

Informetric Analysis on the International Retracted Publication Based on the Web of Science Database

Chunbo Zhang ^{a,*}, Kun Ding ^b and Zeyuan Liu ^c

Institute of Science and Science and S&T Management, Dalian University of Technology, China

Addresses: No.2 Linggong Road, Ganjingzi District, Dalian City, Liaoning Province, China, 116024

^azhangcb@mail.dlut.edu.cn, ^bdingk@dlut.edu.cn, ^cliuzy@dlut.edu.cn

*Corresponding author

Keywords: Retracted publication, Informetric analysis, Web of Science, Scientific norm.

Abstract. Retracted publications are the special literature carriers revealing scientific errors and misconduct. Based on Web of Science, this paper accessed the authoritative and plentiful retracted publications through heterogeneous multi-source data and effective retrieval strategies. Then, the informetric analysis was performed on the retracted publications with a focus on the timeline, geographical locations, disciplines, journals, institutions, and authors. Current findings show that there is a growing number of retracted publications worldwide as well as in mainland China. Recidivist and large-scale centralized publication retractions have exerted numerous influences on the changes in the number of retracted publications. Countries with strong scientific research capabilities have more retracted publications. In addition, the number of retracted publications has significant differences corresponding to different disciplines. Specifically, life science is a typical field that the retraction of scientific publications frequently occurred, which can also be seen in partial interdisciplinary publications. In terms of journals, which are counted as small group but with a high ratio of retraction should be further investigated. The institutions with more publications retracted can be divided into three types: institutions with more and scattered disciplines, institutions with more but concentrated disciplines, and institutions with fewer and concentrated disciplines. Recidivists are the focus at the author level and have quite different specific behavioral patterns. The main findings from the research are of great informetric value for the theoretical research on scientific norms and the practices of scientific research integrity supervision.

Introduction

In recent years, a series of events have made publication retraction a hot topic in the domestic academic community. In response to the large-scale retraction of Chinese publications by international journals and subsequent disclosure of misconducts, the Science and Technology Administration of China has taken decisive action to eradicate scientific misconducts, including the Five Prohibitions for Academic Publications jointly issued by seven departments in December 2015, the Self-disciplinary Regulations for Scientific and Technological Workers released by China Association for Science and Technology in July 2017 as well as the Opinions on Further Strengthening the Construction of Scientific Research Integrity promulgated by the General Office of the Central Committee of the Communist Party of China and the General Office of the State Council in May 2018. Publication retraction itself is also a behavioral mechanism that demonstrates a sense of responsibility for the implementation of scientific research activities and the development of scientific knowledge. As Nature Biotechnology stated in the editorial regarding the retraction of the publication by Han Chunyu, “the retraction decision is the best practice to maintain the integrity of published scientific records.” [1]. For the problems reflected in the scientific research practices and the existing disputes, it is necessary to study the retracted publications as a special research subject. Based on above, this paper utilized the Web of Science database to conduct informetric analysis on global retracted publications, which reveals the rules and characteristics of the retracted publications in different dimensions and provided relevant enlightenment and suggestions for researchers,

journals and the departments of scientific management.

Organization of the Text

With the dramatic increase in the number of retracted publications and the retraction rates [2-3], the characteristics and distributions with respect to this particular literature group have been explored on a large scale. However, due to the differences in data selection, the research conclusions are correspondingly different and even contradictory. Therefore, it is necessary to construct a relatively comprehensive, authoritative and informative research dataset. Based on Web of Science, this study acquired a large number of retracted publications through heterogeneous multi-source data [4] and effective retrieval strategies [5]. Specifically, a total of 14 databases for three types were retrieved to obtain an initial dataset, including the professional citation databases such as Web of Science, PubMed, and Scopus, the full-text journal database which covers ten major international publishers as well as a dedicated publication retraction platform--Retraction Watch Database. Then, the final dataset was obtained by filtering at the Web of Science (Fig. 1). According to the publication time of the research data, 6,397 retracted publications were acquired as of the end of 2017.

