Shema, Hadas; Hahn, Oliver; Mazarakis, Athanasios; Peters, Isabella

Article — Published Version

Retractions from altmetric and bibliometric perspectives

Information - Wissenschaft & Praxis

Kontakt/Contact
ZBW – Leibniz-Informationszentrum Wirtschaft/Leibniz Information Centre for Economics
Düsternbrooker Weg 120
24105 Kiel (Germany)
E-Mail: info@zbw.eu
https://www.zbw.eu/de/ueber-uns/profil-der-zbw/veroeffentlichungen-zbw

Bibliometrie

Hadas Shema, Oliver Hahn, Athanasios Mazarakis* und Isabella Peters

Retractions from altmetric and bibliometric perspectives

Abstract: In the battle for better science, the research community must obliterate, at times, works from the publication record, a process known as retraction. Additionally, publications and papers accumulate an altmetric attention score, which is a complementary metric to citation-based metrics. We used the citations, Journal Impact Factor, time between publication and retraction and the reasons behind retraction in order to find determinants of the retracted papers´ altmetric attention score. To find these determinants we compared two samples, one of retractions with top altmetric attention scores and one of retractions with altmetric attention scores chosen at random. We used a binary choice model to estimate the probability of being retracted due to misconduct or error. The model shows positive effects of altmetric scores and the time between publication and retraction on the probability to be retracted due to misconduct in the top sample. We conclude that there is an association between retraction due to misconduct and higher altmetric attention scores within the top sample.

Descriptors: Altmetrics, Bibliometrics, Scientific misconduct, Publication

Dr. Hadas Shema, ZBW – Leibniz-Informationszentrum Wirtschaft, Forschungsgruppe Web Science, Düsternbrooker Weg 120, 24105 Kiel, E-Mail: h.shema@zbw.eu
Oliver Hahn, ZBW – Leibniz-Informationszentrum Wirtschaft, Forschungsgruppe Web Science, Düsternbrooker Weg 120, 24105 Kiel, E-Mail: O.Hahn@zbw.eu
*Kontaktperson: Dr. Athanasios Mazarakis, ZBW – Leibniz-Informationszentrum Wirtschaft, Forschungsgruppe Web Science, Düsternbrooker Weg 120, 24105 Kiel und Christian-Albrechts-Universität zu Kiel, Christian-Albrechts-Platz 4, 24118 Kiel, E-Mail: a.mazarakis@zbw.eu
Prof. Dr. Isabella Peters, ZBW – Leibniz-Informationszentrum Wirtschaft, Forschungsgruppe Web Science, Düsternbrooker Weg 120, 24105 Kiel und Christian-Albrechts-Universität zu Kiel, Institut für Informatik, Arbeitsgruppe Web Science, Christian-Albrechts-Platz 4, 24118 Kiel, E-Mail: I.Peters@zbw.eu

Rücknahmen wissenschaftlicher Publikationen aus altmetrischer und bibliometrischer Sicht

Zusammenfassung: Es erscheinen manchmal Publikationen, die den hohen wissenschaftlichen Standards nicht entsprechen. Solche werden über einen Prozess des Zurückziehens (engl. Retraction) aus dem Publikationsverzeichnis entfernt. Artikel akkumulieren einen altmetrischen Aufmerksamkeitswert, den Altmetric Attention Score. Wir nutzen Zitationen, Journal Impact Factor, die Zeit zwischen Veröffentlichung und Rückzug einer Publikation und den Grund der Rücknahme, um Determinanten des altmetrischen Aufmerksamkeitswerts für zurückgezogene Veröffentlichungen zu finden. Dazu verglichen wir zwei Stichproben, eine mit hohen altmetrischen Werten und eine zufällige Stichprobe. Ein binomiales Regressionsmodell ermöglicht die Schätzung der Wahrscheinlichkeit mit der eine Publikation wegen Fehlverhalten oder Fehler zurückgezogen wurde. In der Stichprobe mit hohen altmetrischen Werten fanden wir einen positiven Zusammenhang zwischen Altmetric Attention Scores, sowie Veröffentlichungszeit, mit der Wahrscheinlichkeit wegen Fehlverhalten zurückgezogen zu werden.

Deskriptoren: Altmetrics, Bibliometrie, wissenschaftliches Fehlverhalten, Publikation

Un point de vue altmétrique et bibliométrique sur la rétraction de publications scientifiques

Résumé: Parfois, les publications ne répondent pas aux exigences scientifiques. Elles sont alors retirées après un certain temps et retirées du répertoire des publications (en anglais : Retraction). Toutefois, dans l'intervalle, elles ont déjà atteint une mesure altmétrique, l' « Altmetric Attention Score ». La recherche avait pour but de déterminer si une publication avait été retirée pour inconduite ou pour erreur. Nous utilisons les citations, le Journal Impact Factor, le temps écoulé entre la publication et le retrait d'une publication et la raison du retrait, pour trouver les facteurs qui influencent la valeur altmétrique des publications retirées. À cette fin, nous comparons deux échantillons, l'un avec des valeurs altmétriques élevées et un échantillon

aléatoire. Un modèle de régression binomial permet d'estimer la probabilité qu'une publication soit retirée pour cause d'inconduite ou d'erreurs. Dans l'échantillon de scores altmétriques élevés, nous avons trouvé une corrélation positive entre les « Altmetric attention Scores » et la probabilité de retrait pour cause d'inconduite ou d'erreurs.

Descripteurs: Almétrics, Bibliométrie, Inconduite scientifique, Publication

Introduction

To correct science, flaws in scholarly works must be discovered first. Ideally, the research community uses peer review to detect and correct flawed science prior to publication. However, when peer review fails to detect major flaws, the scientific record has to be corrected after publication.

The Committee on Publication Ethics (COPE) guidelines regarding retractions define the retraction as *"a mechanism for correcting the literature and alerting readers to publications that contain such seriously flawed or erroneous data that their findings and conclusions cannot be relied upon."* (Wager, Barbour, Yentis, & Kleinert, 2009, p. 532). Additionally, COPE advices the consideration of retraction in cases of plagiarism, redundant publication, or report of unethical research (Wager et al, 2009).

