Questionable papers in citation databases as an issue for literature review

Mehdi Dadkhah¹ · Mohammad Lagzian¹ · Glenn Borchardt²

Received: 14 November 2016 / Accepted: 30 November 2016 / Published online: 18 February 2017
© The International CCN Society 2017

Abstract In recent years, the academic world has been faced with much academic misconduct. Examples involve plagiarizing papers, manipulating data, and launching predatory or hijacked journals. The literature exposing these activities is growing exponentially, and so is the presentation of criteria or guidelines for counteracting the problem. Most of the research is focused on predatory or hijacked journal detection and providing suitable warnings. Overlooked in all this is the fact that papers published in these journals are questionable, but nevertheless show up in standard citation databases. We need some way to flag them so future researchers will be aware of their questionable nature and prevent their use in literature review.

Keywords Predatory journals · Hijacked journals · Citation databases · Questionable papers · Academic ethics · Literature review

Introduction

Nowadays, academic search engines and citation databases such as Google Scholar, Scopus, and Thomson Reuters Web of Science (WoS) are popular among researchers who hope to

find suitable references for their research efforts by doing a literature review (Osipov et al. 2014). Scopus and WoS are accessible through payment or subscription, but Google Scholar is free for all researchers, so it is a good candidate for an initial literature search. Scopus and WoS only cover selected peer-reviewed journals, but Google Scholar also includes nontraditional sources such as Webpages and self-published articles (Kulkarni et al. 2009). In other words, WoS and Scopus use defined criteria for selecting journals to cover their contents (De Moya-Anegón et al. 2007), but Google Scholar does not (Harzing and Van der Wal 2008). Google merely uses its usual web technologies to find what appears to be academic content.

As most researchers use academic search engines and citation databases to find papers on prior work to complete their literature review, it is important that such databases contain adequately peer-reviewed content. Otherwise, researchers might establish their new research based on bogus and manipulated papers. In support, they may inadvertently use references based on invalid “facts.” Scientific databases should include papers with varying, even opposed interpretations, but the data used for those interpretations should be beyond reproach.

Questionable papers in predatory and hijacked journals

Usually, we consider papers published in predatory and hijacked journals as questionable. Predatory journals were first uncovered by Jeffrey Beall, a librarian at the University of Colorado, Denver, in 2009. He archived each suspicious “call for papers” email sent to him, took a picture of each

✉ Mehdi Dadkhah
Mehdidadkhah@mail.um.ac.ir

¹ Department of Management, Faculty of Economics and Administrative Sciences, Ferdowsi University of Mashhad, Mashhad, Iran

² Progressive Science Institute, Box 5335, Berkeley, CA 94705, USA

journal website pitching the call, and then coined the term “predatory publisher” in 2010. His first list had 20 publishers that he considered to be predatory. He eventually created a website with two continually updated lists, one for predatory publishers, and another for standalone predatory journals (Beall 2013). Predatory journals have a primary distinguishing characteristic: they only exist to make money. They do not have adequate peer review and will publish almost anything—for a price. They typically exploit the open access model that is now becoming popular among legitimate publishers, but do not have the quality and validation to back it up (Bhad and Hazari 2015; Nahai 2015). In response, investigators, following Beall, have highlighted their negative effect on science and have presented guidelines to improve the awareness of prospective authors (Hansoti et al. 2016; Tin et al. 2014; Beall 2016).

Hijacked journals are another problem for scholarly publishing. Hijacked journals are websites that mimic authentic journals by using similar names and ISSN numbers (Jalalian and Dadkhah 2015; Dadkhah and Borchardt 2016). Like predatory journals, these typically exploit the open access model—again, for a price. These journals do not have peer review. The first hijacked journal was created on August 11, 2011. Cybercriminals registered the domain “sciencerecord.com” and launched three hijacked journals including: “Science Series Data Report,” “Innova Ciencia,” and “Science and Nature” (Jalalian and Mahboobi 2014). Mehrdad Jalalian is the first researcher to detect systematic journal hijacking and presented the term “hijacked journal” to the academic world in early 2012 (Jalalian and Dadkhah 2015). Nevertheless, our observations show that there is not sufficient research about hijacked journals in comparison to predatory journals. When we search for the term “hijacked journal” in Google Scholar, 38 results are reachable, but the term “predatory journal” yields 161 results (searched on 15 September 2016).