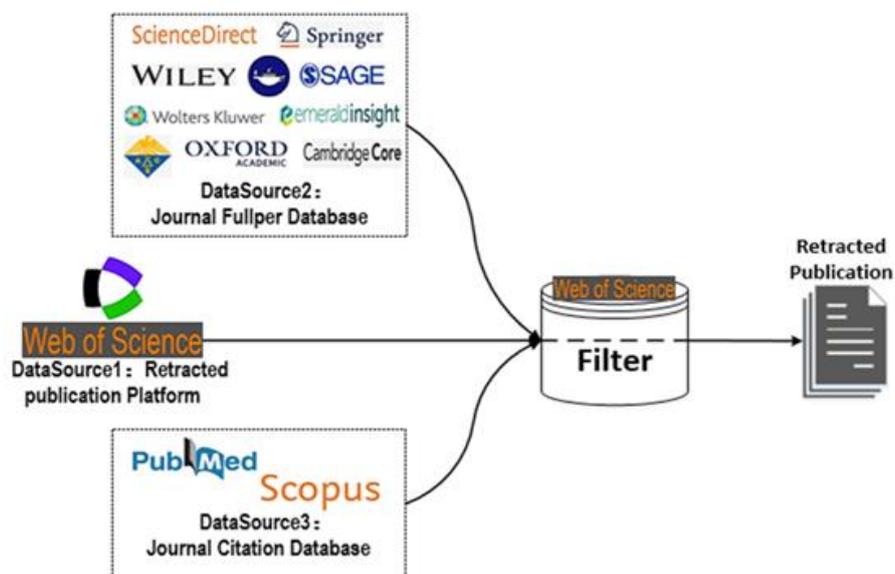


Fig. 1 Acquisition process of research data

Results

According to the structural data included in scientific publications, the analysis was conducted from six aspects: time, geological locations, disciplines, journals, institutions, and authors.

Analysis of Temporal Characteristics

Since the annual statistical data on the retraction time of the publications is fixed whereas the annual distribution of the publication time is variable, the distribution of the retraction time was examined.

Retraction rate refers to the ratio of the number of retracted publications in a certain year to the number of published publications this year [6]. In terms of the calculation of international retraction rate, the denominator was the number of all articles and reviews recorded in the SCI-E and SSCI databases in a certain year while it was limited to the number of the corresponding publications published by institutions located in mainland China in the calculation of Chinese retraction rate. The value was calculated in the unit of part per million, which refers to multiply the ratio by 100,000%.

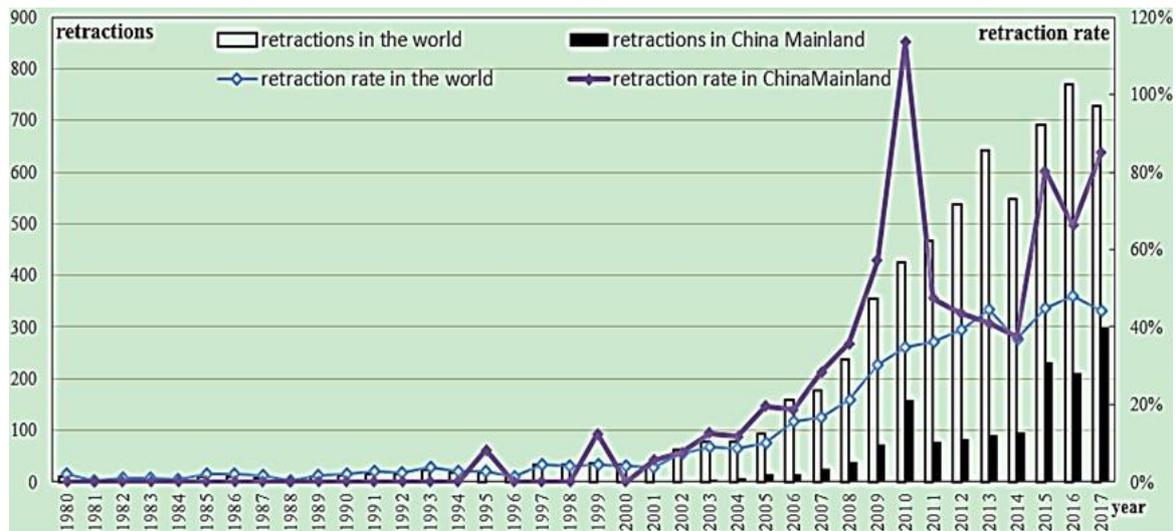


Fig. 2 Annual distribution of retracted publications worldwide and in mainland China (1980-2017)

