

Inaccurate citation, citation cartels, fake citations, citation of retracted papers: why research integrity needs citation integrity

Jodi Schneider (jodi@illinois.edu) & Halil Kilicoglu (halil@illinois.edu)

School of Information Sciences

University of Illinois at Urbana Champaign



OEC talk series
September 21, 2023

Who we are



Jodi Schneider
Associate Professor of Information Sciences
University of Illinois at Urbana Champaign



Halil Kilicoglu
Associate Professor of Information Sciences
University of Illinois at Urbana Champaign

Poll:

- Do you cover citation in RCR training at your institution?
- Have you heard about citation in an RCR training anywhere?

Why research integrity needs citation integrity

- Accurate citation is the bedrock of reliable, trustworthy research.
- Citations are also widely used in measuring research impact.
- But there is little accountability for citations.
- Questionable Research Practices around citation are common – but can be reduced with citation integrity education.

Inaccurate citations can be harmful to our health!

In conclusion, we found that a five-sentence letter published in the *Journal* in 1980 was heavily and uncritically cited as evidence that addiction was rare with long-term opioid therapy. We believe that this citation pattern contributed to the North American opioid crisis by helping to shape a narrative that allayed prescribers' concerns about the risk of addiction associated with long-term opioid therapy. In 2007, the manufacturer of OxyContin and three senior executives pleaded guilty to federal criminal charges that they misled regulators, doctors, and patients about the risk of addiction associated with the drug.⁵ Our findings highlight the potential consequences of inaccurate citation