In recent years, some of the gatekeepings of science have been done by online reviewers, using general social networks as well as designated platforms for post-publication review. Among the platforms available for such activities are the blog Retraction Watch, which routinely covers retractions and their causes, and PubPeer, a post-publication peer review platform where readers can comment anonymously. On a number of occasions, papers have been corrected or retracted following PubPeer comments (see, among others, Palus (2016); Daley (2018); Extance (2018)).

In a very well-known case, social media played an important role in exposing the fraud of the stimulus-triggered acquisition of pluripotency (STAP) cells, following the publication of two papers on the subject in the scientific journal Nature. Pubpeer discussions compared these two papers to a paper by the same author from 2011 and showed that they contained duplications of the same images (Cyranoski, 2014). Additionally, the studies were extensively discussed in Twitter and blogs. A study of the Twitter activity related to STAP cells showed that discussions of possible misconduct took place on Twitter before the discussions in the mainstream media (Sugawara et al., 2017).

These cases and others show the role social media can play in self-correcting of science. However, this role has not been yet studied in full. We chose to examine retractions and focus on determinants of the altmetric attention score, which is an aggregated attention metric to a publication in social media. While many retractions did not get much attention, we have few publications which attract a lot of attention. We conclude the main reason behind very high altmetric attention scores lies in the reason for retraction itself.

Literature review

In the review we will discuss first the use of social media by researchers as represented by alternative metrics (altmetrics). Second, we will discuss the characteristics of retractions and the patterns they follow.

Altmetrics

A growing number of researchers use social media tools in their professional lives. Among these tools we can find general social networks, such as Facebook and Twitter, as well as academy-designated tools along the lines of the reference manager Mendeley and the social network ResearchGate (Van Noorden, 2014; Lemke et al., 2017). Considering these developments, it is no wonder that alternative or complementary metrics of science, based on social media, have become part of the scientific landscape.

Costas and colleagues (Costas, Zahedi, & Wouters, 2015) found that citation, altmetric and journal impact factor scores have a positive, though moderate correlation. Shema, Bar-Ilan, and Thelwall (2014) showed that citing in blog posts correlates with a future higher level of citations. Recently Thelwall and Nevill demonstrated that altmetric scores, particularly those including Mendeley, can predict future, long term citation counts (Thelwall & Nevill, 2018).

There is some evidence that online attention relates to self-correcting procedures in scholarly literature. Brookes (2014) has compared papers whose data integrity has been questioned in public and those whose data integrity has been questioned, but the doubts have not been made public. He found that *"public papers were retracted 6.5-fold more, and corrected 7.7-fold more, than those in the private set."* (Brookes, 2014, p. 1).

Retractions

Perhaps the first English-language retraction in history took place in 1756 (Oransky, 2012). The earliest retraction noted by PubMed (at the time of writing) is for a 1959 paper, and was published in 1966 (Goldstein & Eastwood, 1966). The retraction is thought to be "*an emerging institution that renders scientific misconduct visible*" (Hesselmann, Graf, Schmidt, & Reinhart, 2017, p. 815) and "*a window into the scientific process*" (Oransky & Marcus, 2010).

In the last two decades, the number of retractions has risen at a rate far exceeding the growth in the total number of published articles (see review in Hesselmann et al., 2017). Grieneisen and Zhang (2012) showed that "*The number of articles retracted per year increased by a factor of 19.06 from 2001 to 2010,*" (Grieneisen & Zhang, 2012, p. 1).

Many of the articles containing errors or misconducts have not been retracted so far; in a study of image duplication in over 20,000 life science articles, around 4 percent were found to include "*inappropriately duplicated images*" (Bik, Casadevall, & Fang, 2016). These findings show that the number of retractions can potentially increase in the years to come, depending on the retraction criteria.

The increased number of retractions has not necessarily resulted from an increase in misconduct by scientists, but from increased awareness of misconduct among journal readers and editors (Fanelli, 2013). Fanelli sees the increase in retractions as "*extremely positive changes*" (Fanelli, 2013, p. 6).

In practice, however, some retracted papers remain part of the scientific record. A recent study of ScienceDirect, Elsevier's full text database, showed that retracted papers still accumulate citations after their retraction (Bar-Ilan & Halevi, 2018).

Patterns of retraction

Discipline, the journal impact factor, country of origin and other factors have all been connected with retractions. Ribeiro and Vasconcelos (2018) studying around 1600 retractions covered in the blog Retraction Watch, found that most (63 %) of them came from the biomedical, medical and clinical sciences. 85 percent of the retractions came from 15 countries, with the United States and China accounting together for about 41 percent. Grieneisen and Zhang (2012) found that the percentage of retractions "*in the broad fields of Medicine, Chemistry, Life Sciences and Multidisciplinary Sciences*" (Grieneisen & Zhang, 2012, p. 6) was higher in comparison with their percentage at the Web of Science records.

The journal's impact, as presented by the Journal Impact Factor (JIF) is in certain cases correlated with retractions. A "retraction index" created for leading journals' retractions between 2001-2010, showed a positive relationship between the impact factor and the frequency of retractions in these journals (Fang & Casadevall, 2011). Furthermore, retraction of highly-cited articles occur more frequently (Furman, Jensen, & Murray, 2012).

Also, Fang et al. (2012) found fraud, suspected fraud or error as causes for retraction which correlated with the journal impact factor. However, duplicate publication or plagiarism were only slightly correlated with the JIF.

Additionally, the time between publication and retraction has grown shorter over the years; Bar-Ilan and Halevi (2017) found that the average time between publication and retraction in a sample of 820 Elsevier journals retractions was 2.5 years. Moylan and Kowalczuk (2016), who studied retractions between 2000-2015, found that the time between retraction and publication was slightly less than a year.

Steen and colleagues found that the time between publication and retraction has become shorter for articles published after 2002 averaging about 24 months until retraction in comparison to articles published prior to 2002, which averaged about 50 months time-to-retraction. (Steen, Casadevall, & Fang, 2013).