The newest twist in academic publishing involves computer generated papers and do not present real knowledge. SCIGen (<http://pdos.csail.mit.edu/scigen/#generate>) is a tool developed by the Massachusetts Institute of Technology to test peer review at low quality conferences. It has been used by some researchers to create fake papers in the attempt to publish them in conferences and academic journals (Bohannon 2015a, b). Bohannon found that in 2013, 85 fake papers were accepted at 24 different conferences (Bohannon 2015a). Some legitimate journals also do not provide essential peer review, with fake papers having a good chance of being published. As a test, Bohannon sent a fake paper to 304 selected journals. It was accepted by 157 of them (Bohannon 2013).

These questionable papers may be used by researchers for doing literature review or designing hypotheses. It is clear that using such papers will lead to unreliable and invalid science. Researchers need to know the current situation in their disciplines. They do not need to follow bogus information.

Existence of questionable papers in citation databases

As mentioned, citation databases must not include questionable papers. In particular, they should not include papers published by predatory or hijacked journals. Usually citation databases and academic search engines use advanced techniques to select content (such as journals and proceedings), but our observations show that many questionable papers and journals are included nonetheless.

Predatory journals and citation databases

Google Scholar’s aim is to cover comprehensive journals regardless of their quality and makes no effort to exclude papers published in predatory journals (Beall 2016). The publishers of most predatory journals seem to have no qualms about announcing their being indexed in Google Scholar. Thus, Google Scholar is worthless for detecting questionable papers published in predatory journals. As mentioned, WoS and Scopus use special criteria to select high quality journals for indexing (De Moya-Anegón et al. 2007), although some predatory journals slip through. For example, we searched a custom query in Scopus (Appendix 1) to find the top journals under the phrase “most published papers in Asia.” The fifteenth journal is the International Journal of Applied Engineering Research with 11,002 papers that were indexed between 2010 and 2016. Beall and some others believe that it has a predatory publisher (<https://scholarlyoa.com/publishers>) (Dyrud 2014), so Scopus includes papers published in predatory journals too. Recently, Beall wrote about the Thomson Reuters master list and announced that this list also contains many predatory journals (<https://scholarlyoa.com/2016/04/28/the-tr-master-journal-list-is-not-a-journal-whitelist/#more-7354>). As mentioned in a previous section, Beall believes that predatory journals have flawed peer review and are a source of questionable papers. They should not be included in citation databases of reviewed journals.

Hijacked journals and citation databases

As mentioned before, hijacked journals are fake websites made to look like those of reputable journals. As a result, papers published in association with these websites should not be used by researchers to conduct new research. Normally, citation databases do not index papers published in hijacked journals, but cybercriminals have found ways to include their hijacked journals’ papers in academic search engines and citation databases.

As a test, we used Publish or Perish tools (<http://www.harzing.com/resources/publish-or-perish>) to uncover some papers published in hijacked journals and subsequently indexed in Google Scholar. We used Beall’s updated list

Table 1 Number of indexed papers for each hijacked journal in Google Scholar (2014–2016)

No.	Journal name	Number of indexed papers for each hijacked journal
1	Academie royale des sciences d outre-mer bulletin des seances Authentic URL: http://www.kaowarsom.be/fr/bulletins Hijacked URL: http://ardsj.com	46
2	Ayer Authentic URL: http://www.ahistcon.org/revistaayer.html Hijacked URL: http://ayeronline.com	218
3	Bulletin of the Georgian National Academy of Sciences Authentic URL: http://www.science.org.ge/moambe/moambe.html Hijacked URL: http://www.bnas.org	3
4	JNCC REPORT Authentic URL: http://jncc.defra.gov.uk/page-2132 Hijacked URL: http://jncc.info/ojs-2.4.6/index.php/jncc/about	21
5	MAGNT Research Report Authentic URL: http://www.worldcat.org/oclc/223764115 Hijacked URL: http://brisjast.com/	20
6	MULTITEMAS Authentic URL: http://site.ucdb.br/editora-ucdb/tematica/37/periodicos/multitemas-numero-2-dez-1996/rmt/ Hijacked URL: http://multitemas.com/index.php/MTJournal	181
7	Walia Journal Authentic URL: http://www.worldcat.org/oclc/183351204 Hijacked URL: http://www.waliaj.com/	16

of hijacked journals (<https://scholarlyoa.com/other-pages/hijacked-journals/>) to compile Table 1.

We expect that popular citation databases such as Scopus and Thomson Reuters Web of Science will avoid indexing hijacked journals in the future. Nonetheless, we suspect that cybercriminals eventually will find a way to hack these scientific databases too. In November 2015, Bohannon (2015b) published a paper in Science disclosing that

cybercriminals had hijacked 20 journals. These hijacked journals are accessible in Thomson Reuters master list along with their hijacked URLs. Researchers visiting the Thomson Reuters master list were victimized when they inadvertently entered the URLs of these hijacked journals instead of the URLs of the authentic journals. Again, we must warn that citation databases and academic search engines may include questionable papers from hijacked as well as predatory journals.