The NEW ENGLAND JOURNAL of MEDICINE

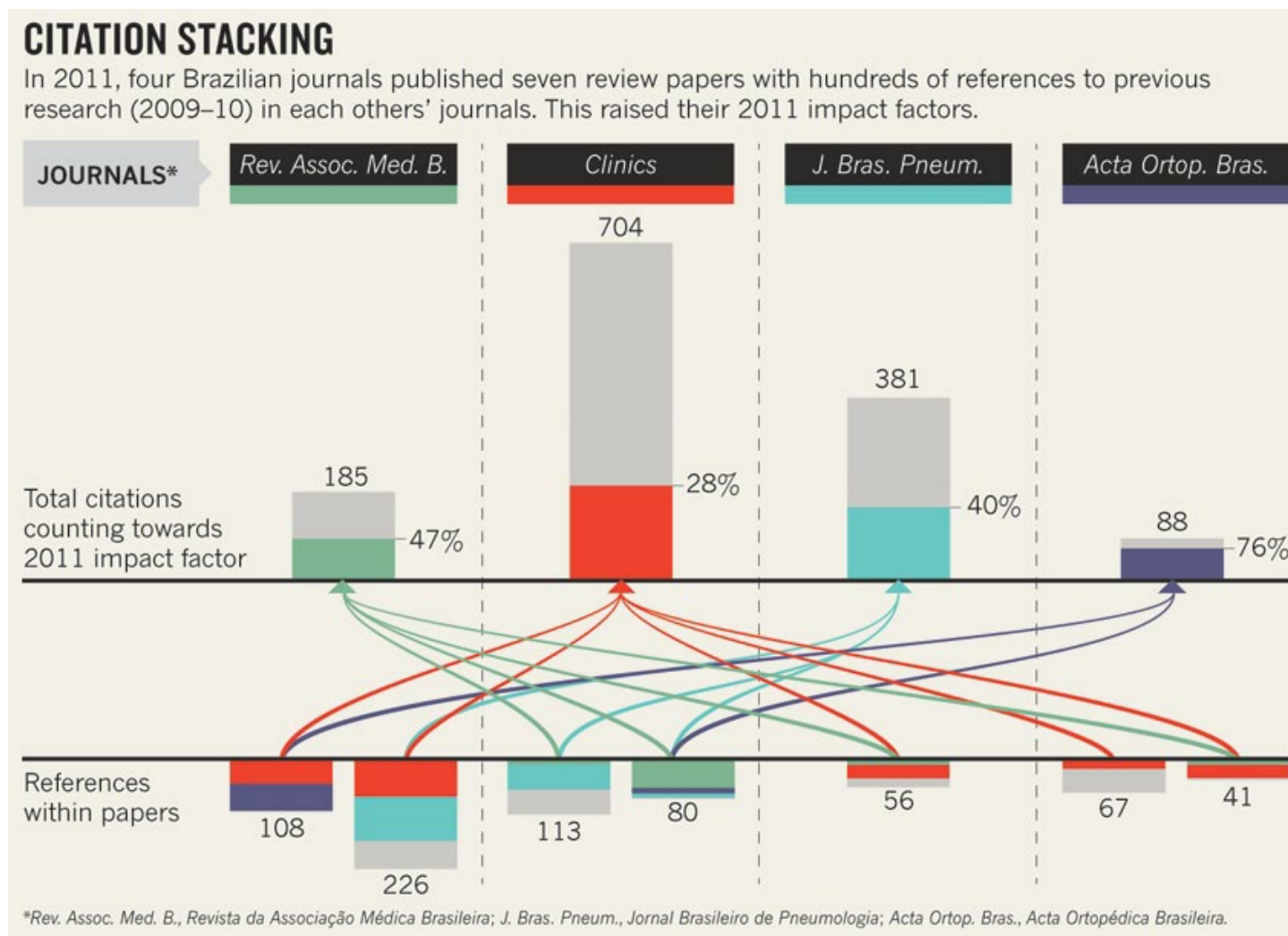
CORRESPONDENCE



A 1980 Letter on the Risk of Opioid Addiction

A 1980 Letter on the Risk of Opioid Addiction
N Engl J Med 2017; 376:2194-2195
DOI: 10.1056/NEJMc1700150

Citation cartels (aka “citation stacking”)



Van Noorden, R. Brazilian citation scheme outed. *Nature* **500**, 510–511 (2013).

<https://doi.org/10.1038/500510a>

Leads to suppression from Journal Citation Reports ([2022 policy](#))

Fake citations

Ways ChatGPT Can Fail

It makes stuff up.

Here is a list of sources on carbon capture provided by ChatGPT.

1. *“Carbon Capture and Storage: An Overview”* by Michael R. Economides and Howard J. Herzog. Published in *Annual Review of Chemical and Biomolecular Engineering*, 2011.
2. *“Carbon capture technology”* by J. David Rogers. Published in *Progress in Energy and Combustion Science*, 2008.
3. *“Carbon capture and storage: An essential technology in the fight against climate change”* by Howard J. Herzog and John Larsen. Published in *Environmental Science & Technology*, 2007.

Each of these sources is fabricated. At a glance, these look like scholarly sources on the topic. Some of these authors exist, some of these journals exist, some of these titles have been used on articles before, but none of these citations represent a published article. In keeping with GPT-3’s pattern, what it produces looks plausible but has no relationship with reality.

Detect fake citations with Edifix

edifix[®]

Vincenzo, B., Caruso, M., Garzia, M., & Licitra, G. (2007). Costochondritis in women taking oral contraceptives. *European Journal of Internal Medicine*, 18(2), 86–88.
<https://doi.org/10.1016/j.ejim.2006.09.001>

Edifix has not updated this reference because the article title in Crossref significantly differs from the author's original. The Crossref reference is Finucane, P., & Kellett, J. (2007). A new direction for medical education in Ireland? *European Journal of Internal Medicine*, 18(2), 101–103. (Ref. "Vincenzo, Caruso, Garzia, Licitra, 2007")

“Dr. OpenAI just made that up”

Fake citation provided by OpenAI, via <https://insidemedicine.substack.com/p/fun-with-openai-medical-charting>

RETRACTION GUIDELINES

Summary

Editors should consider retracting a publication¹ if:

- They have clear evidence that the findings are unreliable, either as a result of major error (eg, miscalculation or experimental error), or as a result of fabrication (eg, of data) or falsification (eg, image manipulation)
- It constitutes plagiarism
- The findings have previously been published elsewhere without proper attribution to previous sources or disclosure to the editor, permission to republish, or justification (ie, cases of redundant publication)
- It contains material or data without authorisation for use
- Copyright has been infringed or there is some other serious legal issue (eg, libel, privacy)
- It reports unethical research
- It has been published solely on the basis of a compromised or manipulated peer review process
- The author(s) failed to disclose a major competing interest (a.k.a. conflict of interest) that, in the view of the editor, would have unduly affected interpretations of the work or recommendations by editors and peer reviewers.

What is retraction?

“Retraction is a mechanism for correcting the literature and alerting readers to articles that contain such seriously flawed or erroneous content or data that their findings and conclusions cannot be relied upon.”