Finally, the time between publication and retraction depends on the cause of retraction. Fang, Steen, and Casadevall (2012) showed that retraction due to fraud averages 46.8 months from publication to retraction, while plagiarism and error averaged only 26 months.

Error or misconduct?

Earlier retraction studies, based solely on retraction notices, found that many, if not most, retractions were the results of errors or results that could not be reproduced (61.8 % in Nath, Marcus, and Druss (2006), 40 % in Wager and Williams (2011)). However, more recent studies classify most retractions as results of misconduct (Fang et al., 2012) and showed that in some retraction notices, the retraction seems to be the result of honest errors, while in reality it is the result of misconduct. After classification of over 2,000 retraction notices indexed by PubMed, they found that "*three-quarters were retracted because of misconduct or suspected misconduct, and only one-quarter was retracted for error.*" (Fang et al., 2012, p. 17028).

Misconduct is not clearly defined in literature. While one study defines misconduct as fabrication, falsification, plagiarism, image/data manipulation and faked data/re-
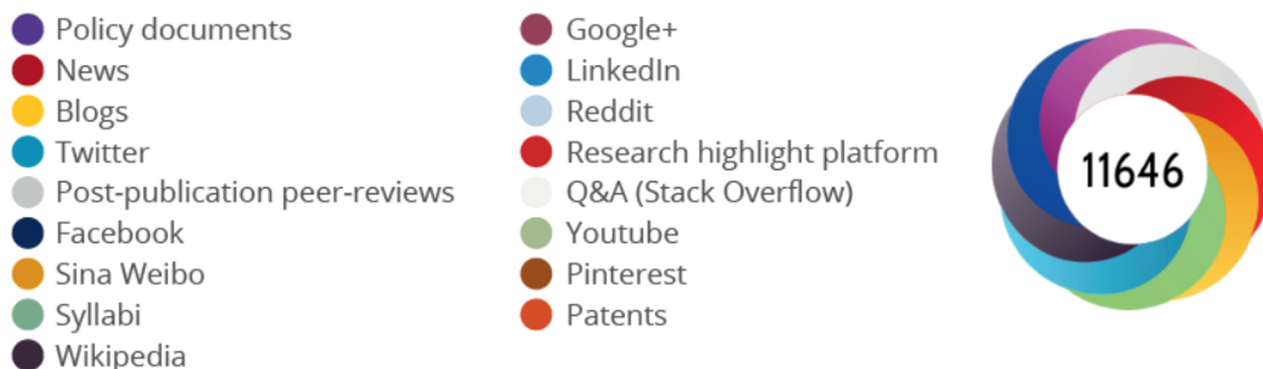
- Policy documents
- News
- Blogs
- Twitter
- Post-publication peer-reviews
- Facebook
- Sina Weibo
- Syllabi
- Wikipedia

- Google+
- LinkedIn
- Reddit
- Research highlight platform
- Q&A (Stack Overflow)
- Youtube
- Pinterest
- Patents

11646

Figure 1: The altmetric donut and its sources ("The donut and Altmetric Attention Score – Altmetric," n.d.).

sults/figures, classifying other problematic issues such as forged authorship and faked peer review as "other" (Ribeiro & Vasconcelos, 2018), others choose to see all categories that are not honest error as misconduct (Moylan & Kowalczuk, 2016).

In Moylan and Kowalczuk (2016)'s study misconduct was the reason behind the retraction of 76 percent of their sample. They found that compromised peer review, plagiarism and data falsification or fabrication were the leading causes of retractions. A study of Chinese retractions found that the most common reasons for retractions of Chinese articles were plagiarism, error, duplicate publication, fake peer review process and authorship dispute (Chen, Xing, Wang & Wang, 2018). The percentage of misconduct and error depends on the definition of each and on the sources studies rely upon in their classification.

In our study, we examine the characteristics of retractions which have been featured prominently in the social media and those of retractions that were chosen in random from a bigger population of retractions which received altmetric attention. We are interested in finding determinants of altmetric values of retracted articles with a focus on explaining extraordinary high altmetric attention scores. A simple binary choice model will be introduced as an approach for measuring the influence of variables on the probability of being retracted by misconduct or error.

Methods

We analyzed two samples, one which includes 100 retractions with high altmetric attention scores and a second sample which includes 100 retractions with random altmetric attention scores. Our variables are the altmetric attention score, number of citations, Journal Impact Factor, the number of days between publication and retraction and the reason for retraction. We begin by investigating

basic descriptive statistics of both samples to describe general differences.

Our focus on the retraction and the relation between its cause and the altmetric scores has led us to create a binary choice model, in order to determine whether the variables affect the probability to be retracted due to misconduct or error. We used the exact Wilcoxon signed rank test for investigating differences in mean. For correlations we relate to the correlation coefficient of Spearman, because we want to minimize the effects of extreme values.

Data description and our two samples

The sample in the study is based on the matching of data from PubMed and Altmetric.com. As mentioned in the literature review, many retractions take place in the life science and medicine disciplines, which is why we relied on data from PubMed, an "*Online version of Index Medicus produced by the US National Library of Medicine (NLM)*" which covers over 25 million records (Rickman, n.d.).

Altmetric.com is a prominent company in the field of altmetrics and its altmetric attention index incorporate blogs, news, social networks, post-publication peer review sites, Wikipedia and more into one score presented as a colorful "donut". The altmetric attention score is "*an automatically calculated, weighted count of all of the attention a research output has received,*" ("The donut and Altmetric Attention Score – Altmetric," n.d.) (Fig. 1). We use the terms altmetric score, altmetric attention score, altmetric value, altmetric attention index or altmetrics synonymously.

