Testing the Keystone Framework by Analyzing Positive Citations to Wakefield's 1998 Paper

Amulya Addepalli[1,3][0000−0003−4871−0290], Karen Ann Subin[2,3][0000−0003−2656−5328], and Jodi Schneider[3][0000−0002−5098−5667]

[1] William Fremd High School Amulya.Addepalli@gmail.com
[2] Naperville North High School kasbin864@gmail.com
[3] University of Illinois at Urbana Champaign jodi@illinois.edu

Abstract. Science is constantly developing as new information is discovered. Papers discredited by the scientific community may be retracted. Such papers might have been cited before they were retracted (as well as afterwards), which potentially could spread a chain of unreliable information. To address this, Fu & Schneider (2020) introduced the keystone framework for auditing how and whether a paper fundamentally depends on another paper, and proposed that an alerting system be developed to flag papers that fundamentally depend on retracted papers. The need for expert labor is the main challenge of such alerting in such systems. This paper tests whether a flowchart process for non-experts could accurately assess dependencies between papers, reducing the need for expert assessment. We do this by developing such a process and testing it on citations to one highly cited retracted paper. In our case study, non-experts using our process can resolve the question of dependency in about half the cases. Two annotators had 92.9% agreement on 85 papers annotated, with 100% agreement after discussion. In future work we will assess the reliability of non-experts' decisions as compared to experts, and identify possibilities for automation.

Keywords: retracted papers · knowledge maintenance · keystone citations · Wakefield · misinformation in science.

1 Introduction

Modern science evolves through the centuries, moving along the dual path of reforming and reinventing; it builds itself on trial and error. Papers discredited by the scientific community may be retracted when they are found to be in error or even fraudulent; about 1 paper in 2500 is retracted [1]. Such papers might have been cited before they were retracted (as well as afterwards), which potentially could spread a chain of unreliable information to a wide body of literature. Ideally, scientific information that becomes retracted would also be removed from or updated in earlier publications, and current guidelines specify that "Articles that relied on subsequently retracted articles in reaching their own conclusions... may themselves need to be corrected or retracted," [4]. However, in practice, no

such auditing process is in current use, even though papers are commonly cited after they are retracted [10, 3]. This can allow disproven information to spread among the scientific community creating a domino effect which could invalidate the argument of a new paper written. But are there any corrective measures?

Recently, Fu & Schneider introduced the keystone framework [7], which aims to determine how a paper is affected when relying upon unsound findings. It is an approach for auditing how and whether existing papers depend on a particular paper, through the use of argumentation theory and citation context analysis. In particular, a keystone citation is defined as a citation that a paper heavily bases its argument or knowledge around. Fu & Schneider proposed that an alert system could selectively notify authors whose work significantly depended on a retracted paper. Although alerting is promising, a key limitation is that the proposed approach requires slow, laborious work done best by experts. Ideally, some of this work could be scaffolded by an expert, creating a process that a group of non-experts could apply, to reduce the expert labor required.

In this paper, our goal is to explore the effectiveness of a system which can accurately assess the dependency of a paper on unreliable data, using non-experts. This is accomplished by testing the process on a single highly cited retracted paper, which we introduce next.

2   Case Study

To analyze the impact a retracted paper can have on work that cited it, we focused on a paper written in 1998 by Andrew Wakefield and colleagues called Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children [12]. Not long after its publication in *The Lancet*, the paper came under fire [6]. Wakefield's paper was partially retracted in 2004 and fully retracted in 2010. The main claim of the paper, which has been fully discredited [5], is that the measles, mumps and rubella (MMR) vaccine causes children to be more susceptible to behavioral regression and a form of autism called pervasive developmental disorder. The Wakefield paper contributed to mistrust of vaccines and low vaccination rates, setting the stage for measles outbreaks in the UK, US, and Canada [6]. The Wakefield paper is a good case for us to study keystone citations because citing papers that were published before the retraction, between 1998-2004, could not have known that it would be retracted.

3   Related Work

In 2019, scholarly communications librarian Elizabeth Suelzer and her team conducted a study to inspect the features of citations that mention Wakefield's 1998 paper. They found 1153 papers in English that cite Wakefield [11]. Suelzer's team used a stepwise approach to group the articles into categories such as negative, perfunctory, affirmative, and assumptive [8]. Alongside the abundance of affirmative citations, Suelzer and her team discovered a number of authors who didn't record the Wakefield paper as retracted.

