typesafe/jev), Clef (@cf/cloudflare/clef) and Gemma 4 26B (@cf/google/gemma-4-26b-a4b-it). None of them is told the paper was retracted. Each answers two questions:

- stance: anywhere in the passage, including the sentences either side, is the work called retracted, withdrawn, fabricated, discredited, erroneous or flawed, are its data called confounded or unreliable, or are its findings said to have failed to replicate? Plain disagreement ("our results differ from") does not count.
- use: does the sentence use a specific finding, number, method or conclusion of the work (the codebook's test: "if its data had been made up, would this sentence lose its support?"); is the work counted or pooled as data in a review; is the sentence general background; or is the work only described?

A fixed rule turns each model's answers into a label (the codebook's internal names, substantive and perfunctory, follow Moravcsik and Murugesan 1975). Flagged as retracted or unreliable: the model says the text disputes the work. Cited as support: the work is counted as data in a review and the passage has a counting word (studies, trials, included, pooled, meta-analysis and similar), or the work is the only reference in its group and the sentence uses its finding, number, method or conclusion. Mentioned in passing: everything else (shared citations, background, description). Passages with formula markup are unclear. The panel label is flagged if any model says the work is disputed and the passage contains a retraction, withdrawal, fabrication, falsification, misconduct, plagiarism, error or image-integrity word, or if two models say so. Otherwise it is the majority of the three labels, with a model's lone flagged vote counted as support; when all three differ, Jev's label decides. A citing pair's label is the strongest across its mentions. Gemma replies cut off at the default length were retried once with a larger cap.

Development and validation

The extractor, codebook and decision rule were developed on 321 citing papers labelled by a separate language model (Claude, Anthropic) working from the codebook and blind to the panel's output ("reference labels" below). Tie-breaker models (GLM 5.2, GPT-OSS 120B, Qwen 3.8) were tried on the citations where the panel disagreed; none beat the panel's own majority, so none is used. Version 3.1 was then frozen and tested on a blind holdout: a simple random sample of 300 citing papers from the census (seed 31337), labelled before any model output was read. It agreed on 277 (92.3%, 95% CI 88.8–94.8). The errors showed two counting bugs and missed criticism in neighbouring sentences; these were fixed (version 3.3), and because that used the holdout, version 3.3 was frozen and tested again on a second fresh blind sample of 150 (seed 4242): 142 agreed (94.7%, 95% CI 89.8–97.3); where all three models agreed (126 of 150) 123 were right (97.6%, 93.2–99.2). The panel's rate of cited as support ran above the reference labels (by 2.3 points on the holdout with version 3.1, and 2.7 points on the confirmation sample), so a rate corrected for the measured misclassification (Rogan-Gladen; Rogan and Gladen 1978) is reported beside the raw rate.

For the flagged label we also compare with people: Hsiao and Schneider (2021) screened post-retraction citation contexts for cue words or a co-cited retraction notice and confirmed the flagged contexts by hand (data: Illinois Data Bank, doi:10.13012/B2IDB-8255619_V2). On the citing pairs both studies hold we report the agreement of our flagged label with theirs (Cohen's kappa) and our flagged rate over their citing years.

Retraction reasons and fields

Retraction Watch reason tags were grouped: misconduct, fabrication or duplication (including paper mills, plagiarism, compromised peer review and computer-generated content); error or unreliable results; other or unstated; and administrative, where no fault was found in the science (withdrawn as out of date, retracted and replaced by a corrected version, withdrawn to publish elsewhere, a publisher's duplicate, temporary removal). Process tags (investigation, notice details, updates) are ignored. A paper is administrative only if all its remaining tags are

administrative; otherwise administrative tags are dropped before grouping. Headline rates exclude administrative retractions; they are reported separately. Field is the set of the retracted paper's Retraction Watch subject codes: basic life sciences only, health sciences only, both, or a code outside those two.

Statistics

The unit is a citing pair: one citing paper's use of one retracted paper. Rates are counts over pairs that could be judged, rounded once. 95% confidence intervals come from a cluster bootstrap over retracted papers (citations to the same paper are not independent; 400 resamples, 200 for the robustness checks). The corrected rate uses the Rogan-Gladen estimator with the sensitivity and false-positive rate measured on the confirmation sample; its interval also resamples that sample. For citations where the three models did not all agree we also report a strict and a lenient rate. Validation accuracy is given with Wilson intervals. As a check on the interval, a two-way bootstrap also resamples citing papers.