COPE Council. COPE Guidelines: Retraction Guidelines. Version 2, Nov 2019
<https://doi.org/10.24318/cope.2019.1.4>

Inconsistent notification styles make it difficult to find out that publications are retracted.

✓ **SCIENTIFIC REPORTS**


Article | [Open Access](#) | Published: 01 July 2019


RETRACTED ARTICLE: Doubled haploid production in alfalfa (*Medicago sativa* L.) through isolated microspore culture

Dengxia Yi , Jifeng Sun, Yanbin Su, Zongyong Tong, Tiejun Zhang & Zan Wang 

Scientific Reports **9**, Article number: 9458 (2019) | [Cite this article](#)

1470 Accesses | **1** Citations | [Metrics](#)

 This article was [retracted](#) on 13 September 2019

 This article has been updated

▼ **SCIENTIFIC REPORTS**

[Open Access](#) | Published: 21 February 2012

A therapeutic method for the direct reprogramming of human liver cancer cells with only chemicals

Hisashi Moriguchi, Yue Zhang, Makoto Mihara & Chifumi Sato

Scientific Reports **2**, Article number: 280 (2012) | [Cite this article](#)

197 Accesses | **5** Citations | **75** Altmetric | [Metrics](#)

 An [Addendum](#) to this article was published on 18 October 2012

 An [Addendum](#) to this article was published on 18 October 2012

 A [Retraction](#) to this article was published on 09 November 2012

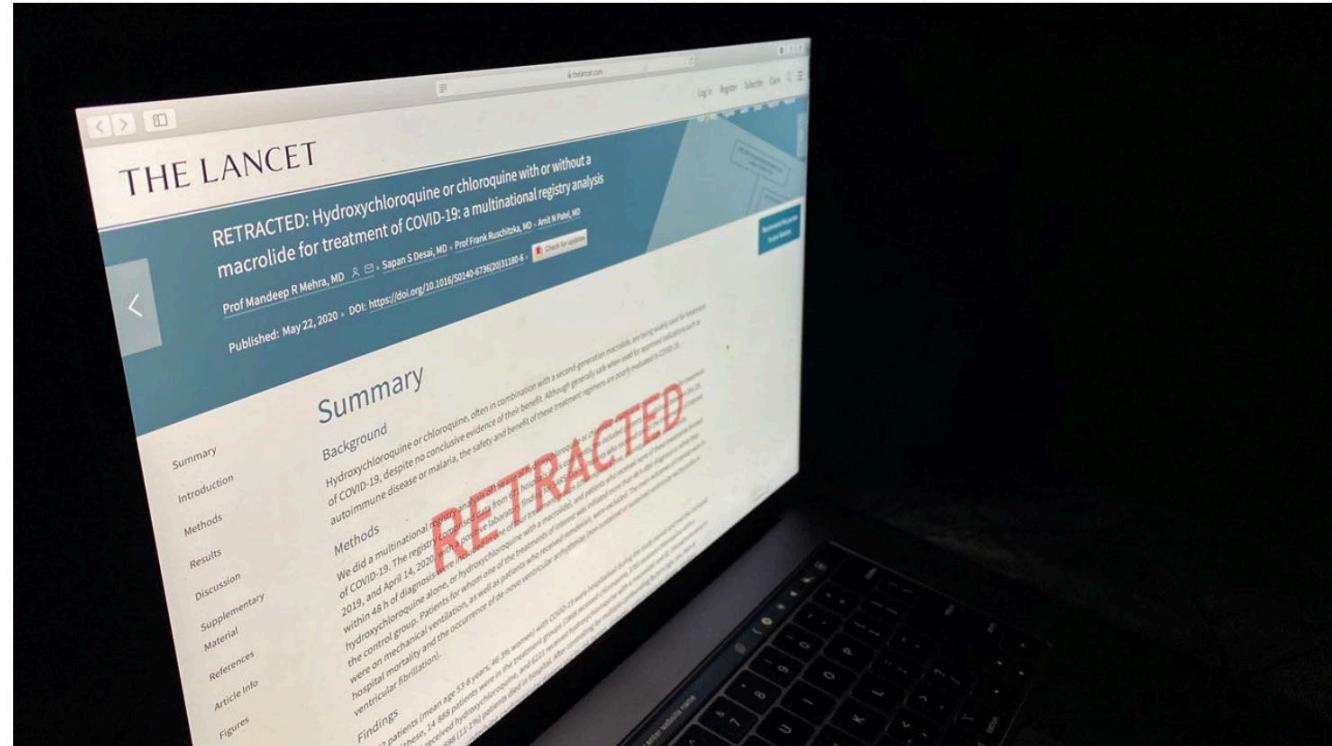
 This article has been updated

Elizabeth M.Suelzer, Jennifer Deal, Karen L. Hanus. Challenges in discovering the retracted status of an article. <http://hdl.handle.net/2142/108367>
See also their article: *JAMA Netw Open*. 2021;4(6):e2115648 <http://doi.org/10.1001/jamanetworkopen.2021.15648>

Continued citation of retracted papers

Two COVID-19 articles retracted a month after they were published have over 1200 citations each.

Science magazine examined 200 of the post-retraction citations to these papers and concluded that over half inappropriately cited the retracted articles.



E. PETERSEN/SCIENCE

Many scientists citing two scandalous COVID-19 papers ignore their retractions

By **Charles Pillar** | Jan. 15, 2021, 8:00 AM

<https://doi.org/10.1126/science.abg5806>

Why research integrity needs citation integrity

- Accurate citation is the bedrock of reliable, trustworthy research.
- Citations are also widely used in measuring research impact.