More recently, Ivan Heibi and Silvio Peroni [8] analyzed open citations to Wakefield's paper found in COCI, the OpenCitations Index of Crossref DOI-to-DOI citations [9]. They annotated in-text citations and used both citation intent analysis and topic modeling to understand how Wakefield's paper has been cited. Over time, more papers mentioned the retraction, with the largest percentage (61%) of mentioning papers in the most recent (2017) publication year in their dataset. Following the partial retraction, publications beyond medicine became more prominent, with an increase in citations from social sciences papers and from new areas such as economics and environmental sciences.

4 Methods

4.1 Our Dataset

Our study focused on papers that cited and used information from the Wakefield article in an affirmative light: we hypothesized that these articles would be the most problematic since they were more likely to depend on the Wakefield paper. Starting with the dataset that Suelzer [11] provided, we retained papers matching the following criteria: 2004/2010 retractions were not referenced, classified as affirmative, published in any year (pre- or post-retraction). This gave us 89 papers that cite Wakefield's 1998 paper positively and fail to acknowledge the partial or full retraction. The subset was created to have a reasonable number of papers: large enough for any trends/patterns in the articles to be identified and yet small enough that we could analyze the dataset multiple times if needed as our theories developed. We used Zotero to store and organize the full-text of the articles we were analyzing; in the study period we did not find the full-text of 4 papers. Hence, ultimately our dataset was 85 papers.

4.2 Annotation Design

The goal of the annotation process was to operationalize the keystone framework, in order to enable non-experts to identify keystone citations to Wakefield's retracted 1998 paper. That is, the objective was to be able to distinguish whether or not a paper that cites the Wakefield paper can stand without the support of the information from the Wakefield paper. One person (AA) iteratively designed the annotation process in consultation with the larger group. During the initial, year-long process, one annotator manually used a systematic data analysis methodology, first categorizing the papers manually on an Excel spreadsheet without using a guideline. After recognizing the common steps taken to group each paper, a nuanced method of analyzing each paper was created which is represented the annotation manual excerpted in Fig. 1, available in our data deposit [2]. The process guides the annotator using questions such as "Does the article title correlate with the ideas presented in Wakefield?" and "How many times was Wakefield mentioned in the paper?". Using this process a non-expert should be able to determine if a paper is independent of the cited paper, dependent on it, or should be shown to a professional to make that determination.

Follow these steps:

I.   Read the Wakefield paper and gain a basic understanding of the main ideas from its method and findings.

    A. This is an important step because the background knowledge gained from reading the original article will help classify the cited articles much more easily.

II.   Choose an article that cites the Wakefield paper from the dataset of 89 articles.

III. Read the title of the article and briefly skim over the main idea of the article to determine where the citations of Wakefield are located.

IV.   Look at the authors of the article because if Andrew Wakefield is a co-author it can be considered a self-citation and the paper has to be looked at more carefully.

V.   By looking at the article's title use the range "Article title range" and determine which branch of the given flowchart will be used. If its "Article title range" is between (3-5) it is at higher risk of being a keystone citation.

VI.   If the "Article title range" is between 1-2 look at the number of times the paper cites Wakefield in both singletons and clusters. If the "Article Wakefield citation count" is between 3-5 read the rest of the article to determine the main ideas of the article to determine where to go next.

VII. When an article is determined independent the Wakefield citation does not pose a threat to the article's credibility so it can be left alone. (We expect a large percentage of articles to be deemed independent.)

VIII. Keep going down the flowchart and if the question of "How much information is being used" is reached, look at the range of "Range for the amount of information being used" on the side of the flowchart to determine where to go next.

IX.   Once a conclusion is reached for which category the article is categorized in, record the findings in an Excel spreadsheet that has been separated into the columns of "Knowledge Claim", "Independent, dependent or "Up to a Professional", "Main idea of the Paper", and "Author/Title." The Excel sheet can be filled out while analyzing the paper to make it easier to make conclusions.

X.   Annotate which pathway was taken to get to the conclusion of either Dependent, Independent and Up to Professional under the column of "Pathway used from Flowchart"

    A. Example: 0-1-3-5

XI.   Repeat the previous steps for all 89 articles.

Fig. 1: A part of our annotation manual [2] which is intended to provide a replicable process for non-experts to recognize keystone citations to Wakefield's paper.