There is a growing trend of retracted publications at a global level and a regional level (Fig. 2). The annual distribution of retracted publications worldwide can be divided into three stages based on the changes in relevant ratios. The first stage started before 2001 in which the publications were only occasionally retracted. Although the academic misconduct and publication retraction of some scholars were visible at that time, it was still uncommon. The exposure of such activities was likely to cause concerns but not sufficient to raise the introspection within the scientific community. The second stage was from 2002 to 2013 which marked the outbreak period of the publication retraction. As a major scandal in the history of science, the Schon Incident reached the first tipping point of publication retraction [7]. While being shocked by the Schon Incident, members of the scientific community, including societies, institutions, and journals, profoundly introspected and strengthened the scrutiny, disclosure and penalties for scientific misconducts. Thus, publication retraction has become a regular mechanism for handling scientific misconducts long with an increased number of retracted publications. Until 2013, an unprecedented burst of publication retraction appeared. For the scholar with the most retracted publications, Yoshitaka Fujii, 84 publications of his total 126 retracted records were made in 2013. The third stage began from 2014 to the present in which publication retraction has frequently occurred. In addition to multiple regular publication retraction incidents in the United States, several “centralized” large-scale retraction in the mainland China and Iran has greatly boosted the number of retracted publications in the world.

Analysis of Geological Characteristics

Geographically referenced information is another basic characteristic of the literature and can reflect the distribution of the regions where the publications in a particular type or on a certain topic are produced. As mentioned earlier, this paper only conducted statistical analysis on the institutions of the first authors (Co-) and the corresponding authors (Common) who only serve as the core instructors and key implementers [8, 9] and their countries/regions.

Table 1. Countries and regions with more than ten retracted publications

Country/Region	Number	Country/Territory	Number	Country/Region	Number
USA	1535	Egypt	70	Denmark	19
China Mainland	1446	Brazil	69	Thailand	19
Japan	513	Switzerland	54	Austria	18
India	369	Sweden	54	Serbia	18
Germany	334	Singapore	43	Tunisia	18
South Korea	280	Pakistan	43	South Africa	17
Iran	244	Saudi Arab	38	Algeria	17
UK	222	Malaysia	36	Portugal	16
Italia	151	Poland	31	Romania	16
China Taiwan	142	Hong Kong	31	Finland	15
Canada	124	Israel	30	Norway	15
Netherlands	103	Belgium	24	Argentina	14
Turkey	97	Czech	24	Mexico	13
Spain	95	Greece	23	Hungary	12
Australia	94	Russian	22		
France	89	Ireland	21		

Publication retraction occurred in a total of 95 countries and regions. Table 1 lists the countries and regions where more than ten publications were retracted. It can be seen that these countries and regions are widely distributed, ranging from traditional developed countries to emerging powers represented by China and India, as well as some countries that play an important role in regional development. Therefore, it can be initially judged that the number of retracted publications is directly related to the number of local scientific research outputs, while the regional scientific research culture and environment may also have a certain effect.

Analysis of Disciplinary Characteristics

The way of knowledge production and academic citation habits differ with the disciplines, and their scientific misconducts may also be different. This paper selected the ESI disciplines as the criteria for disciplinary classification. In the meanwhile, the AI index [10] was used to construct the disciplinary retraction index RI in order to examine the relative tendency of publication retraction in a certain discipline. R_i is the number of retracted publications in the discipline i , $\sum R$ is the total number of retracted publications in all the disciplines, namely 6,397; P_i is the total number of publications in the discipline i (1996 to 2017), and obtained through the Documents in JIF Journals on the In Cites platform, and $\sum P$ is the total number of publications in all the disciplines.

$$RI = \frac{R_i / \sum R}{P_i / \sum P} \tag{1}$$

Fig. 3 shows the number of retracted publications in different disciplines. The number of retracted publications (the size and value of the dots) and the retraction index are closely related to the disciplines. All the disciplines were divided into six clusters according to two dimensions.