- But there is little accountability for citations.
- Questionable Research Practices around citation are common – but can be reduced with citation integrity education.

What we're doing about citation integrity in our ongoing research projects

- Office of Research Integrity-funded project: Natural Language Processing to Assess and Improve Citation Integrity in Biomedical Publications, ORIIR220073
- Alfred P. Sloan-funded project: Reducing the Inadvertent Spread of Retracted Science, G-2022-19409

Citation accuracy and integrity in biomedicine

- Citations are rarely examined for accuracy in the peer review process
 - Metadata errors
 - Citation content errors (quotation errors)



Quotation errors may be harmful

- Difficult to detect for readers, journals, and peer reviewers
- Estimated ~25% of medical articles contain quotation errors, half of them severe (Jergas and Baethge, 2015)
- Inaccurate citations to a letter published in the NEJM in 1980 may have contributed to the opioid crisis (Leung et al., 2017)
- Citation distortions and biases have led to unfounded claims to be accepted as beliefs in Alzheimer's disease research (Greenberg, 2009)



Quotation error

- Citation: *This is coherent with the fact that hACE2 expression were not observed in the gut of the mice used in that study [37].*
- Reference: [37] *In the gastrointestinal tract of K18-hACE2 mice, hACE2 was expressed most abundantly in the colon, which correlated with infection seen at later time points.*



CONTRADICTION

Citation invention via quotation error

Conversion of hypothesis to fact with citation alone (Greenberg, 2009)



[1] *the accumulation of intracellular A β in muscle fibers may be an initiating event in IBM...*

[2] *It is of particular interest that in IBM, abnormal accumulation of β APP and β APP-mRNA seems to precede other abnormalities...*

[3] *We have previously demonstrated that accumulation of A β PP epitopes precedes other abnormalities in IBM muscle fibers^{1,2}*

**IBM*: Inclusion body myositis

* *β APP*: Beta-amyloid precursor protein

Multi-citation

- *Recent studies proved that PPARs functions as a key regulator for adipocyte development and IMF regulation [9, 24, 25].*
- *[24] PPAR γ is a major driver of the accumulation and phenotype of adipose-tissue Treg cells*

