

Project Proposal

Project Title: Impact of retracted journal articles: online COVID-19 misinformation

Author: Andy Hu

Date: 12/11/2020

Project Definition:

The overall aim of this project is to demonstrate if retracted scientific journal articles on COVID-19 have an impact on discourse on social media as well as if that impact is positive or negative. To test this, a few chosen scientific databases will be searched to create a list of retracted articles relating to COVID-19. A website called "RetractionWatch" is also very useful as the creators of that website have already compiled a decently sized list. These articles will then be evaluated based on eligibility for inclusion in this study. One criteria to consider are things like reason for retraction, as plagiarized articles are not necessarily going to be spreading disinformation. Following this, Altmetric will be used to obtain information regarding online citation information of these retracted articles. Then, a sentiment analysis program returning positive/negative values will be used to determine the context in which these articles are cited. A positive sentiment analysis score will indicate the citation was in favor/supporting the cited retracted article. A negative will indicate the opposite. The same process will be repeated for a chosen list of articles that have not been retracted, preferably with similar citation metrics. This project will ultimately provide a method for identifying whether or not retracted articles continue to influence online discourse even post-retraction. Based on background reading, I expect that there will be a select few retracted scientific articles relating to COVID-19 that are cited far more often and with polarized sentiment due to how the virus has become politicized.

Background:

P1:

How does the influence of retracted articles to non-retracted articles compare on social media in terms of frequency of citation in support of the article?

P2: If retracted articles are cited the same or more often than non-retracted articles, then retracted articles will degrade the quality and correctness of discourse on social media because false information will continue to be purported as correct despite acknowledgement from the scientific community stating otherwise.

Overview of Misinformation

With the advent of the internet, a source for information has become commonplace for much of the modern world. However, this wealth of information has also contributed large sums of disinformation(Lewandowsky, 2012). Information that is correct may be overwritten by false information that is sensationalized and designed to trigger emotional responses(Lewandowsky, 2012). In a 2019 review of the literature on health related misinformation, it was found that there was a common theme that misinformation tended to be more popular than correct information(Wang, 2019). This is troubling, as it indicates that the spread of misinformation will outpace that of correct information.

Example of impact

Misinformation has many different sources, one of them being scientific articles themselves. Misinterpretation of scientific papers and the sharing of retracted scientific papers can both contribute to public health crises, such as what has happened with the anti-vaxx movement(Hargreaves, Lewis, & Speers, 2003).

Retracted Scientific Articles

The number of retracted scientific articles per 10,000 papers has been increasing steadily in the field of nursing(Fig.1). However, the rates of retractions in nursing are low overall, leading towards the conclusion that there was a general lack of ethical conduct in the field of nursing.

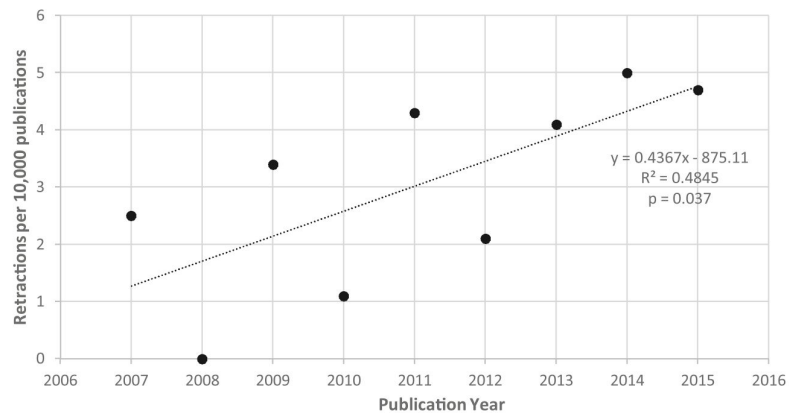


Fig. 2. Association between the normalised number of retracted publications per 10,000 publications against publication year.

Fig 1. Al-Ghareeb, 2018

Another study looking at retracted papers in general found that for 714 retracted articles published from 1973 to 2002, retraction took an average of 49.82 months. For 1333 retracted articles published after 2002, retraction took 23.82 months(Steen,2013). This is good, as it means that false information will be around for less time on average and have less of a chance to influence other researchers' work. However, some studies have found that post-retraction, for particularly contentious papers like the anti-vaxx paper mentioned above, the citations on social media will continue to remain high, therefore influencing public opinion and beliefs.

Overall, the study of retracted articles contributing to misinformation is not new, however, as more studies are published overall and better techniques are developed for analyzing them, better conclusions can be derived from them. In particular is the study of health related misinformation, which has so far mostly been done in relation to infectious diseases. However, more study is warranted as the growing usage and importance of the internet, especially with the global pandemic, promotes greater spread of misinformation around the globe. It can be difficult to predict exactly what will happen as a result of misinformation, but the anti-vaxx movement provides insight into one real-world example that has spiraled into a bit of a health-crisis.

COVID-19 Misinformation

Misinformation about COVID-19 is so great, that the general effects of COVID-19 have been described as not only a pandemic, but an infodemic as well. The virus hit hard and fast in its early days, and some governments responded better than others. However, all governments and people in general were mostly in the dark about the effects of the virus. Information was relatively scarce as COVID-19 was a novel disease. As such, researchers rushed to find information, to perform experiments, and to eventually publish their results. The usually rigorous process of vetting journal articles for potential flaws lapsed a bit, as the rates of retractions on COVID-19 far surpass the normal rate (Yeo-The & Tang, 2020). Due to the high-profile of the virus, media outlets and social media users in general latched on to the newest research that was being done. Some of this research, later found to be in need of retraction, had already spread far and wide, influencing the minds of many.