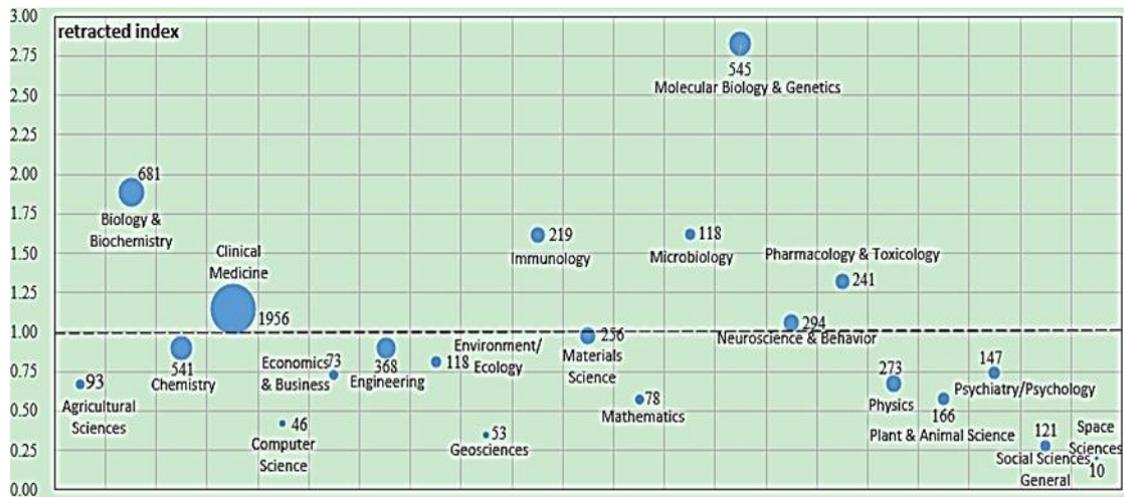


Fig. 3 Distribution of ESI disciplines for retracted publications (1963-2017)

Cluster 1 refers to the life science discipline in which the number of publications involved is large and the retraction index is high. Cluster 2 refers to the space science and earth science discipline in which the number of publications involved is small and the retraction index is low. These fields mainly present the characteristics of “grand science” in terms of the pattern of knowledge production. Usually, they require multi-person collaboration and large-scale equipment to monitor, collect and measure scientific research data, and have high accuracy requirements for data and results. Cluster 3 refers to the interdisciplinary community in the Natural Engineering Sciences in which the number of publications is large and the retraction index is also high. This disciplinary group includes the interdisciplinary subjects (such as materials science, engineering science, environmental and ecological science) that study a certain subject through multidisciplinary approaches and the interdisciplinary subjects that consist of multiple branches of their own and apply their methods and thoughts to a variety of subjects (such as chemistry). Cluster 4 refers to the computational disciplines in which the number of publications involved is small and the retraction index is low. Mathematics and computer science are typical computational disciplines. The formula derivation and data processing procedures were fully presented in the publications. Therefore, they reproduce the knowledge production process the most vividly in the sciences. Cluster 5 refers to the community of agricultural, plant and animal disciplines in which the number of retracted publications is large but with a relatively low the retraction index. Biological and chemical approaches have become the main research methods and experimental standards for this disciplinary community. Cluster 6 refers to the community of disciplines of social science. Although the number of retracted publications is not large, the retraction indexes of sub-fields still are obviously different. Economic, business, mental and psychological sciences all focus on data acquisition, processing, and use. In these fields, the number of retracted publications is large.

Analysis of Journal Characteristics

The 6397 retracted publications involved 2,175 journals, including a few journals with high retraction rates and a large number of journals with low retraction rates. There were 99 journals with ten or more retracted publications, accounting for 36%, while there were 1,971 journals with 5 or fewer retracted publications, accounting for 52.13%. Therefore, attention should be paid to a few journals with high retraction rates.