Robustness

The primary rate is recomputed without citing papers that are themselves retracted (S1), with only citations two or more years after retraction (S2), with only pairs the three models judged unanimously (S3), with the first passage only (S4), and with only references matched by PubMed ID or DOI (S5). We report agreement between the models (Fleiss' kappa) on the label, on stance, on the raw use answer and on passages where the work is cited alone; a test-retest of 500 passages re-asked of every model at the census settings, which measures how repeatable the hosted models are; and, as exploratory analyses, self-citation (shared surname and first initial, which can over-match common surnames) and timing: the citing paper's JATS received date (or accepted date) against the Retraction Watch retraction date, with the rate recomputed without citing papers received before the retraction (S6).

Open-access coverage

For every later citation, read or not, Europe PMC metadata gave the citing year and publication type. We compare read and unread citations by citing year, publication type, retraction reason and field, and reweight the primary rate so that each retracted paper counts by all its later citations, and to the citing-year mix of all later citations.

Reading beyond open access

Amendments 6 to 11 registered, before each measurement, the reading of later citations whose citing paper is not open access in Europe PMC. Each new source passed an extraction gate fixed in advance (usable yield at least 85% on two samples for PDF copies; for PMC author manuscripts, status agreement with the census of at least 95%, passage agreement of at least 0.95, and a hand check of 30 extracted pairs) before any model call. Every set was read by the same frozen panel and codebook, analysed with the census rules, deduplicated so that each citation is read once, and reported with its own denominator; a combined rate is a rate over the

readable corpus only. Copies were fetched only where openly offered, with robots rules and rate limits obeyed.

Exploratory: retracted studies counted as data in syntheses

Every pair the panel labelled as a meta-analysis, systematic review or guideline counting the retracted study as data was read again at passage level by a language model (Claude) to check that the study was included in the synthesis rather than listed as excluded, mentioned as a previous review, or used for another purpose. Inclusion dates are compared with the retraction by calendar year and, where the citing paper gives one, by received date.

Deviations from the registered protocol

Each amendment is dated in the protocol and was posted before the measurement it governs, unless stated.

- 13 Oct 2026 (amendment 1): the public codebook file was regenerated from the code actually used; added the interval for the corrected rate and the scripts for the planned analyses. No rule changed.
- 14 Oct 2026 (amendment 2): added the robustness checks S1 to S5, model agreement, self-citation and test-retest; disclosed one interim run on partial labels. No rule changed after it.
- 14 Oct 2026 (amendment 3, after the final analysis): corrected the reason and field grouping; ran two promised analyses the code had skipped; added the exploratory analyses marked as such. Pairs still lacking a Gemma label were set aside rather than decided by the other two models as amendment 2 said; counting them leaves the rate unchanged. No label changed.
- 14 Oct 2026 (amendment 4): fixed the scoring of the planned human labelling round before any human answer existed.
- 15 Oct 2026: rates are computed from their counts and rounded once; intervals stored to six places.
- 16 Oct 2026 (amendment 5): released all code; no analysis change.
- 19 to 20 Oct 2026 (amendments 6 to 11): registered the extended reading described above; withdrew one HTML-based set whose source did not permit batch download; required every labelled set to pass a panel integrity check; added the frame check that removes Retraction Watch records whose PubMed ID points to a different work, applied to every set, this census included.

Use of AI

Language models (Jev, Clef, Gemma) produced the citation labels as described above. Claude (Anthropic) wrote most of the analysis code, produced the development and validation reference labels, read the synthesis cases again, ran an internal review of the analysis, and helped draft and

edit this manuscript. The author checked all of it and takes responsibility for the content. No AI tool is an author.

Cost and openness

Every model call and its tokens are logged. Cost is computed from logged usage at Cloudflare's list prices (Clef $0.24 per million input tokens; Gemma from the neurons Cloudflare reports, at $0.011 per 1,000; Jev $0.042 per million input tokens), and for Jev it is also given from Cloudflare's AI Gateway billing record for the run (13–14 October 2026). All code (MIT), the frame, every label and every citing passage are public (Data availability), and every figure in this paper is generated from the result files.