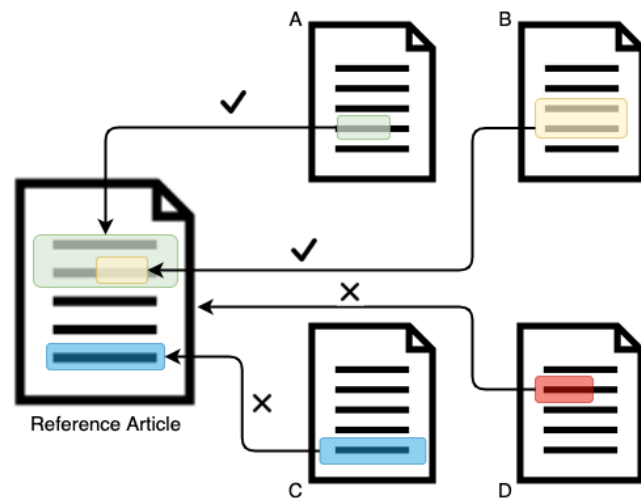
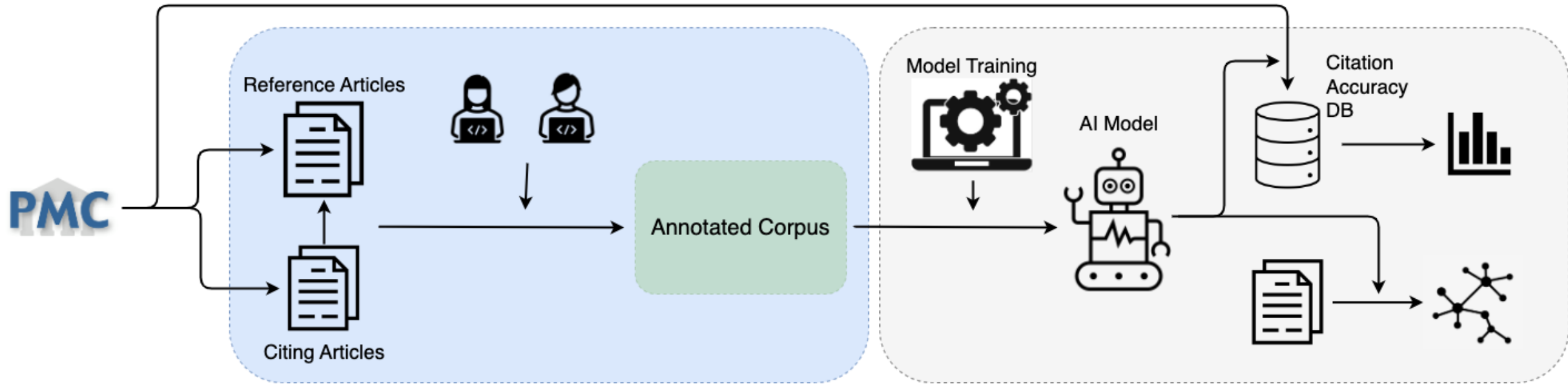


Natural Language Processing for citation accuracy

- Assessing citations for accuracy requires considerable manual effort
 - Needs to be done pre-publication
- Natural Language Processing techniques could support tools for citation accuracy checking
 - A type of Artificial Intelligence (AI)
 - Flag problematic citations for closer scrutiny
 - Trace the provenance of misleading claims/misinformation
- Automated tools can
 - Support journal workflows
 - Raise awareness about citation errors
 - Support meta-research



Our work

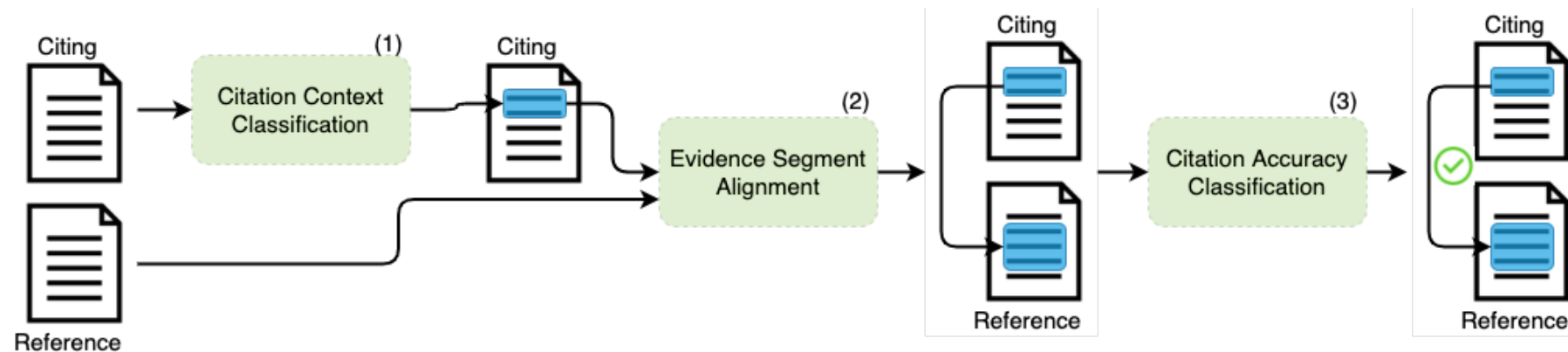


Citation accuracy categories

- ACCURATE
- Major errors
 - NOT_SUBSTANTIATED, IRRELEVANT, CONTRADICT
- Minor errors
 - ETIQUETTE, OVERSIMPLIFY, INDIRECT, MISQUOTE
- 3105 instances, 1251 errors (~41%)
- 563 major, 688 minor