We have downloaded the data of retracted publications from PubMed for publications with an official publication and retraction date between January 1, 2012 and August 2, 2017. We focused on data from 2012 onwards because Altmetric.com has started its data gathering on July 2011. The
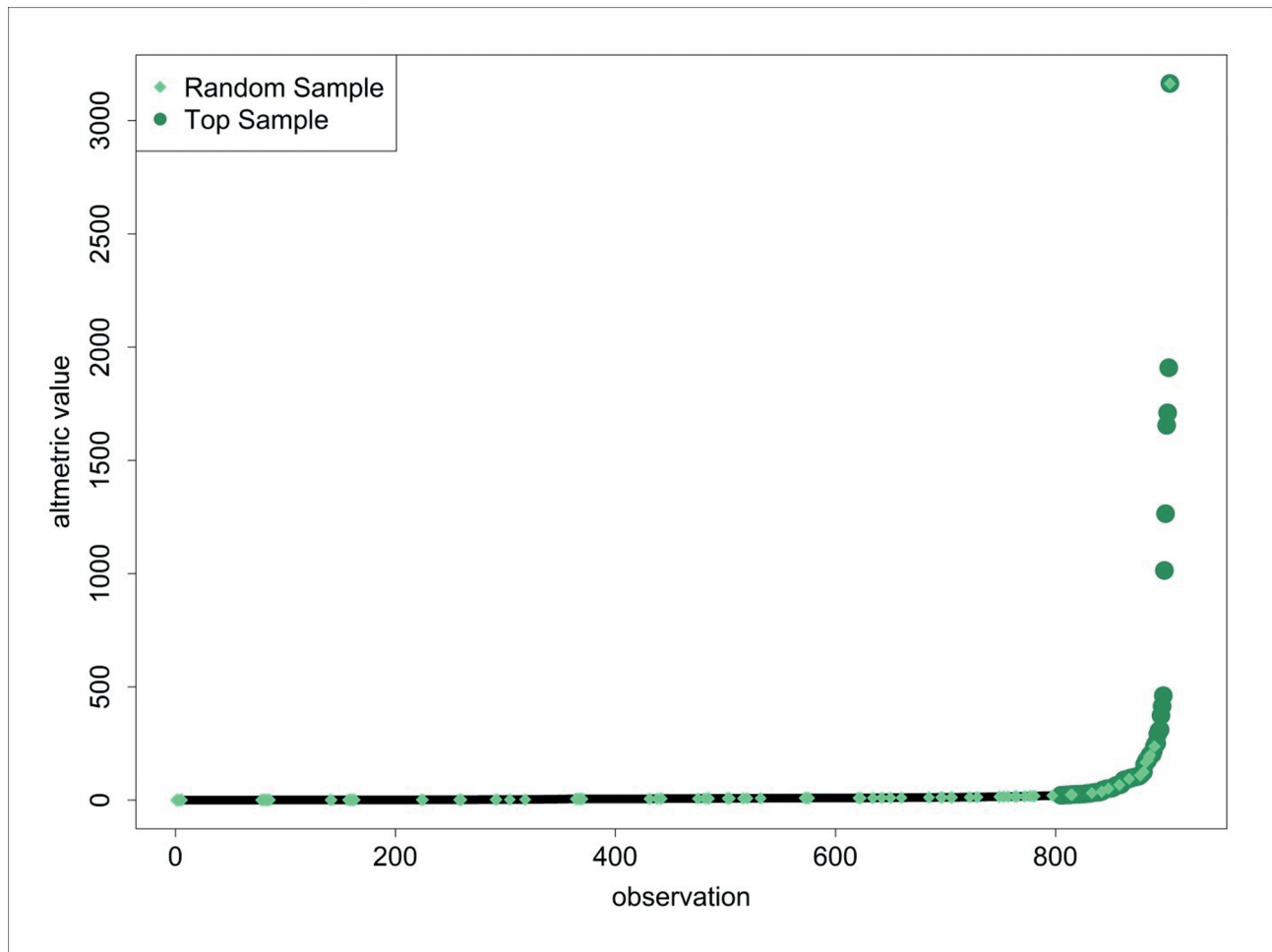
Figure 2: Altmetric attention score for each of the PubMed retractions from PubMed.

PubMed data included 1700 retractions, out of which we have matched altmetric data for 904 publications.

Figure 2 depicts all 904 publications sorted, starting with the lowest altmetric value. Every point represents one of our 904 observations. As seen, altmetric values increase linearly until a threshold, where further increase of altmetric values becomes exponential. We intend to find possible reasons for the dissonance hypothesizing that publications with very high altmetric attention scores and publications with random scores differ in other aspects as well. This motivated us to create the two samples for our analysis.

We decided to take a sample of those papers with top altmetric values from the point where we believed the threshold lies. This decision was based on visual analysis. We chose 100 retractions, which we considered to be the top sample and are represented by dark green dots in Figure 2. We also chose at random 100 retractions shown in Figure 2 as light green diamonds. Sixteen observations appear in both samples. We collected the altmetric scores using the R package "rAltmetrics" (Ram, 2017) in September 2018.

Normally, altmetrics do not decrease over time except in rare occasions, (e.g. if a Twitter post was deleted) ("Why has the Altmetric Attention Score for my paper gone down? : Altmetric Support," n.d.). This also applies for retracted publications. A crucial assumption is that further increase of the altmetric scores is neglectable, given that enough time has passed after the retraction. We assume a diminishing interest in that paper over time, so that there is only a small, if at all, increase of the altmetric score after our data collection. Therefore, we sampled only retractions which have been already been retracted for more than a year at the time of our analysis.

The same reasoning applies for the number of citations. There is evidence of retracted articles being cited after retraction (Bar-Ilan & Halevi, 2018), but the growth of the accumulated number of citations slows down over time. Nevertheless, we consider papers accumulating citations after retraction a limitation.

We collected the number of citations to the retractions in the sample which were published between the years 2012–2017 using the "All Databases" search. Those cita-

tions are from Web of Science, collected between the 24th of July 2018 and the 1st of August 2018.

When observing the time between publication and retraction, we collected the publication date ahead of print rather than on the official publication date, since altmetric indices begin to accumulate from the moment an article with an object identifier, such as DOI, is online.

We collected the publication and retraction dates using PubMed. In cases where PubMed did not offer an exact date, we searched the publisher´s website for the dates. In cases where there were no clear publication and/or retraction dates, but only mention of a month, we took the 15th of the month as the date. In case of a version publication, as in F1000 papers, we took the online publication date of the first version. The time is measured in days.

The Journal Impact Factor was measured in 2016 and data came from Web of Science. Eight retractions in the top sample and eleven retractions in the random sample were published in journals which were not indexed in Web of Science and therefore did not have an impact factor.