Results

Unless stated otherwise, a "citation" below is one citing paper's use of one retracted paper (a citing pair), and 95% confidence intervals come from a cluster bootstrap over retracted papers. "Cited as support" is the label for a citing sentence that uses a specific finding, number, method or conclusion of the work, or a review that counts the work as data. "Flagged as retracted or unreliable" is the label for text next to the citation that calls the work retracted, withdrawn, fabricated or unreliable, or says it failed to replicate. "Mentioned in passing" is everything else.

Frame and coverage

Of the 22,483 retracted papers, 15,992 (71.1%) were cited in a calendar year after their retraction, 130,229 citations in all. For 58,329 citations the citing paper had open full text. We located the citation in the text for 56,167 (96.3%); the rest failed because the retracted paper could not be found in the reference list (1,013), had no in-text citation (967), or the full text or body was missing (182). We set aside 485 retraction notices and related items and 6 citations for which one model never returned an answer, leaving 55,676 read citations to 12,857 retracted papers (42.8% of all later citations). Headline rates exclude the 4,044 citations to administrative retractions (reported separately) and 18 the panel could not label: 51,614 citations to 12,321 retracted papers.

The read citations are more recent than the unread ones (41.8% vs 28.4% published 2023–2026), because open access has grown, but they are alike in citing paper type (research articles 64.4% vs 58.9%, reviews 27.5% vs 27.2%), retraction reason (misconduct 57.4% vs 54.1%, error 28.2% vs 32.1%) and field (basic life sciences only 33.4% vs 33.2%, health sciences only 26.3% vs 21.1%). Weighting the read citations so that each retracted paper counts in proportion to all its later citations gives a rate of cited as support of 39.5% (38.4–40.3); weighting to the citing-year mix of all later citations gives 40.0% (39.0–40.8).

How retracted work is cited

Label	Share of 51,614 citations	95% CI
Cited as support	40.7% (21,032)	39.9–41.5
Mentioned in passing	55.9% (28,857)	55.1–56.7
Flagged as retracted or unreliable	3.3% (1,725)	2.7–4.2

Corrected for the misclassification measured on the blind confirmation sample (sensitivity 0.965, false-positive rate 0.065), the rate of cited as support is 38.1% (33.5–41.8); with that sample limited to the headline's scope (137 non-administrative pairs) it is 37.0%. The three models agreed on 46,847 of the 55,676 read citations (84.1%), two of three on 8,806, and none on 23. Over the 8,829 citations they did not all agree on, counting a citation as support only when no model called it passing gives 34.5% (strict); counting it when any model called it support gives 46.7% (lenient). A bootstrap that also resamples citing papers (some cite several retracted papers) gives 39.7–41.9.

Of the 12,321 non-administrative retracted papers with a judged citation, 8,193 (66.5%) were cited as support at least once after retraction and 3,983 (32.3%) at least twice.

Time since retraction

Years after retraction	Citations	Cited as support	Flagged as retracted or unreliable
1–2	25,914	43.5% (42.6–44.4)	2.7% (2.1–3.7)
3–5	13,700	39.1% (38.0–40.2)	3.4% (2.9–3.9)
6–10	8,271	37.4% (35.9–38.8)	3.8% (2.8–5.1)
11 or more	3,729	34.9% (32.5–37.0)	6.4% (3.1–12.8)

Cited as support falls from 43.5% in the first two years to 34.9% after more than ten, and the share flagged as retracted or unreliable rises but stays low.

Many citations a year after retraction could have been written before it. For 21,951 citations (42.5%) the citing paper's full text gives the date it was received (or, failing that, accepted). Only 1,483 of these (6.8%) were received before the retraction; leaving them out gives 40.6% (39.8–41.4). Among citing papers received after the retraction, the rate was 43.8% in the first calendar year after it, 40.5% in the second and 37.8% later, so the decline is not an artefact of papers written before the retraction (exploratory).

By decade of retraction, the rate rose from 34.3% (1990s, 405 citations) to 44.4% (2020s, 23,032 citations); this overlaps with the time effect, since citations of recent retractions are mostly early ones.

Reason for retraction

Reason group	Citations	Cited as support	Flagged as retracted or unreliable
Misconduct, fabrication or duplication	31,942	41.2% (40.2–42.4)	1.9% (1.2–3.0)
Error or unreliable results	15,711	40.0% (38.9–41.2)	6.5% (5.1–8.2)
Other or unstated	3,961	40.2% (37.7–42.3)	2.6% (2.0–3.4)
Administrative (no fault found; not in headline)	4,044	38.0% (33.9–42.9)	1.4% (1.0–1.9)

The rate of cited as support was the same whatever the reason. Citations of papers retracted for misconduct, which includes paper mills, fabrication, plagiarism and compromised peer review, flagged the problem in 1.9%; those retracted for error, in 6.5%.