Natural Language Processing models



- Citation context classification: sentence classification (0.98 F_1)
- Evidence segment alignment (0.55 F_1)
- Citation accuracy classification (accurate/inaccurate/not enough info ; 0.69 macro- F_1)
- Development ongoing

Continued citation for 11+ years after retraction



CHEST

A human trial article published in 2005, retracted in October 2008 because an author falsified data, has been cited more AFTER retraction than before.

My team examined the post-retraction citations, 2009-2019, and concluded that 96% inappropriately cited the retracted article.

RETRACTION

The article “Effects of Omega-3 Polyunsaturated Fatty Acids on Inflammatory Markers in COPD” (Matsuyama W, Mitsuyama H, Watanabe M, et al; Chest 2005; 128:3817–3827) in the December 2005 issue is being retracted by CHEST because the university that employs the authors determined that one of the authors, Wataru Matsuyama (now deceased), falsified data.

Schneider, Jodi, Di Ye, Alison M Hill, Ashley S Whitehorn. 2020. “Continued post-retraction citation of a fraudulent clinical trial report, eleven years after it was retracted for falsifying data.” In *Scientometrics*. <http://doi.org/10.1007/s11192-020-03631-1>



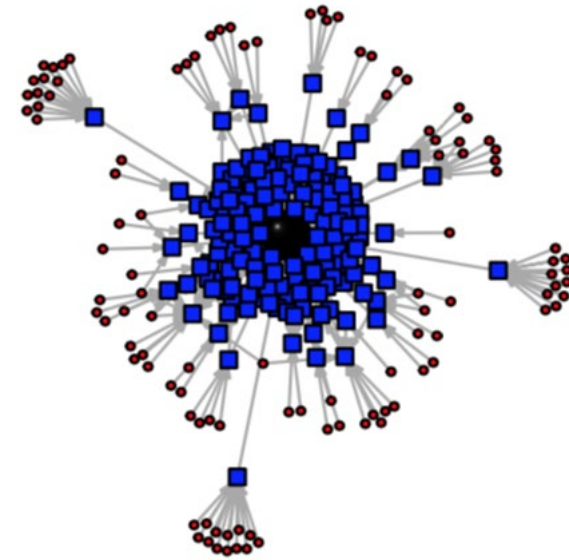
Jodi Schneider
jodi@illinois.edu

Continued citation for 11+ years after retraction

A human trial article published in 2005, retracted in October 2008 because an author falsified data, has been cited more AFTER retraction than before.

My team examined the post-retraction citations, 2009-2019, and concluded that 96% inappropriately cited the retracted article.

2019
Only 5 mention the retraction



148 direct citations

108 indirect published 2019

Schneider, Jodi, Di Ye, Alison M Hill, Ashley S Whitehorn. 2020. "Continued post-retraction citation of a fraudulent clinical trial report, eleven years after it was retracted for falsifying data." In *Scientometrics*. <http://doi.org/10.1007/s11192-020-03631-1>

Continued citation is not rare

My team examined the post-retraction citations in PubMed Central to 7813 retracted papers in PubMed as of August 2020 and concluded that 94% inappropriately cited the retracted article.

Tzu-Kun Hsiao & Jodi Schneider. Continued Use of Retracted Papers- Temporal Trends in Citations and (Lack of) Awareness of Retractions Shown in Citation Contexts in Biomedicine. In *Quantitative Science Studies* https://doi.org/10.1162/qss_a_00155



Jodi Schneider
jodi@illinois.edu

Continued citation is not rare

My team examined the post-retraction citations in PubMed Central to 7813 retracted papers in PubMed as of August 2020 and concluded that 94% inappropriately cited the retracted article.