Table 2. Top 50 journals with the most retracted publications

Journal	Number	Journal	Number	Journal	Number
J Biol Chem	177	Mol Cell Biol	27	Mater Sci Eng A	18
Tumor Biol	144	Eur J Anaesth	26	Tissue Eng Med	18
Acta Crys Sect E	116	J Am Chem Soc	26	Clin Cancer Res	17
PNAS	88	Br J Anaesth	24	Diabetes	17
Science	79	Cancer Res	24	Eur Rev Pharma	17
PLoS One	69	Int J Cancer	24	J Hazard Mater	17
Anesth Analg	65	Genet Mol Res	23	J Pers Soc Psychol	17
Nature	60	J Neurosci	23	J Virol	17
J Immunol	51	Embo J	22	Faseb J	15
J Vib Control	49	Int J Clin Exp Med	22	Mol Neurobiol	15
J Clin Invest	39	Eur J Med Res	20	Mol Cells	15
Cell	35	Immunop Immu	20	Neuroscience	15
Can J Anesth	34	Renew Sust Rev	19	Phys Rev B	15
Biochem Biophys	33	Appl Phys Lett	18	Arch Biol Sci	14
Blood	32	Biomed Res Int	18	Int J Clin Exp Pathol	14
Diagn Pathol	31	Infect Immun	18	Met Mater-Int	14
Mol Biol Rep	31	J Cell Sci	18	OncoTargets Ther	14

Table 2 shows that the journals with high retraction rates exhibit significant stratified characteristics. First, the high-impact journals have indeed become one of the main battlefields for publication retraction. Twelve of the 51 journals listed are from the Nature Index corpus. Second, a small number of journals with lax reviews and even obvious regulatory loopholes have carried out large-scale centralized retraction of problematic publications. These journals have also been severely criticized by the scientific community, and many have been excluded from the SCI catalog because of the possible predatory nature [11]. The third type covers journals with large publication volume, high fees, and relatively low publication difficulties, which are called “water journals”. Finally, the journals of anesthesiology have their special circumstances. The quality of these journals is not poor, but very few recidivists are involved in large-scale fraudulent behaviors. In the case of four anesthesiology journals listed in the table, 109 of the 127 retracted publications were written by Yoshitaka Fujii, Joachim Boldt, and Scott Reuben. Therefore, these journals should claim the responsibility for publishing these articles [12].

Analysis of Institutional Characteristics

Compared with journals, institutions with retracted publications were more widely distributed, but there were still some dominant institutions with higher retraction rates (Table 3). These institutions with higher retraction rates were divided into three basic types according to the disciplinary distribution of the retracted publications.

Table 3. Top 50 institutions with the most retracted publications

Institution	Number	Institution	Number
University of Tsukuba	91	Pingtung University	30
Jinggangshan University	89	University of Ryukyus	29
Ludwigshafen Hospital	85	CSIR-India	28
Harvard University	79	University of Toronto	28
Chinese Academy of Sciences	71	UCSD	27
Shanghai Jiao Tong University	68	University of Campinas	27
Shandong University	67	University of Groningen	27
University of Tokyo	62	Mayo Clinic	26
Islamic Azad University	52	NUS	26
National Institutes of Health-USA	51	Stanford University	26
China Medical University	44	UCLA	26
Duke University	43	UTMD Anderson Center	26
Kaohsiung University of S&T	43	Central South University	25
CNRS-France	39	Tilburg University	24
Fudan University	39	UCSF	24
Toride Kyodo General Hospital	39	Wuhan University	24
HUST	38	Yonsei University	24
Cheng Kung University	37	Harbin Medical University	23
Tongji University	37	Nanjing Medical University	23
Zhejiang University	37	Second Military Medical University	23
JST	35	Spanish National Research Council	23
Wayne State University	32	The PLA General Hospital	23
Seoul National University	31	University of South Florida	23
University of Tehran	31	Capital Medical University	22
Bell Laboratories	30	Indian Institute of Technology	22
Columbia University	30		

The institutions with more and scattered disciplines are mainly state-level scientific research institutions that build disciplinary layouts and (jointly) establish branches in various parts of the country, such as the Chinese Academy of Sciences and the French National Scientific Research Center. These institutions are small in number, which is related to their special research management system. The institutions with more but concentrated disciplines are mainly world-famous or domestic well-known comprehensive universities. Therefore, they involve many disciplines, such as Harvard University and Shandong University. The institutions with fewer and concentrated disciplines are mainly medical institutions and institutions that are greatly influenced by recidivists such as the University of Tsukuba. The biomedical field at the disciplinary level and the recidivist phenomenon at the author level are the two main factors resulting in the publication retraction within the institutions.