Retraction classification

As we have noted in the literature review, the definitions of misconduct and error change from one study to another. In this study, we have chosen to classify as misconduct the articles, where there has been an intention of one author or more to deceive the readers and as error, where it was clear that the articles' flaws were unintentional. We also classified the retractions inside the "misconduct" category according to the type of misconduct which took place (e.g. plagiarism).

We classified the retractions according to their retraction notices, Retraction Watch and the Retraction Watch database entries, if those exist. Two coders each classified the samples separately, after which the classifications were compared and differences have been discussed until agreement has been reached. In cases where agreement between the coders could not be reached, the retractions have also been classified as "unclear".

The "unclear" category was also used when the two coders both considered the reason behind the retraction as unknown, based on our sources, and the "other" category was created for cases where the article has been retracted not due to error or misconduct of the authors but different reasons (such as concerns of legal actions against the journal unless the article is retracted, though the article itself is considered valid). The "influenced by a third party" category was used for misconduct where the authors have not been at fault, such as a commentary piece which has

been retracted because the article it commented about was fraudulent.

Binary Choice Model

The binary choice model is a regression in which the dependent variable is a nominal variable. To create the model we limited our samples to all retractions which were retracted because of misconduct or error, leaving out the reasons "other", "unclear" and "influenced by third party". So, the reason of retraction has been reduced to a binary variable which is encoded as 1 if the reason for retraction is "misconduct" and 0 for publications which got retracted because of "error".

This is possible as we focus only on the main reason for retraction; even though the occurrence of both reasons simultaneously is generally possible, if misconduct happened, it is usually the main reason for retraction. If error has been identified as the main reason, then misconduct did not happen. Therefore, we can use a binary choice model where "misconduct" and "error" are mutually exclusive.

Having this binary variable implemented as the dependent variable, we estimated if the probability of being retracted by misconduct increases or decreases due to the impact of an independent variable, also known as regressor. As we use a complementary log-log link function, which is a non-linear function, only average partial effects are given. The real effect varies depending on the regressors' values. For this reason, we focus on the direction of the effect and on the statistical significance of the estimators. The regressors are the number of citations, the Journal Impact Factor and the number of days between publishing and retraction. We conducted this regression separately for each of our two samples.

Results

Descriptive Statistics

In the following Tables (1-4) we will compare some descriptive statistics of the two samples. In Table 1, the altmetric values of the random sample range from a minimum of 0.25 to a maximum of 3163.53. The median is 7.54; even though the random sample includes the retraction with the highest altmetric value, the sample's mean is 53.57. There are many observations with low altmetric values and only a few with a high altmetric value, therefore the skewness is high (9.54) and the standard deviation is 317.49 (Tab. 1).

The top sample altmetric attention scores range between 21.25 and 3163.28. The difference between the top value in the samples is due to the data collection of the top sample being a few days after the random sample, nevertheless it is the same retraction. Compared to the random sample, the top sample has a higher median ($Mdn = 61.57$) and a higher mean ($M = 193.48$). The standard deviation increased, and the skewness is smaller than in the random sample (4.48) (Tab. 1). The difference in means between the samples is statistically significant ($Z = 7.06$, $p < .001$). We expected those differences due to the sampling process. Nevertheless, a sample similar to the top sample could have been chosen at random. The statistical significance indicates that the random sample drawn is indeed different than the top sample.

Table 1: Altmetric attention scores, random and top sample.

Altmetrics	Minimum	Median	Mean	Maximum	SD	Skewness
Random	0.25	7.54	53.57	3163.53	317.49	9.54
Top	21.25	61.57	193.48	3163.28	446.98	4.48

The measurements of citations, collected from Web of Science (WoS), are shown in Table 2. Citations in the random sample have accumulated between 0 and 140 citations, the median being 5.50 citations and the mean being 11.89 citations per retraction. The random sample deviates on average about 19.05 citations from the mean and has due to the lower bound of zero citations a positive skewness (4.07) (Tab. 2).

While the lowest number of citations in the top sample is also 0, the retraction with the most citations in the top sample has 265 citations. Like the altmetric values, the number of citations in the top sample has a lower positive skewness (2.89) and the median is higher ($Mdn = 13.50$) than in the random sample. The top retractions accumulated on average 29.08 citations, more than double the random sample's average. The top sample's standard deviation is 38.78 (Tab. 2). The difference between the samples' means are again statistically significant ($Z = 3.989$, $p < .001$).

Table 2: Web of Science citations, random and top sample.

Citations	Minimum	Median	Mean	Maximum	SD	Skewness
Random	0	5.50	11.89	140.00	19.05	4.07
Top	0	13.50	29.08	265.00	38.78	2.89

Table 3 shows similar differences between the samples using the variable "Journal Impact Factor". The random sample's Journal Impact Factor range is between 0.73 and 44.41. 44.41 is also the highest JIF in the top sample, and the minimum top sample JIF is 1.27. The random sample's median is lower ($Mdn = 3.65$) than the median of the top sample ($Mdn = 6.48$), and the random sample's mean is also lower ($M = 7.47$) than the mean of the top sample ($M = 14.61$). The standard deviation of the top sample is 14.43, compared with the standard deviation of 10.83 in the random sample. The skewness is smaller in the top sample (0.97) than the one in the random sample (2.65). Here the difference between the means is significant as well ($Z = 3.562$, $p < .001$).

Table 3: Journal Impact Factor, random and top sample.

JIF	Minimum	Median	Mean	Maximum	SD	Skewness
Random	0.73	3.65	7.47	44.41	10.83	2.65
Top	1.27	6.48	14.61	44.41	14.43	0.97

So far, we have shown rather large differences between the samples, but the differences between the samples for the variable "days being published", on the other hand, are less obvious (Tab. 4). Both samples include a single paper which was only published for two days. The longest duration of publication before retraction in the random sample was 1859 days, and in the top sample, 1818 days. The median of the random sample is higher ($Mdn = 450$) when compared with the 428.5 days median of the top sample. However, the mean of "days being published" is higher in the top sample ($M = 548.40$) days, compared to 512.20 days of the random sample. The standard deviation of the random sample is smaller ($SD = 359.76$) when compared with the top sample ($SD = 447.85$) and the skewness is higher, with the random sample having skewness of 1.06 and the top sample having skewness of 0.85. The average deviation from the mean is 359.76 compared with 447.85 in the random sample. The skewness is lower in the top sample (0.85) compared with 1.06 in the random sample. The difference between the samples mean for this variable is not statistically significant ($Z = 0.246$, $p = 0.807$).