Fig. 1 Conceptual Google Scholar search showing special flags for questionable and authentic papers

Fake papers and citation databases

As mentioned, computer-generated and other kinds of fake papers have been published in academic proceedings and journals. Most of these were nonsensical, but got through the submission process due to low quality or nonexistent review.

We planned a small experiment by using SCIdgen to generate 10 fake papers. Then, following Google Scholar instructions for indexing academic papers (<https://scholar.google.com/intl/en/scholar/inclusion.html>), we were able to index our own fake papers to be included in Google's search results. Others did a similar experiment by creating fake citations for selected papers (Lopez-Cozar et al. 2012). The main problem is this: Google crawls web contents and looks for PDF files that seem to be academic documents. According to Google's instructions for indexing a document, each PDF file that is linked to a web page with a title, author name, and references, is a candidate for indexing in Google Scholar.

Some general guidelines for authors

As mentioned in previous sections, many questionable papers are indexed in citation databases and academic search engines. Obviously, authors should not cite questionable papers in their research. The use of such citations tends to invalidate their own research. Authors must use only legitimate journals in their work. By using Beall's lists for predatory and hijacked journals, researchers will be able to detect the most questionable papers. In the future, we hope that citation databases will flag questionable papers. Figure 1 shows the conceptual image of such a citation database. In this figure, the first paper has been retracted, so it easily classifies as a questionable paper. The other results come from reputable journals that are not predatory or have not been hijacked. A "trusted sign" has been shown for them. At present, citation databases only flag retracted papers that have flaws in terms of reliability.

Conclusion

Nowadays, scholarly publishing is faced with much misconduct leading to questionable papers throughout the academic literature. In this paper, we reveal the most common types of questionable papers and show that citation databases and academic search engines are not immune to including them in their indices. Researchers can avoid their use of such bogus references by checking lists of known predatory and hijacked journals. Also, a new duty now arises for citation databases and academic search engines: they need to find solutions for detecting questionable papers, thus preventing their spread throughout academic literature without their true nature being known to prospective users. In some research methods, such

as meta-analysis and meta-synthesis, the quality of selected sources is important. Research that exposes fraudulent papers obviously is critical for such assessment.

Appendix 1.

Custom searched query in Scopus for finding top journals by using the phrase "most published papers in Asia"

```
( AFFILCOUNTRY ( Afghanistan ) OR
AFFILCOUNTRY ( Armenia ) OR
AFFILCOUNTRY ( Azerbaijan ) OR
AFFILCOUNTRY ( Bahrain ) OR
AFFILCOUNTRY ( Bangladesh ) OR
AFFILCOUNTRY ( Bhutan ) OR AFFILCOUNTRY ( Brunei )
OR AFFILCOUNTRY ( Myanmar ) OR
AFFILCOUNTRY ( Cambodia ) OR
AFFILCOUNTRY ( China ) OR AFFILCOUNTRY ( India ) OR
AFFILCOUNTRY ( Indonesia ) OR AFFILCOUNTRY ( Iran )
OR AFFILCOUNTRY ( Iraq ) OR AFFILCOUNTRY ( Japan )
OR AFFILCOUNTRY ( Jordan ) OR
AFFILCOUNTRY ( Kazakhstan ) OR
AFFILCOUNTRY ( North Korea ) OR
AFFILCOUNTRY ( South Korea ) OR
AFFILCOUNTRY ( Kuwait ) OR
AFFILCOUNTRY ( Kyrgyzstan ) OR
AFFILCOUNTRY ( Laos ) OR AFFILCOUNTRY ( Lebanon )
OR AFFILCOUNTRY ( Malaysia ) OR
AFFILCOUNTRY ( Maldives ) OR
AFFILCOUNTRY ( Mongolia ) OR AFFILCOUNTRY ( Nepal )
OR AFFILCOUNTRY ( Oman ) OR
AFFILCOUNTRY ( Pakistan ) OR
AFFILCOUNTRY ( Philippines ) OR
AFFILCOUNTRY ( Qatar ) OR AFFILCOUNTRY ( Saudi
Arabia ) OR AFFILCOUNTRY ( Singapore ) OR
AFFILCOUNTRY ( Sri Lanka ) OR AFFILCOUNTRY ( Syria )
OR AFFILCOUNTRY ( Taiwan ) OR
AFFILCOUNTRY ( Tajikistan ) OR
AFFILCOUNTRY ( Thailand ) OR AFFILCOUNTRY ( Timor-
Leste ) OR AFFILCOUNTRY ( Turkmenistan ) OR
AFFILCOUNTRY ( United Arab Emirates ) OR
AFFILCOUNTRY ( Uzbekistan ) OR
AFFILCOUNTRY ( Vietnam ) OR AFFILCOUNTRY ( Yemen )
OR AFFILCOUNTRY ( Israel ) ) AND PUBYEAR AFT 2010.
```