Who cites retracted work

Citing paper type	Citations	Cited as support	Flagged as retracted or unreliable
Research article	33,642	12,149 (36.1%)	816 (2.4%)
Review	14,238	7,283 (51.2%)	588 (4.1%)
Clinical trial	1,117	397 (35.5%)	30 (2.7%)
Meta-analysis	865	487 (56.3%)	84 (9.7%)
Systematic review	659	358 (54.3%)	58 (8.8%)
Case report	660	241 (36.5%)	7 (1.1%)
Letter, editorial or comment	399	99 (24.8%)	141 (35.3%)
Guideline	30	16	0
Other	4	2	1

Types are Europe PMC publication types; rows under 50 citations show counts only. Reviews and evidence syntheses cited retracted work as support most often: more than half of their citations did. By the retracted paper's Retraction Watch subject codes, the rate was highest for basic life sciences only (44.4% of 18,380), then basic life and health sciences together (41.3% of 16,723), health sciences only (37.2% of 11,432) and papers with a code outside those two, often alongside them (34.0% of 4,997); 82 citations went to papers with neither code.

When the reference list already says "retracted"

For 3,628 citations the citing paper's own reference entry contained the word "retracted". In 2,664 (73%) it was part of the title ("Retracted article: ..."), as NISO recommends and many publishers now do (NISO 2024); in 523 (14%) it was a note pointing to the retraction notice. Even so, 40.4% (37.4–43.5) of these citations were cited as support and only 12.9% (7.6–18.0) were flagged as retracted or unreliable in the text.

Retracted studies counted as data in evidence syntheses (exploratory)

Inclusion of retracted trials in systematic reviews and guidelines is well documented (Kataoka et al. 2022; Fanelli et al. 2022; Xu et al. 2025; Graña Possamai et al. 2025). Here we asked how often it shows up across all retracted papers and study designs, found by reading the citing text. The panel judged 570 citations to be a meta-analysis, systematic review or guideline counting the retracted study as data. A second reading of each passage confirmed that the study was included in the synthesis in 507 (89%): 306 meta-analyses, 198 systematic reviews and 3 guidelines, 472 citing documents in all, drawing on 395 retracted papers, 226 of them retracted for misconduct, fabrication or duplication. In 51 the study was not included, 8 were unclear, and 4 were papers auditing the retracted trials, which the panel should have labelled flagged.

Some of these inclusions were unavoidable, because the retraction came after the review's search. 299 of the 507 came two or more calendar years after the retraction. Of the 193 with a received date, 21 were received before the retraction, 26 within six months after it, and 146 six months or more after it.

Agreement with published labels

Hsiao and Schneider (2021) screened post-retraction citation contexts for cue words (retract*, withdr*, error) or a co-cited retraction notice and confirmed the flagged contexts by hand; contexts without a cue were counted as not acknowledging. Of their 9,548 post-retraction citing pairs, 9,300 fall in our frame and we read 8,944 (96.2%). Our flagged label agreed with their acknowledgement label on 97.7% (Cohen's kappa 0.74). We found 81.2% of the citations they marked as acknowledging the retraction. Of the citations we flagged, 70.6% were marked so by them; our label is broader by design, since it also counts text calling the work fabricated or unreplicated without the word "retracted". On these shared pairs they found acknowledgement in 4.4% and we in 5.1%. Over the citing years their data cover (to 2019), our census rate is 5.1% (4.0–6.5).

Reliability and robustness

Across 72,111 passages the three models agreed well on the label (Fleiss' kappa 0.82) and on whether the text disputes the work (0.80). Part of the label agreement is built in, because code, not the models, labels a citation shared with other references as passing; on the raw use answer kappa is 0.57, and on the label for the 35,278 passages where the work is cited alone it is 0.62. Asked again, on 500 random passages, each model gave the same label 98.8%–100% of the time and the panel 99.2%. For 3 of the 150 confirmation pairs the census run gave a different label from the confirmation run, in line with that figure.