4.3   Testing the Annotation

To test our annotation manual [2], we recruited a second annotator (KAS) to provide suggestions on how to make the manual more understandable to a wider range of people and help locate areas of concerns on specific pathways on the flowchart. After comparing the analysis of each paper, we were able to agree on specific trends found in the results from the flowchart and made revisions to the flowchart accordingly. The final annotation was stored in Excel spreadsheets [2], completed independently by two annotators (AA, KAS), with some discussions regarding the level of expertise needed to make determinations of up to a professional articles. For instance, after seeing a number of articles using the same pathway (0-2-12-13), we updated the flowchart to classify these as independent; this pathway selects articles whose main topic autism, gastrointestinal problems, or the specific disorder (2); which include information about the Wakefield article (12); but which cite it outside the methods section (13). We compared the decisions, resolved disagreements, and came to a final determination for each article. We computed the percentage agreement and we analyzed the disagreements in an error analysis.

Left flowchart (independent):

Does the article title correlate with the ideas presented in Wakefield 1998?

Minimally related

Article Title: *"Gluten and casein free diets in autism: a study of the effects on food choice and nutrition"*

How many times was Wakefield mentioned in the paper?

1 or 2 citations to Wakefield 1998

3

Are the articles in singleton or cluster?

Singleton

6

What kind of information is being taken from Wakefield?

Knowledge Claim from Wakefield: *"Wakefield et al. (1998) reported an increased incidence of ileal lymphoid nodular hyperplasia" in children with autism*

Talking about information from the paper

8

Where in the article is Wakefield being cited?

If anywhere else

9

Independent

Reasoning: *The citing article merely discussed the effects of the Wakefield paper in the scientific community without directly quoting false information from it, so the paper was classified as Independent.*

Right flowchart (dependent):

Does the article title correlate with the ideas presented in Wakefield 1998?

Overlap with a main topic: **autism in general, gastrointestinal problems, behavioral problems, Ileal-lymphoid-nodal hyperplasia non-specific colitis**

Article Title: *Intestinal pathophysiology in autism*

2

What kind of information is being taken from Wakefield?

Knowledge Claim: *"In the study by Wakefield et al. (14) examination of histological sections of intestinal tissue from autistic patients revealed a near doubling of the number of lymphoid follicles, as well as follicle enlargement and merging of adjacent follicles."*

Talking about information from the paper

12

Where in the article is Wakefield being cited?

If in method

14

How much of the citing paper uses information from Wakefield 1998?

An entire section uses info taken from Wakefield 1998 or The entire paper is based on Wakefield 1998

Make sure evidence being used is correct

If incorrect

18

Dependent

Reasoning: *Due to the sizable amount of data cited from Wakefield's paper within the methodology of this paper, the information was reviewed to determine the validity of the statements. Upon further inspection, the information was forund to be inaccurate and the citing paper was deemed to be Dependent.*

Fig. 2: Determining that an article was (a) independent (left) or (b) dependent (right).

4.4   Annotator Background

The annotators were high school students with honors biology coursework and a strong interest in science. A deep professional/scientific background in biomedicine or a related area, including up-to-date biomedical knowledge and knowledge of medical terminology, would be needed to fully understand the research papers annotated. Our focus was on the triage process, to determine which papers needed further vetting from a professional in order to determine the validity of a paper's scientific arguments and dependence on the Wakefield paper.

Figure 2 illustrates the annotation process for two articles from our subset. In each case, each annotator answered a series of questions, recording the pathway taken through the flowchart by noting the sequence of numbers encountered. First, the annotator read through the article and determined the most important points of the article. Second, the annotator examined how reliant each citation was on the Wakefield paper.

The paper classified as independent, shown on the left in Figure 2, mentioned the Wakefield paper just once and the annotator judged that, since the citing article simply referenced the consequences of the Wakefield paper on the scientific community, it could be classified as independent.

The paper classified as dependent, shown on the right in Figure 2, cited Wakefield multiple times. The annotator deemed that the paper directly cited data from the Wakefield paper, and in sizable amounts as well. Since the annotator assessed that a portion of the evidence taken from the Wakefield paper was mentioned in the methodology of the citing paper, and noted that substantial section of the paper referenced the Wakefield article. To check the validity of the statements, the annotator researched the topics discussed and determined that the information cited was inaccurate. Consequently, the paper was categorized as dependent on Wakefield's paper.