Analysis of Author Characteristics

The statistical data on the first authors and the corresponding authors (common) of the retracted publications demonstrates two main phenomena: first-offense phenomenon in the number of authors and recidivist phenomenon in the number of publications. Among the 7,722 authors, 6,663 were “first offenders” who have only one publication retracted. The recidivists with multiple publications retracted should be the focus. This study defined 153 authors who have assumed a major role in no fewer than five retracted publications as recidivists. Although recidivists are all involved in scientific misconducts, their specific behavioral patterns are different. Table 4 lists the top 50 recidivists with the most retracted publications. Only the publications in which they acted as the main authors were included in the number of retracted publications while the time span included the years in which their publications were retracted. It can be seen that some recidivists were involved in long-term misconducts, from students to disciplinary heads of famous universities, such as Diederik A Stapel [13]. Some recidivists were concentrated in the short-term misconducts with a time span of mostly 2

to 5 years, which almost marked the evaluation period for tenure acquisition or title promotion. A few centralized misconducts were also associated with the journals. Due to the journals or relationships with specific authors, some authors were engaged in batch “irrigation-type” publications in these journals.

Table 4. Top 50 authors with the most retracted publications

Author	Number	Span	Author	Number	Span
Yoshitaka Fujii	116	11	Pattium Chiranjeevi	13	5
Joachim Boldt	44	26	Noriyuki Takai	13	11
Hua Zhong	41	3	Suresh Radhakrishnan	11	8
Shigeaki Kato	32	12	Gilson Khang	11	3
Jan Hendrik Schon	30	5	Javad Javanbakht	11	3
Naoki Mori	30	12	Denis Lima Guerra	11	4
Hyung-In Moon	29	8	Yoshihiro Sato	11	14
Chen-Yuan Chen	29	5	Alirio J Melendez	11	10
Tao Liu	27	2	Li-Mao Wu	11	2
Cheng-Wu Chen	27	4	Chang-Suk Han	10	8
Diederik A Stapel	23	16	Jon Sudbo	10	5
Swen N Piper	20	12	Jagadeesh Sridhara Rao	10	8
Dipak K Das	17	11	Naohito Aoki	10	15
Scott S Reuben	17	12	Karel Bezouska	9	19
Wataru Matsuyama	16	5	Yuhji Saitoh	9	15
Bharat B Aggarwal	16	6	Judith M Thomas	9	8
Anil K Jaiswal	15	9	Mariko Tomita	9	6
Robert A Slutsky	15	3	Rony Seger	9	13
Khalid Zaman	15	4	Anil Potti	9	4
Friedhelm Herrmann	15	9	Mario Jose Abdalla Saad	9	17
Fazlul H Sarkar	15	8	John Roland Darsee	9	3
Jin-Quan Cheng	15	11	Ill-Min Chung	9	2
Larry R Pease	14	8	Eric J Smart	9	11
Ulrich Lichtenthaler	14	5	James E Hunton	9	16
Yan Sui	13	3	Elena Bulanova	9	11

Conclusions and Discussion

The number of retracted publications in the world and the mainland China has generally increased. Recidivists and large-scale centralized revocation have a great impact on the number of publication retractions. Since 2000, the retraction rate and the retraction ratio in the mainland China have been significantly higher than the average in the world. The Jinggangshan University Incident in 2010 and the concentrated publication retraction incident in recent years directly raised the number of retracted publications and the retraction rate. The countries with strong scientific research capabilities have more publications retracted, including traditional developed countries and emerging powers represented by China and India, as well as some countries and regions that play an important role in regional development.

The number of retracted publications demonstrates obvious disciplinary differences. Life science is a typical field frequent to publication retraction. In addition, interdisciplinary and application-oriented basic research will also lead to many problematic studies while producing significant innovative results. These disciplines include applied interdisciplinary subjects (such as materials science and environmental science) as well as basic interdisciplinary subjects (biology, chemistry, and physics). Although the number of overall retracted publications in the social sciences is not large, the retraction tendency of “hard disciplines” (such as psychological, management, and economic sciences) that focus on data and measurement is not low and deserves attention. Although many journals have retracted publications, most of them have retracted only a few publications, and thus the publication retraction was not common. Therefore, the focus should be placed on the minority of journals with high retraction rates. These journals with high retraction rates also exhibit

obvious stratified characteristics, involving both high-impact journals, a few journals with lax reviews and even regulatory loopholes, as well as irritation journals with large volumes, high fees, and low publication difficulties. The journal of anesthesiology also exists as a special type of journal.