Experimental Design/Research Plan Goals:

Major Parts of the Project (rough outline) will continue to evolve over time and should be updated frequently. Make sure the goals are SMART oriented.

- Find as many retracted articles relating to COVID-19 as possible
- Clean up the list of articles by determining eligibility by reading the reason for retraction
- Get data on altmetric citations prior compared to post-retraction
- Obtain the context of the citation online prior and post-retraction and compile into database
 - For Facebook
 - Not sure how to do yet
 - For Twitter
 - Not sure how to do yet
 - For Reddit
 - Use Reddit API
- Learn how sentiment analysis works, ins and outs and such
- Train sentiment analysis program and test on predefined phrases that have been developed by other researchers
- Get accuracy to an acceptable rate (rates TBD)
- Use sentiment analysis on the context to determine positive(in support), neutral, or negative in relation to the retracted article for context prior and post-retraction
- Analyze resulting data, does the overall sentiment change at all?

IDV:

- Whether or not an scientific article is retracted

DV

- The sentiment towards the scientific article on social medias

Controls

- Will do same process for non-retracted articles relating to COVID-19 with ideally similar amounts of altmetric citations

Iterations

- Can expand to a different area of study and look at retracted articles there.
- Can hold off on looking at different platforms for now
- Can hold off on looking at post compared to prior to retraction
- Bare-bones would just be looking at retracted scientific journal articles dealing with COVID-19 compared to non-retracted scientific journal articles dealing with COVID-19.

Materials List

- Need my computer

- Access to some specific programs, may cost some money to use

Risk/Safety Concerns:

- No serious safety concerns to address

Data Analysis:

- I can use python programming libraries like pandas to analyze data.

Potential Roadblocks: (with action steps identified of how you might solve these):

- Sentiment analysis is not entirely reliable(Need more research into this)? Struggles with things like negations, metaphors, and contrasting information.(Figure out if it is possible to train program to take these things into account)
- May still not be enough retracted articles to have significant results even with increased rates of retractions dealing with COVID-19.

References: (In APA Format with in-text citations):

Al-Ghareeb, A., Hillel, S., Mckenna, L., Cleary, M., Visentin, D., Jones, M., . . . Gray, R. (2018). Retraction of publications in nursing and midwifery research: A systematic review. *International Journal of Nursing Studies*, 81, 8-13. doi:10.1016/j.ijnurstu.2018.01.013

Eggertson L. Lancet retracts 12-year-old article linking autism to MMR vaccines. *CMAJ*. 2010;182:E199–200

Hargreaves, I., Lewis, J., Speers, T. (2003). *Towards a better map: Science, the public and the media*. London, England: Economic and Social Research Council.

Kardeş, S., Levack, W., Özkuk, K., Aydın, E. A., & Karabulut, S. S. (2020). Retractions in Rehabilitation and Sport Sciences Journals: A Systematic Review. *Archives of Physical Medicine and Rehabilitation*. doi:10.1016/j.apmr.2020.03.010

Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and Its Correction. *Psychological Science in the Public Interest*, 13(3), 106-131. doi:10.1177/1529100612451018

Smith T. C. (2017). Vaccine Rejection and Hesitancy: A Review and Call to Action. *Open forum infectious diseases*, 4(3), ofx146. <https://doi.org/10.1093/ofid/ofx146>

Steen, R. G., Casadevall, A., & Fang, F. C. (2013, July 08). Why has the number of scientific retractions increased? Retrieved October 26, 2020, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3704583/#:~:text=The increase in retracted articles,authors with a single retraction.>

Wang, Y., Mckee, M., Torbica, A., & Stuckler, D. (2019). Systematic Literature Review on the Spread of Health-related Misinformation on Social Media. *Social Science & Medicine*, 240, 112552. doi:10.1016/j.socscimed.2019.112552

Yeo-Teh, N. S., & Tang, B. L. (2020). An alarming retraction rate for scientific publications on coronavirus disease 2019 (COVID-19). *Accountability in Research*, 1-7. <https://doi.org/10.1080/08989621.2020.1782203>

Timeline: (with action steps identified- sub-deadlines will continue to evolve):

Brainstorming			
Summer Brainstorming	100%	7/1/20	8/26/20
Brainstorming done in school	100%	8/26/20	9/15/20
Background and Project proposal			
Background Reading(12-15 papers)	100%	9/16/20	10/16/20
Email experts(ongoing?)	40%	9/25/20	11/1/20
Extra background reading	5%	10/16/20	10/25/20
Project proposal	70%	10/16/20	11/1/20
Paperwork	0%	10/16/20	11/1/20
Execution			
Literature Review	0%	11/1/20	12/10/20
Ad. Hoc. Testing	40%	10/10/20	10/24/20
Gather articles and obtain citation information	0%	11/1/20	11/8/20
Determine citation context	0%	11/8/20	11/15/20
Extra time for troubleshooting	0%	11/15/20	11/25/20
Prepare for Dec. fair	0%	11/25/20	12/11/20
Analysis and poster			
Complete data collection		11/15/20	12/20/20
Analyze		12/20/20	1/21/21
Write up results		12/21/20	1/21/21
Finish poster		1/22/21	2/10/21