Check	Cited as support	95% CI	Citations
Primary	40.7%	39.9–41.6	51,614
S1 Citing paper not itself retracted	40.7%	40.0–41.6	51,189
S2 Two or more years after retraction	39.0%	37.9–39.9	35,506
S3 Only citations whose every passage the three models agreed on	38.7%	37.9–39.4	43,551
S4 First passage only	36.1%	35.3–36.8	51,604
S5 Reference matched by PubMed ID or DOI	40.9%	40.1–41.6	45,815
S6 Citing papers received before the retraction removed (exploratory)	40.6%	39.8–41.4	50,131

Every row of this table comes from one bootstrap run of 200 resamples, so the primary interval here differs slightly from the 400-resample headline interval. Every check lies between 36.1% and 40.9%. S4 is lower because a citation's label is the strongest of up to three passages. In an exploratory analysis, 972 citations (1.9%) were self-citations by a shared surname and initial; they were cited as support slightly more often (43.1% vs 40.7%) and flagged about as often (4.2% vs 3.3%).

Reading beyond open access (registered extension)

After the census, protocol amendments 6 to 11 extended the reading to later citations whose citing paper is not open access in Europe PMC: free legal copies located through Unpaywall and read from PDF, and author manuscripts from the PMC Article Datasets licensed for text mining. Each set was gated on its own extraction accuracy before any model call and is reported with its own denominator. The free-copy set gave 38.6% (1,019 of 2,642); a small fetch-recovery set gave 6 of 19; the author-manuscript set gave 33.1% (1,345 of 4,064; 95% CI 31.4–34.7). With the census, 63,026 later citations have been read (48% of all later citations), and across the 58,339 judged the rate of cited as support is 40.1% (39.3–40.9). The two larger new sets run lower than the open-access census; they come from a different mix of papers, and we report the difference without explaining it. The rest of this paper keeps to the open-access census.

Cost

The census took 216,633 model calls. From logged usage at Cloudflare's list prices it cost $32.05 (Clef $21.91, Gemma $5.68, Jev $4.46); Cloudflare's billing record for the Jev calls was $1.835, because fewer tokens were billed than Jev reports per call.

Discussion

About two in five post-retraction citations of retracted biomedical papers, in the open-access literature we could read, cite the retracted work as support (40.7%), and about one in 30 is flagged as retracted or unreliable (3.3%). A third of retracted papers that are still cited are cited

as support at least twice. These rates hold under every robustness check, after correcting for measured error, after removing citing papers submitted before the retraction, and after reweighting the read citations to match all later citations by year and by paper.

Earlier studies found that most post-retraction citations treat the work as valid (Budd et al. 1998; Frasco et al. 2022). On that definition, every citation not flagged as retracted or unreliable, our figure is 96.7%. Cited as support is a stricter subset: the citing sentence uses the work, not just mentions it. Our acknowledgement rate is close to the published one where the two can be compared: on the citations both studies read, Hsiao and Schneider found 4.4% and we found 5.1%, and over their citing years our census rate is 5.1%. Across the full census it is lower (3.3%), because the census is dominated by recent citations, which flag the retraction less often (1.9% of those published in 2024, against 7–14% in 2009–2014).

Three findings bear on why notices fail. First, use fades slowly: more than a decade on, 34.9% of citations still cite the retracted work as support. Second, the warning often reaches the bibliography but not the text: when the citing paper's own reference list says "Retracted", 40.4% of those citations still cite the work as support. NISO recommends the title prefix so that readers consider it when deciding how to use the content (NISO 2024); in our data the prefix arrived, probably with the title when the reference was imported, and the text did not change. This fits the finding that post-retraction citations come mostly from authors who find papers by quick search outside their own field (Woo and Walsh 2024). Third, evidence syntheses use retracted work most: meta-analyses and systematic reviews cited it as support in more than half their citations, and 472 counted a retracted study as data, in line with studies of retracted trials (Xu et al. 2025 found 1,330 retracted trials included in 847 systematic reviews).

Flagging alone has not been enough. Reference managers have warned about retracted items since 2019, and submission tools check reference lists against Retraction Watch. Telling citing authors about a retraction did not reduce later citations in a randomised trial (DeVito et al. 2025), and alerts to the authors of reviews and guidelines rarely led to corrections (Avenell et al. 2024). What is missing is a check that editors must act on: a retracted reference should need a stated reason before acceptance, and a synthesis should not be published while counting a retracted study. Reading every citing passage, as we did, cost $32.05 for the whole census at list prices, so the checking itself is not the obstacle.