References

- Beall J (2013) Medical publishing triage—chronicling predatory open access publishers. *Ann Med Surg* 2(2):47–49. doi:10.1016/S2049-0801(13)70035-9

- Beall J (2016) Dangerous predatory publishers threaten medical research. *J Korean Med Sci* 31(10):1511–1513. doi:[10.3346/jkms.2016.31.10.1511](https://doi.org/10.3346/jkms.2016.31.10.1511)
- Bhad R, Hazari N (2015) Predatory journals in psychiatry: a note of caution. *Asian J Psychiatr* 16:67–68. doi:[10.1016/j.ajp.2015.06.008](https://doi.org/10.1016/j.ajp.2015.06.008)
- Bohannon J (2013) Who's afraid of peer review. *Science* 342(6154):60–65. doi:[10.1126/science.342.6154.60](https://doi.org/10.1126/science.342.6154.60)
- Bohannon J (2015a) Hoax-detecting software spots fake papers. *Science* 348(6230):18–19. doi:[10.1126/science.348.6230.18](https://doi.org/10.1126/science.348.6230.18)
- Bohannon J (2015b) How to hijack a journal. *Science* 350(6263):903–905. doi:[10.1126/science.350.6263.903](https://doi.org/10.1126/science.350.6263.903)
- Dadkhah M, Borchardt G (2016) Hijacked journals: an emerging challenge for scholarly publishing. *Aesthet Surg J* 36(6):739–741. doi:[10.1093/asj/sjw026](https://doi.org/10.1093/asj/sjw026)
- De Moya-Anegón F, Chinchilla-Rodríguez Z, Vargas-Quesada B, Corera-Álvarez E, Muñoz-Fernández F, González-Molina A, Herrero-Solana V (2007) Coverage analysis of Scopus: a journal metric approach. *Scientometrics* 73(1):53–78. doi:[10.1007/s11192-007-1681-4](https://doi.org/10.1007/s11192-007-1681-4)
- Dyrd AM (2014). Predatory online technical journals: A question of ethics. In 121st ASEE Annual Conference & Exposition, Indianapolis, IN, U.S.A., June 15–18, 2014
- Hansoti B, Langdorf MI, Murphy LS (2016) Discriminating between legitimate and predatory open Access Journals: report from the International Federation for Emergency Medicine Research Committee. *West J Emerg Med* 17(5):497. doi:[10.5811/westjem.2016.7.30328](https://doi.org/10.5811/westjem.2016.7.30328)
- Harzing AWK, Van der Wal R (2008) Google scholar as a new source for citation analysis. *Ethics in science and environmental politics* 8(1): 61–73
- Jalalian M, Dadkhah M (2015) The full story of 90 hijacked journals from august 2011 to June 2015. *Geographica Pannonica* 19(2):73–87
- Jalalian M, Mahboobi H (2014) Hijacked journals and predatory publishers: is there a need to re-think how to assess the quality of academic research? *Walailak J Sci & Tech* 11(5):389–394. doi:[10.14456/WJST.2014.16](https://doi.org/10.14456/WJST.2014.16)
- Kulkarni AV, Aziz B, Shams I, Busse JW (2009) Comparisons of citations in web of science, Scopus, and Google scholar for articles published in general medical journals. *JAMA* 302(10): 1092–1096. doi:[10.1001/jama.2009.1307](https://doi.org/10.1001/jama.2009.1307)
- Lopez-Cozar ED, Robinson-Garcia N, Torres-Salinas D (2012) Manipulating Google Scholar citations and Google Scholar metrics: Simple, easy and tempting. *arXiv preprint arXiv: 1212.0638*
- Nahai F (2015) The rise of predatory journals: what difference does it make? *Aesthet Surg J* 35(8):1042–1043. doi:[10.1093/asj/sjv085](https://doi.org/10.1093/asj/sjv085)
- Osipov G, Smirnov I, Tikhomirov I, Sochenkov I, Shelmanov A, Shvets A (2014) Information retrieval for R&D support. *Lect Notes Comput Sci* 8830:45–69. doi:[10.1007/978-3-319-12511-4_4](https://doi.org/10.1007/978-3-319-12511-4_4)
- Tin L, Ivana B, Biljana B, Ljubica IB, Dragan M, Dušan S (2014) Predatory and fake scientific journals/publishers—a global outbreak with rising trend: a review. *Geographica Pannonica* 18(3):69–81