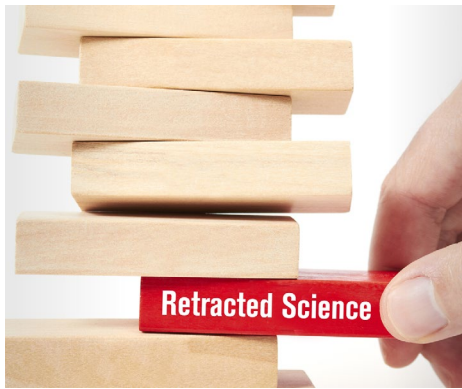
No awareness that these aren't "normal" citations:

- A clinico-histopathologic study in rabbits confirmed that PRP treatment can achieve a faster wound healing rate [retracted cite].
- However, to date, only one human study has demonstrated an induction of SIRT1 mRNA level in PBMCs [retracted cite].

Tzu-Kun Hsiao & Jodi Schneider. Continued Use of Retracted Papers- Temporal Trends in Citations and (Lack of) Awareness of Retractions Shown in Citation Contexts in Biomedicine. In *Quantitative Science Studies* https://doi.org/10.1162/qss_a_00155



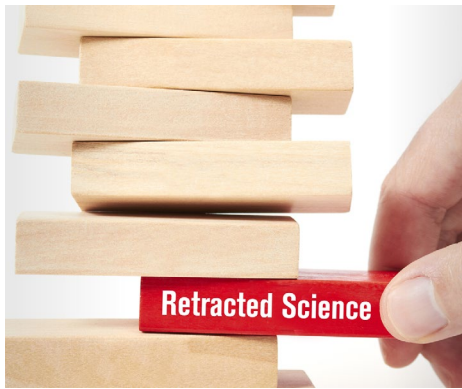
Jodi Schneider
jodi@illinois.edu



RISRS2020 recommendations

<http://doi.org/10.31222/osf.io/ms579>

1. Develop a systematic cross-industry approach to ensure the **public availability of consistent, standardized, interoperable, and timely information about retractions.**
2. Recommend a **taxonomy of retraction categories/classifications and corresponding retraction metadata** that can be adopted by all stakeholders.
3. Develop **best practices for coordinating the retraction process** to enable timely, fair, unbiased outcomes.
4. **Educate stakeholders** about publication correction processes including retraction and about pre- and post-publication stewardship of the scholarly record.



RISRS2020: Reducing the Inadvertent Spread of Retracted Science

<https://infoqualitylab.org/projects/risrs2020/>

Collaboration across diverse stakeholders: *funders, editors, peer reviewers, authors, publishers, database providers, research integrity officers, science journalists*

Annette Flanagin
RN, MA, FAAN



JAMA

Daniele Fanelli, PhD



London School of
Economics and Political
Science

C.K. Gunsalus, JD



National Center
for Principled
Leadership &
Research Ethics

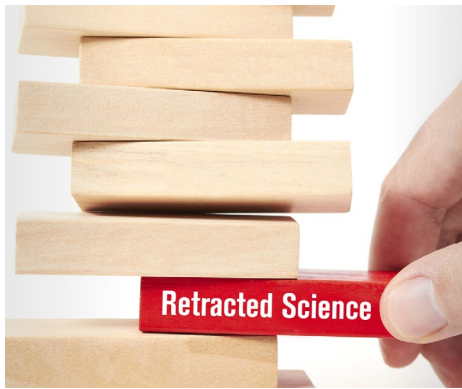
Ivan Oransky, MD



Retraction Watch



Jodi Schneider
jodi@illinois.edu



RISRS2020: Reducing the Inadvertent Spread of Retracted Science

A stakeholder consultation & environment scan

- 47 interviews of stakeholders in the scientific publishing ecosystem
- ~70 participants in a 3-part online workshop to interact, react, and reflect in real-time on the problem retractions pose and possible solutions
- Literature review of empirical research about retraction
- Citation analysis of retracted research

Reducing the Inadvertent Spread of Retracted Science
<https://infoqualitylab.org/projects/risrs2020/>



NISO Communication of Retractions, Removals, and Expressions of Concern (CREC) Working Group

The draft standard will be released for public comment in October 2023

- Additions to existing **metadata** to support awareness of retractions
- Best practices on population of retraction notices
- Outlining responsibilities of creators and consumers of retraction information
- Consistent **display standards**: labels, user experience, signaling
- Workflow definition for circulating information



Jodi Schneider
jodi@illinois.edu

Why research integrity needs citation integrity

- Accurate citation is the bedrock of reliable, trustworthy research.
- Citations are also widely used in measuring research impact.
- But there is little accountability for citations.
- Questionable Research Practices around citation are common – but can be reduced with citation integrity education.