5   Results

From the annotation spreadsheets [2], we calculated the number of articles which had been classified into each category. Two annotators had 92.9% agreement on 85 papers annotated, with 100% agreement after discussion. All 6 disagreements are shown in Table 1. As seen in Fig. 3, using our process, we were able to classify 89.4% of the articles as "independent", meaning that they take a minimal to no amount of information from the Wakefield paper and it doesn't damage their paper's credibility. Only 2.4% of the articles were classified as "dependent", meaning that they use the Wakefield paper in a fundamental way. And 8.2% of the articles were classified into the "up to a professional" category, meaning that they need further review. Table 2 shows examples of some of the deciding factors in making these determinations.

Table 1: Error analysis

ID	Error	Final Annotation	Annotator 1	Annotator 2	Pathway 1	Pathway 2	Comment
Deisher 2015	Different pathways	Up to a professional	Up to a professional	Independent	0-1-4-6-8-10	0-2-11	Public health-related analysis, talks extensively about the "Wakefield Scare" and the relationship between the MMR vaccine and autism prevalence
Levy 2007	Different pathways	Up to a professional	Up to a professional	Independent	0-1-4-6-8-10	0-1-4-6-8-9	Seeks to determine whether gastrointestinal symptoms are related to diets, comparing autistic and non-autistic children
Cohly 2005	Recording error	Independent	Up to a professional	Independent	0-2-12-13	0-2-12-13	Immunological findings in autism
Kawashima 2000	Flowchart evolution	Up to a professional	Up to a professional	Independent	0-2-12-14-15	0-2-12-14-15	Similar to Wakefield's study in that the children all had ileal lymphoid nodular hyperplasia hyperplasia and nonspecific colitis
Sabra 1998	Flowchart evolution	Up to a professional	Up to a professional	Independent	0-2-12-14-15	0-2-12-14-15	Mechanisms similar to that of ileal-lymphoid-nodular hyperplasia
Horton 1998	Flowchart insufficient	Up to a professional	Up to a professional	Independent	0-2-12-14-16-17	No current classification	"We did not prove an association between measles, mumps, and rubella vaccine and the syndrome described."

Table 2: Examples of key indicators that helped categorize papers

| Decision	Reason	Description	Paper ID
Independent	Cluster Citation	There is only one citation to Wakefield, which supports the same information as multiple different papers: "Preliminary evidence has suggested that some children with autism and other PDDs may experience gastrointestinal problems (Horvath, Papadimitriou, Rabsztyn, Drachenberg, and Tildon, 1999; Lightdale et al., 2001; Wakefield et al., 1998; Williams et al., 2000)"	Kerwin, 2005
Up to a professional	Multiple uses, including in the methods and results section	Used Wakefield's paper as part of their background, method and results. The paper was released shortly after Wakefield's and constantly mentions the fact that the new study was done based on the findings of Wakefield.	Sabra, 1998
Up to a professional	The paper uses Wakefield to make a conclusion about a medical phenomenon that we don't have the expertise to determine whether or not it is correct.	"Wakefield et al have introduced strong data to support our hypothesis that FA is the pivotal causative factor that produces the lesions in the terminal ileum that consist of greatly enlarged lymphoid nodules containing large collections of lymphocytes in the GI lymphoid tissues adjacent to Peyer patches"	Bellanti, 2003
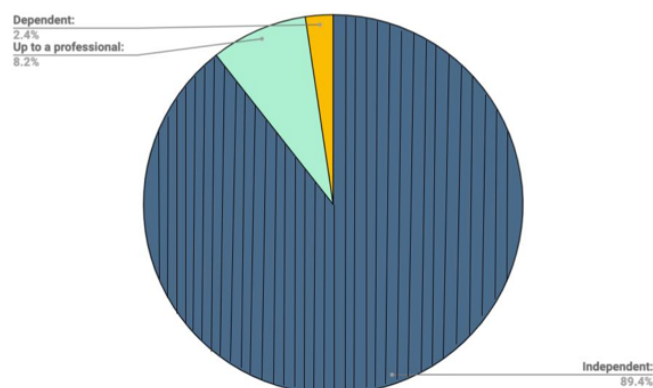
Fig. 3: The percentage of articles from our subset that we classified as dependent on versus independent of Wakefield's 1998 paper.