According to the disciplinary distribution of the retracted publications, the institutions with high retraction rates can be divided into three types. The institutions with more and scattered disciplines are mainly state-level scientific research institutions that carry out disciplinary layout and (jointly) establish branches in various parts of the country, such as the Chinese Academy of Sciences and the French National Scientific Research Center. These institutions are small in number, which is related to their special research management system. Most institutions involve more but concentrated disciplines, or fewer and concentrated disciplines. The biomedical field at the disciplinary level and the recidivist phenomenon at the author level are the main influencing factors. Regarding the distribution of authors, the statistical data demonstrates two main phenomena: first-offense phenomenon in the number of authors and recidivist phenomenon in the number of publications. The recidivists with multiple publications retracted should be the focus. Although recidivists are all involved in scientific misconducts, their specific behavioral patterns are different. Some recidivists were involved in long-term misconducts while some recidivists were concentrated in the short-term misconducts.

This study contributed to maintain the comprehensiveness of research data through heterogeneous multi-source databases, which cover most of the international journal articles. Through the filtering at Web of science, the authoritativeness of research data was guaranteed as the included SCI-E and SSCI represent the mainstream of international natural science and social science research. Therefore, the informetric conclusions of this paper on the characteristics of the retracted publications are of great value for the theoretical research on scientific norms and the practices of scientific research. In the meanwhile, the authors will continue to analyze the reasons for the publication retraction in the future research, and provide more specific practical measures for researchers, journals and departments of scientific management based on the informative characteristics of the publications.

Acknowledgement

This research was financially supported by the High Level STI Think Tank Project (Doctor project): “Informetric analysis on the International Retracted publication From the Perspective of Scientific norm”. The project is granted by China Association for Science and Technology.

References

- [1] Editorial. Time for the Data to Speak. *Nature Biotechnology*. 35.8 (2017):689.
- [2] M. Cokol, O. Fatih, R. Rodriguez-Esteban. Retraction rates are on the rise. *EMBO reports*. 9.1 (2008): 2-2.
- [3] Van Noorden R. Science publishing: The trouble with retractions. *Nature*. 478.7367 (2011): 26-28.
- [4], M.L. Grieneisen, M.H. Zhang. A comprehensive survey of retracted articles from the scholarly literature. *PLoS one*. 7.10 (2012): e44118.
- [5] C.M. Chen. Science mapping: a systematic review of the literature. *Journal of Data and Information Science*. 2.2 (2017): 1-40.
- [6] C.M. Chen, Z.G. Hu, J. Milbank, et al. A visual analytic study of retracted articles in scientific literature. *Journal of the American Society for Information Science and Technology*. 64.2 (2013): 234-253.
- [7] E.S. Reich. *Plastic fantastic: How the biggest fraud in physics shook the scientific world*. Macmillan, New York, 2009.

- [8] V. Yank, R. Drummond. Disclosure of researcher contributions: a study of original research articles in The Lancet. *Annals of internal medicine*. 130.8 (1999): 661-670.
- [9] J.D. Wren, K.Z. Kozak, K.R. Jonathan D, et al. The write position. *EMBO reports*. 8.11 (2007): 988-991.
- [10] J.D. Frame. Mainstream research in Latin America and the Caribbean. *Interciencia*. 2.3 (1977): 143-148.
- [11] S.B. Demir. Predatory journals: Who publishes in them and why?. *Journal of Informetrics*. 12.4 (2018): 1296-1311.
- [12] U.M. McHugh, S.M. Yentis. An analysis of retractions of papers authored by Scott Reuben, Joachim Boldt and Yoshitaka Fujii. *Anaesthesia*. 74.1 (2019): 17-21.
- [13] D.M. Markowitz, J.T. Hancock. Linguistic traces of a scientific fraud: The case of Diederik Stapel. *PLoS one*. 9.8 (2014): e105937.