Table 4: Days between publication and retraction, random and top sample.

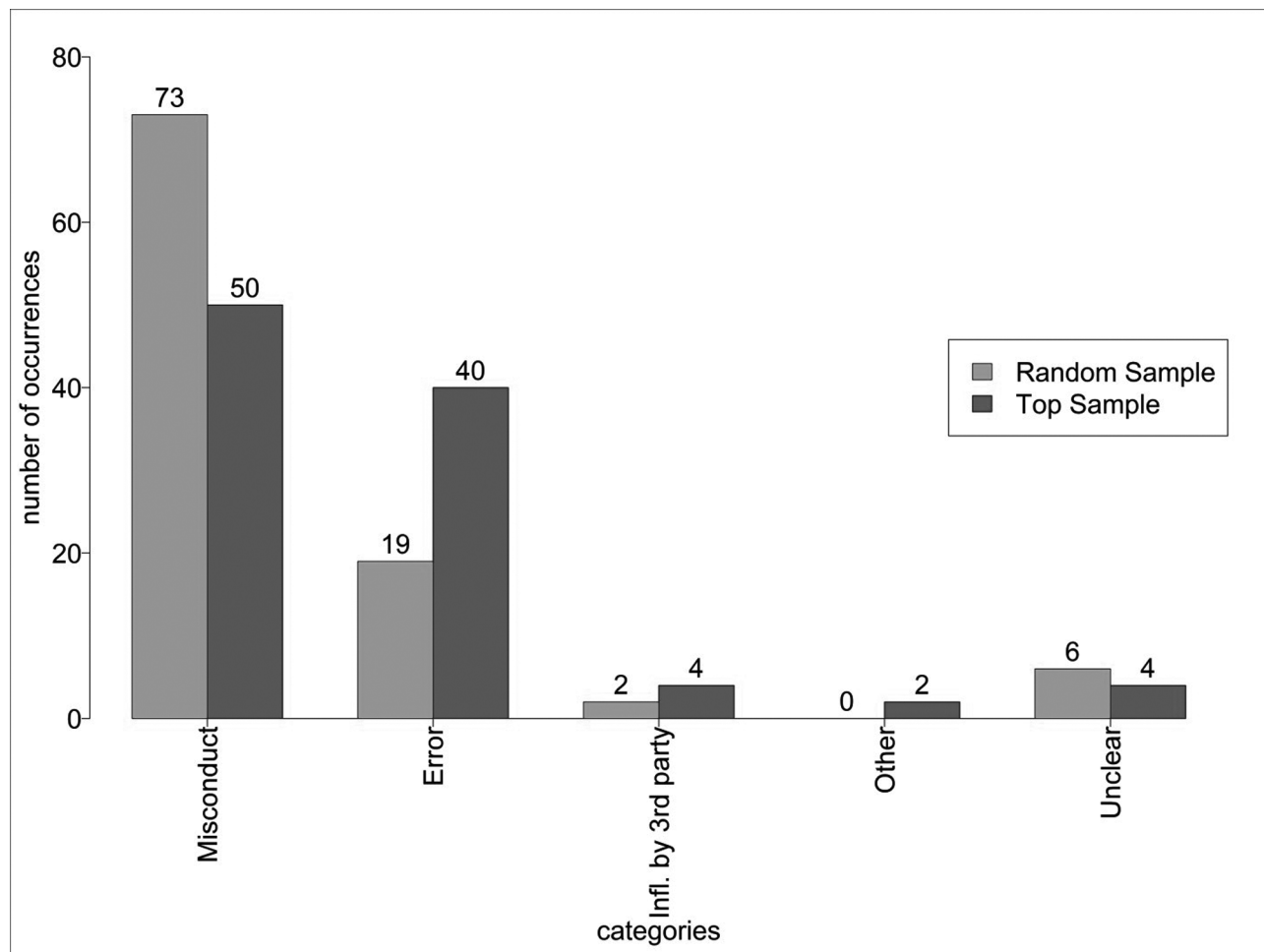| Days being published	Minimum	Median	Mean	Maximum	SD	Skewness
Random	2.00	450.00	512.20	1859.00	359.76	1.06
Top	2.00	428.50	548.40	1818.00	447.85	0.85

Figure 3: A comparison between the random sample and the top sample of the occurrences of the reasons behind retractions.

As seen in the previous tables, the top sample has higher means for all variables as well as higher standard deviations and smaller skewness, when compared with the random sample. The only exception is the "days being published" variable, where values are similar.

The correlations between altmetric attention scores and citations ($r_s = .31$, $p = .002$) and between the altmetric scores and the Journal Impact Factor values ($r_s = .43$, $p < .001$) are statistically significant within the random sample. However, these correlations have no statistical significance at the 5 percent level within the top sample ($r_s = .10$, $p = .305$; $r_s = .19$, $p = .072$).

It appears that the correlations between the variables in the random sample disappear in the top sample. A possible reason is that citations and the journal impact factor are not following the extensive increase in altmetrics, such that even if there is positive correlation it does only apply to a bound of number of citations or level of Journal Impact Factor, which is not crossed even if the altmetric attention further increases. If such a bound is already reached before the altmetric values are high enough for

the top sample, we have no further significant correlation in the top sample.

We have also a lack of significant correlation in the random sample between "days being published" and altmetric attention scores ($r_s = -.08$, $p = .425$). Likewise, there is no statistically significant correlation between these variables in the top sample ($r_s = -.16$, $p = .110$).