Further reading: Fraud by Numbers: Metrics and the New Academic Misconduct

- “regardless of whether citation-based indexes are a paragon of objectivity or mere numerical simulacra of imaginary value, they have an uncanny ability to spawn **new and expanding forms of research misconduct**”
- “**Citations** have now become tokens of value or impact that, like other data in the age of its monetization, **can be effectively bought and sold**, turned into indexes, repackaged and resold to new users for uses far removed from the evaluation of a specific publication or author.”



HOME > ESSAYS

Fraud by Numbers: Metrics and the New Academic Misconduct

September 7, 2020 • By Mario Biagioli



<https://lareviewofbooks.org/article/fraud-by-numbers-metrics-and-the-new-academic-misconduct/>



Further reading: *Gaming the Metrics: Misconduct and Manipulation in Academic Research*

GAMING THE METRICS

Misconduct and Manipulation
in Academic Research

EDITED BY Mario Biagioli AND Alexandra Lippman

Open Access Book online:

<https://doi.org/10.7551/mitpress/11087.001.0001>



Further reading: Diversity in citation practices

<https://www.universityaffairs.ca/career-advice/ask-dr-editor/diversity-in-citation-practices-auditing-your-list-of-references-contributes-to-better-science/>

Home

[Ask Dr. Editor](#)

Diversity in citation practices: auditing your list of references contributes to better science

As a researcher, you should intentionally consider voices and perspectives that you might have previously omitted or ignored.

BY LETITIA HENVILLE | DEC 09 2022



Post a comment



Question:

I cite the best and most relevant works for my research. So why is my editor suggesting that I diversify the race and gender of the scholars that I cite? That seems to me like politics interfering with science.

– Anonymous, Mechanical Engineering

Dr. Editor's response:

It'll be no surprise to those familiar with [my previous articles](#) that I see science as a human endeavour and therefore as inherently political. When you review, consider, and seek to diversify who you cite, you're not bringing politics into science. Instead, you're recognizing the politics that are already there — notably, the systemic barriers that are keeping you from locating and considering good work done by women, BIPOC scientists, and other folks who are disproportionately marginalized in your field.

Example best practices for citation

- Cite as clearly and specifically as possible. **Be cautious with multicitation:**
 - ❌ lions, tigers and, bears (Hu, 2000; Lee, 2005; Majidi, 1998).
 - ✅ ...lions (Hu, 2000), tigers (Majidi, 1998), and bears (Lee, 2005).
- Cite the best available evidence
 - ❌ Be cautious in citing narrative reviews, opinion pieces, and abstracts.
 - ✅ Cite the strongest research available (best design, most informative, etc.)
 - ✅ Cite original research
- ✅ Use software to manage references
- Before finalizing a paper
 - ✅ Check for the most recent version of preprints, in press, and informal citations
 - ✅ Verify statements with the original source

Pavlovic, V., et al. (2021). How accurate are citations of frequently cited papers in biomedical literature?. *Clinical Science*, 135(5), 671-681. <https://doi.org/10.1042/CS20201573>

Example best practices for citation

“Confirm references are legitimate

- Check whether cited works have been corrected or retracted
- Ensure referenced material is relevant and avoid superfluous cited references
- Maintain a level of self-citation that is appropriate to the field or discipline”

Szomszor, M., & Quaderi, N. (2020). Research Integrity: Understanding our shared responsibility for a sustainable scholarly ecosystem. *Institute for Scientific Information*.

What all RIOs and RCR instructors can do to improve citation integrity

1. Learn best practices for citation and why they matter.
2. Train researchers in best practices for citation.
3. Support PIs in creating “a clear citation management plan so that trainees are learning best practices” (Pavlovic et al., 2021).
4. Develop examples of “known problems with chains of inaccurate citations” (Pavlovic et al., 2021).
5. Discuss citation inaccuracies in RCR training (Pavlovic et al., 2021).
6. Test and use automated tools!

Pavlovic, V., et al. (2021). How accurate are citations of frequently cited papers in biomedical literature?. *Clinical Science*, 135(5), 671-681. <https://doi.org/10.1042/CS20201573>