Retraction is also a narrow filter. When Cochrane reviewers assessed the trustworthiness of the trials in a set of their reviews, they set aside far more trials than journals had retracted (Weeks et al. 2023). Our census measures only the part that journals have already marked, and even that part keeps being used.

Limitations

- Coverage. The census reads PubMed-indexed retractions and open-access full text only; 57% of later citations are in papers we did not read. They are counted, and reweighting by year and paper moves the rate by about one point, but their use is not observed. The

registered extension to other copies gives a combined rate close to the census rate, and its larger sets run lower.

- Citations in the retraction year are excluded, so citation counts are lower bounds; rates are not affected in a known direction.
- A citing paper's label is taken from up to three passages, the strongest one counting; the rate rises with the number of passages read, and the first-passage-only rate is 36.1%.
- Labels judge the citing text, not whether the citing paper's conclusions depend on the work. Cited as support also covers citations that credit a method or say results agree. Support for a paper retracted for duplication or plagiarism need not spread an error (Schmidt 2024); the label measures use, not harm. The flagged label sees only the passage around the citation; a paper may note the retraction elsewhere.
- Validation. The reference labels used to develop and test the panel, and the second reading of the synthesis cases, were made by a language model (Claude), not by people. The flagged label agrees closely with Hsiao and Schneider's hand-confirmed labels. The cited-as-support label is pending human review: a round of two independent human labellers on a fresh random sample, with its scoring fixed in advance (protocol amendment 4), has not yet been run. On the confirmation sample the panel's rate ran 2.7 points above the reference labels (40.7% vs 38.0%), so we report the corrected rate beside the raw one.
- Retraction Watch reason tags and subject codes are used as recorded. Papers retracted and replaced that also carry an error tag count in the error group, and few citations fall outside the life and health sciences.
- Known error patterns, not yet counted across the census: citing papers that recount a well-known retraction as history are sometimes labelled support rather than passing; a citing paper can attach a reference to the wrong sentence, which the panel cannot detect; and a synthesis that excludes a study for being retracted can be labelled support.

Declarations

Competing interests. The author founded SaneApps LLC, a small software company, and runs SaneCite, an independent public-interest research project that checks how published papers cite retracted work and publishes its counts, code and data. SaneCite sells nothing. In the past 36 months the only third-party support for this work was in-kind: Cloudflare startup-program compute credits (see Funding). Applications for grant support are pending, including one to Anthropic, whose model Claude was used in this work as described under Use of AI; none had been awarded at submission. No journal or publisher has paid the author or SaneApps. The author declares no other competing interests.

Funding. This work received no grant funding. Model inference for the census ran on Cloudflare Workers AI: the Clef and Gemma calls were covered by Cloudflare startup-program credits granted to SaneApps LLC (in-kind compute), and the Jev calls ($1.835 as billed) were paid by the author. Cloudflare had no role in the study design, data collection, analysis, interpretation or writing.

Ethics. The study analyses published articles and public databases (Retraction Watch via Crossref, Europe PMC and the PubMed Central open-access subset). It involves no human participants and no personal data beyond published bibliographic records, so ethics committee approval and consent were not required. Results are reported in aggregate; no individual author, paper or journal is named.

Trial registration. Not applicable (not a clinical trial). The analysis protocol was posted before the final analysis at github.com/MrSaneApps/sanecite-retraction-census (13 October 2026).

Author contributions. Stephan Joseph: conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing (original draft, review and editing), project administration. Use of AI tools is described under Methods, "Use of AI".

Data availability. All data and code are public at github.com/MrSaneApps/sanecite-retraction-census and archived on Zenodo (all versions: doi:10.5281/zenodo.23358099): the protocol and its amendments, the frame, every extracted citing passage, every model label, the validation sets and the result files. Code is under the MIT license and data under CC BY 4.0; quoted passages remain under each source article's license, and Retraction Watch records under the terms of Crossref and Retraction Watch. The scripts that produce the paper's figures, which re-run the registered analysis on the checked frame, are released in the repository's paper/ folder with version 1.3.0 on 29 October 2026; release v1.2.0 (doi:10.5281/zenodo.23539192) gives the same headline counts (data/census_results.json, corrected_wrong_records). Hsiao and Schneider's data are at the Illinois Data Bank (doi:10.13012/B2IDB-8255619_V2). Every citing passage and label can also be searched at sanecite.com/data.html.