The lack of correlation between "days being published" and "altmetrics" suggests that altmetric attention does not accumulate equally over time. Otherwise, there would not be any publications with high altmetric attention but short life spans. Therefore, we assume there are short time frames during the life cycle of retractions in which altmetric attention accumulation concentrates and which are independent of the life span's length. Publication and retraction are such possible events, because they happen in each retracted paper's life cycle.

Misconduct or error?

Misconduct was the main reason for retraction for 73 of the articles in the random sample, while error occurred in only 19 of the sample's papers. The "influenced by a third party" category included two articles, the "other" category none and for six papers the reason was "unclear" (Fig. 3).

In the top sample the main retraction reason for 50 of the articles was misconduct. Error was the reason behind the retraction of 40 articles. The "influenced by a third party" category included four articles, the "other" category two articles and the "unclear" category four retractions (Fig. 3).

73 percent of all observations in the random sample were retracted because of misconduct as the main reason. However, only half of all retractions in the top sample were retracted due to misconduct. The results of a chi-squared test for categorical data rejects the null hypothesis of independence (χ^2 =14.84, df = 4, p = 0.005), thus the different distribution of reasons between the two samples depends on the sample itself.

For further analysis we only consider "misconduct" and "error". Based on these data we carried out the binary choice estimation for both samples, so we can pinpoint the influence of the variables on the probability to be retracted by misconduct or error. If "altmetrics" has a significant influence on this probability, we have identified the reason for retraction as another factor affecting altmetric values. The dependent variable is now a binary variable, coded as "misconduct" = 1 and "error" = 0, therefore we measure the probability of being retracted because of "misconduct".

In Table 5 we present our results for the random sample. Altogether we ran three regressions per sample; the first regression includes all variables (altmetric attention score, the number of days between publication and retraction, the JIF and the number of citations) as independent variables. The second regression excludes the Journal Impact Factor and the third regression excludes citations. We ran regressions without those variables, because "citations" and "Journal Impact Factor" are significantly correlated in both samples and we expected both explaining partly the same. Therefore, in Table 5 there are four columns: the first column shows the independent variables; the second column, marked (1), shows the regression results with all variables included; the third column marked (2) shows our results excluding "Journal Impact Factor" and the fourth column shows the regression when "citations" are excluded. We repeat the exact analysis in Table 6 using the top sample data.

Table 5: Binary choice model results, random sample.

	Dependent variable		
	Reason		
	(1)	(2)	(3)
Altmetric score	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)
Citations WoS	-0.010	-0.019*	
	(0.013)	(0.011)	
JIF	-0.016		-0.024
	(0.018)		(0.016)
Days being published	0.001	0.001	0.001
	(0.0004)	(0.0004)	(0.0004)
Constant	0.215	0.328	0.180
	(0.278)	(0.247)	(0.278)
Observations	83	92	83
Log Likelihood	-41.767	-44.233	-42.052
Akaike Inf. Crit.	93.535	96.467	92.105
Note:	*p<0.1;	**p<0.05;	***p<0.01

We have not found any dependencies within the random sample between the altmetric attention scores and the probability of retracting a publication due to misconduct (Tab. 5). Even though there was a large percentage of retraction due to misconduct in the random sample, none of the variables are associated with the reason of retraction. When we exclude the "Journal Impact Factor" in our second regression, the number of citations is significant on the p = 10 percent level. Surprisingly, the effect's sign is negative. If the number of citations has any influence on our model in this sample, then an increase of citations means a lower probability of being retracted due to misconduct.

Table 6: Binary choice model results, top sample.

	Dependent variable		
	Reason		
	(1)	(2)	(3)
Altmetric score	0.002**	0.002**	0.001*
	(0.001)	(0.001)	(0.001)
Citations WoS	-0.013*	-0.013**	
	(0.007)	(0.006)	
JIF	-0.006		-0.019
	(0.014)		(0.013)
Days being published	0.002***	0.002***	0.001***
	(0.0005)	(0.0004)	(0.0003)
Constant	-0.901***	-0.958***	-0.695**
	(0.339)	(0.297)	(0.319)

Table 6: (continued)

	Dependent variable		
	Reason		
	(1)	(2)	(3)
Observations	83	90	83
Log Likelihood	-49.262	-53.449	-51.719
Akaike Inf. Crit.	108.524	114.898	111.437
Note:	*p<0.1;	**p<0.05;	***p<0.01

The results for our top sample are distinct from the random sample (Tab. 6), as we have found several estimators to be statistically significant. On the p = 5 percent level, the estimators "altmetrics" and "days being published" are significant and have a positive effect. The number of citations again is negative and significant on the p = 10 percent level. Leaving out "Journal Impact Factor" results in a significant negative estimator for "citations" on the p = 5 percent significance level. The JIF is not significant in any case.

The shortcoming of our model is simultaneity between independent and dependent variables, which impedes conclusions about causality. Considering the nature of our samples and the correlations we examined before, we still assume an increase of the altmetric values because of misconduct due to the following reasons:

We have found significant correlations within the random sample, but those disappear in the top sample. The binary choice model, on the contrary, shows significant estimators within the top sample, but not in the random sample. It seems that the Journal Impact Factor, as well as the number of citations, are drivers of the altmetric attention score when it is still low. Retracting a publication because of misconduct does not affect the altmetric value in the random sample, as we have no significant estimator in our binary choice model for the random sample.

For our top altmetrics sample, nevertheless, the JIF and the number of citations do not correlate significantly with altmetric values and cannot explain a further increase of such.

The binary choice model, on the other hand, reveals a significant influence of the retraction reason on altmetric values, while the JIF and the citations are either not significant or have a negative impact. Thus, retraction because of misconduct triggers a further increase of altmetric attention for publications which already accumulated altmetric attention before.

Another significant regressor is the number of citations. In this case, the effect is negative: Researchers tend to cite misconducts less than erroneous articles. Further research with panel data could reveal if we have discov-

ered an effect similar to that of Camerer and colleagues (Camerer et al., 2018), who found that researchers can predict reproducibility prior to the replication attempt. Our study can only show that a citation decrease is associated with retraction due to misconduct.

The JIF is not a significant regressor in any of the samples. Therefore, it does not affect the probability of an article to be retracted due to misconduct or error.