6   Discussion and Conclusion

In this paper, we explore and analyze how current research builds on previous work for one specific retracted paper. According to our analysis, only 9 out of the 85 papers we analyzed have a significant chance of needing correction due to the Wakefield retraction. This greatly reduces the number of papers that would need attention from authors or from editors in order to avoid the spread of scientific misinformation. Further work is needed to check whether experts would agree with our assessment.

Our future work lies in three areas. First, we need to validate our approach against an expert's view, particularly to determine whether the papers that we deemed dependent and up to a professional are in fact more likely to require correction than those that we deemed independent of Wakefield. Second, we would like to generalize this process, in order to determine how to handle any paper that cites any retracted paper. This is likely to require some domain analysis based on the topics, but perhaps not the full argumentation analysis proposed in the keystone framework [7]. Ultimately, we hope to help researchers determine the credibility of papers, since there are no guidelines for authors or editors to follow when a retracted paper has already been cited, and no systematic analysis of which papers might need reexamination. Third, we hope to create a computer algorithm that can go through the steps of our flowchart to avoid manual work and minimize human error. This would have the advantage of scaling the work so that it could be used in practice to support knowledge maintenance in digital libraries.

Data Availability

The annotation manual and flowchart, and annotated data are available at [2].

Acknowledgements

References

1. 'tis but a scratch: Zombie research haunts academic literature long after their supposed demise. The Economist **439**(9251), 89 (Jun 2021)
2. Addepalli, A., Subin, K.A., Schneider, J.: Dataset for testing the keystone framework by analyzing positive citations to Wakefield's 1998 paper (2021), https://doi.org/10.13012/B2IDB-2532850_V1
3. Bar-Ilan, J., Halevi, G.: Post retraction citations in context: a case study. Scientometrics **113**(1), 547–565 (Oct 2017), http://doi.org/10.1007/s11192-017-2242-0
4. COPE Council: Retraction guidelines (Nov 2019), http://doi.org/10.24318/cope.2019.1.4
5. Dudley, M.Z., Salmon, D.A., Halsey, N.A., Orenstein, W.A., Limaye, R.J., O'Leary, S.T., Omer, S.B.: Do vaccines cause autism? In: The Clinician's Vaccine Safety Resource Guide, pp. 197–204. Springer International Publishing, Cham (2018), https://doi.org/10.1007/978-3-319-94694-8_26
6. Eggertson, L.: Lancet retracts 12-year-old article linking autism to MMR vaccines. CMAJ : Canadian Medical Association Journal **182**(4), E199–E200 (Mar 2010), https://doi.org/10.1503/cmaj.109-3179
7. Fu, Y., Schneider, J.: Towards knowledge maintenance in scientific digital libraries with the keystone framework. In: Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020. pp. 217–226. ACM, Virtual Event China (Aug 2020), https://dl.acm.org/doi/10.1145/3383583.3398514
8. Heibi, I., Peroni, S.: A qualitative and quantitative analysis of open citations to retracted articles: the Wakefield 1998 et al.'s case. Scientometrics **126**, 8433–8470 (Aug 2021), https://doi.org/10.1007/s11192-021-04097-5
9. Heibi, I., Peroni, S., Shotton, D.: Software review: COCI, the OpenCitations Index of Crossref open DOI-to-DOI citations. Scientometrics **121**(2), 1213–1228 (Nov 2019), https://doi.org/10.1007/s11192-019-03217-6
10. 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 **online first** (2021), https://doi.org/10.1162/qss_a_00155
11. Suelzer, E.M., Deal, J., Hanus, K.L., Ruggeri, B., Sieracki, R., Witkowski, E.: Assessment of citations of the retracted article by Wakefield et al with fraudulent claims of an association between vaccination and autism. JAMA Network Open **2**(11), e1915552 (Nov 2019), https://doi.org/10.1001/jamanetworkopen.2019.15552
12. Wakefield, A., Murch, S., Anthony, A., Linnell, J., Casson, D., Malik, M., Berelowitz, M., Dhillon, A., Thomson, M., Harvey, P., Valentine, A., Davies, S., Walker-Smith, J.: RETRACTED: Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. The Lancet **351**(9103), 637–641 (Feb 1998), http://doi.org/10.1016/S0140-6736(97)11096-0