Reporting guideline. STROBE for cross-sectional studies; the completed checklist is a supplementary file.

Acknowledgements. We thank Retraction Watch and Crossref for open retraction data, Europe PMC for open full text, and Hsiao and Schneider for publishing their hand-confirmed labels.

References

- Avenell A, Bolland MJ, Gamble GD, Grey A. A randomized trial alerting authors, with or without coauthors or editors, that research they cited in systematic reviews and guidelines has been retracted. Account Res 2024;31(1):14–37. doi:10.1080/08989621.2022.2082290
- Bar-Ilan J, Halevi G. Post retraction citations in context: a case study. Scientometrics 2017;113(1):547–565. doi:10.1007/s11192-017-2242-0
- Budd JM, Sievert M, Schultz TR. Phenomena of retraction: reasons for retraction and citations to the publications. JAMA 1998;280(3):296–297. doi:10.1001/jama.280.3.296
- Crossref. Crossref acquires Retraction Watch data and opens it for the scientific community. 12 September 2023. https://www.crossref.org/blog/news-crossref-and-retraction-watch/

- DeVito NJ, Cunningham C, Goldacre B. Notifying authors that they have cited a retracted article and future citations of retracted articles: the RetractoBot randomized controlled trial. International Congress on Peer Review and Scientific Publication, Chicago, September 2025 (abstract). doi:10.48448/sxnj-0j24
- Fanelli D, Wong J, Moher D. What difference might retractions make? An estimate of the potential epistemic cost of retractions on meta-analyses. Account Res 2022;29(7):442–459. doi:10.1080/08989621.2021.1947810
- Frasco PE, Smith BB, Murray AW, Khurmi N, Mueller JT, Poterack KA. Context analysis of continued citation of retracted manuscripts published in anesthesiology journals. Anesth Analg 2022;135(5):1011–1020. doi:10.1213/ANE.0000000000006195
- Graña Possamai C, Cabanac G, Perrodeau E, Ghosn L, Ravaud P, Boutron I. Inclusion of retracted studies in systematic reviews and meta-analyses of interventions: a systematic review and meta-analysis. JAMA Intern Med 2025;185(6):702–709. doi:10.1001/jamainternmed.2025.0256
- Hsiao T-K, Schneider J. Continued use of retracted papers: temporal trends in citations and (lack of) awareness of retractions shown in citation contexts in biomedicine. Quantitative Science Studies 2021;2(4):1144–1169. doi:10.1162/qss_a_00155. Data: Illinois Data Bank, doi:10.13012/B2IDB-8255619_V2
- Kataoka Y, Banno M, Tsujimoto Y, et al. Retracted randomized controlled trials were cited and not corrected in systematic reviews and clinical practice guidelines. J Clin Epidemiol 2022;150:90–97. doi:10.1016/j.jclinepi.2022.06.015
- Moravcsik MJ, Murugesan P. Some results on the function and quality of citations. Soc Stud Sci 1975;5(1):86–92. doi:10.1177/030631277500500106
- National Information Standards Organization. Communication of Retractions, Removals, and Expressions of Concern (CREC). NISO RP-45-2024. doi:10.3789/niso-rp-45-2024
- Retraction Watch database, distributed by Crossref; downloaded 13 October 2026.
- Rogan WJ, Gladen B. Estimating prevalence from the results of a screening test. Am J Epidemiol 1978;107(1):71–76. doi:10.1093/oxfordjournals.aje.a112510
- Schmidt M. Why do some retracted articles continue to get cited? Scientometrics 2024;129(12):7535–7563. doi:10.1007/s11192-024-05147-4
- von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet 2007;370(9596):1453–1457. doi:10.1016/S0140-6736(07)61602-X
- Weeks J, Cuthbert A, Alfirevic Z. Trustworthiness assessment as an inclusion criterion for systematic reviews: what is the impact on results? Cochrane Evid Synth Methods 2023;1(10):e12037. doi:10.1002/cesm.12037
- Woo S, Walsh JP. On the shoulders of fallen giants: what do references to retracted research tell us about citation behaviors? Quantitative Science Studies 2024;5(1):1–30. doi:10.1162/qss_a_00303
- Xu C, Fan S, Tian Y, et al. Investigating the impact of trial retractions on the healthcare evidence ecosystem (VITALITY Study I): retrospective cohort study. BMJ

2025;389:e082068. doi:10.1136/bmj-2024-082068