The "days being published" regressor is positive and significant in the top sample. As we have shown before, the average length of time between publication and retraction is similar in both samples (Tab. 4), but as the regressor is significant and positive in our binary choice model only for the top sample, we have higher average publication time for misconducts there.

Limitations and outlook

The first limitation of our study is the sample size. However, a bigger sample could transform the top sample into an average sample rather than one with exceptionally high altmetric values. Still, the sample size could be increased by including other disciplines in future studies.

Another limitation is simultaneity in our binary choice model, which not only impedes conclusions about causality but also our results because of homogeneity. Panel data, especially of altmetric values, would help to control for time, in order to infer causality and in order to seek additional variables which could decrease possible bias.

The study included articles that have been republished. Therefore, it is possible that their altmetric score might be related also to the new version. Altmetric.com recognizes social media attention by unique identifier numbers (e.g. DOI), but those frequently do not show, for example in news stories. Therefore, there is a chance that altmetric coverage for the new version will be added to the attention score of the old version and vice versa. We kept the republished papers since we did not want to bias the research towards misconduct (republished papers are usually retracted due to errors).

It is possible that other papers published during the years we sampled will still be retracted in the future and might receive greater altmetric attention than those in the current sample. It is also possible that papers in our sample have received additional attention and/or citations after our data collection. One way to check control this limitation is to replicate our study in three to five years again and to look for differences.

A comparison with articles which have not been retracted could give more insight of the altmetric attention

score distribution (s. Fig. 2). If altmetric values of papers which have not been retracted behave similarly, then determinants other than the reason of retraction are responsible for high altmetric scores.

Summary

In this study we examined two samples of retracted papers, one with top altmetric scores and one with random altmetric scores. They differ from one another in the result that the top group has significantly higher averages of citations and Journal Impact Factors. However, the average time between publication and retraction in both groups was similar. In the random group we have found correlations between the altmetric attention score and citations, as well as between the score and the JIF. These correlations were not present in the top sample.

Many retracted papers have low altmetric attention scores; however, after passing a threshold the altmetric scores sharply increase. This sudden exponential aberration is not explained by the number of citations, the JIF or the number of days between publication and retraction. Even though these variables create a basis, which is necessary for reaching very high altmetric attention, they are not the cause for the increased altmetric attention.

The lack of correlation between altmetrics and "days being published" could be the result of most altmetric attention accumulating during short, specific time frames instead of in a steady, slow increase. One possible time frame is around retraction, which we have examined more thoroughly. The main reasons of "misconduct" and "error" behind retractions were utilized in our binary choice model.

The binary choice model shows that in the top sample the higher the altmetric attention scores are and the longer the paper is published, the probability of being retracted for misconduct increases. It also shows that a higher probability is associated with a lower number of citations for the papers in question, and that the JIF is not an estimator for the reason of retraction.

We suggest that for papers already in the spotlights, which were retracted for misconduct, the impact represented by the altmetric attention score is different in nature than the impact represented by formal citations. The public interest in acts of misconduct does not automatically translate into formal influence on the scientific community, but rather the opposite, it is associated with decreased formal interest.

Our main finding is that in the case of papers in the top sample, which have been retracted because of misconduct, the public takes an unusual interest in the retractions. These are papers which have already received relatively high attention, because of their subject (e.g. genetically modified plants, vaccines), their mimetic value (Dawkins, 2006), being published by famous authors and journals and so forth. Then, when they are retracted for misconduct, their altmetric attention score increases, representing the public´s interest in scandals, so to speak. We show that being retracted due to misconduct drives already influential papers' altmetric attention scores even higher.

References

Bar-Ilan, J., & Halevi, G. (2017). Post retraction citations in context: a case study. *Scientometrics, 113*(1), 547–565. https://doi.org/10.1007/s11192-017-2242-0.

Bar-Ilan, J., & Halevi, G. (2018). Temporal characteristics of retracted articles. *Scientometrics, 116*(3), 1771–1783. https://doi.org/10.1007/s11192-018-2802-y.

Bik, E. M., Casadevall, A., & Fang, F. C. (2016). The Prevalence of Inappropriate Image Duplication in Biomedical Research Publications. *MBio, 7*(3). https://doi.org/10.1128/mBio.00809-16.

Brookes, P. S. (2014). Internet publicity of data problems in the bioscience literature correlates with enhanced corrective action. *PeerJ, 2*, e313. https://doi.org/10.7717/peerj.313

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., ... Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour, 2*(9), 637–644. https://doi.org/10.1038/s41562-018-0399-z.

Chen, W., Xing, Q.-R., Wang, H., & Wang, T. (2018). Retracted publications in the biomedical literature with authors from mainland China. *Scientometrics, 114*(1), 217–227. https://doi.org/10.1007/s11192-017-2565-x.

Costas, R., Zahedi, Z., & Wouters, P. (2015). Do "altmetrics" correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology, 66*(10), 2003–2019. https://doi.org/10.1002/asi.23309.

Cyranoski, D. (2014). Acid-bath stem-cell study under investigation. *Nature*. https://doi.org/10.1038/nature.2014.14738.

Daley, J. (2018). Another Retraction for Discredited Researcher. *The Scientist*. Retrieved from https://www.the-scientist.com/the-nutshell/another-retraction-for-discredited-researcher-36648 [8.3.2019].

Dawkins, R. (2006). The selfish gene: with a new introduction by the author. *UK: Oxford University Press.(Originally Published in 1976)*.

Extance, A. (2018). Indian institute investigates nanoscientists for indiscriminate image manipulation. *Chemistry World*. Retrieved from https://www.chemistryworld.com/news/indian-institute-investigates-nanoscientists-for-indiscriminate-image-manipulation-/3009159.article [8.3.2019].

Fanelli, D. (2013). Why Growing Retractions Are (Mostly) a Good Sign. *PLoS Medicine, 10*(12), e1001563. https://doi.org/10.1371/journal.pmed.1001563.

Fang, F. C., & Casadevall, A. (2011). Retracted science and the retraction index. *Infection and Immunity, 79*(10), 3855–3859. https://doi.org/10.1128/IAI.05661-11.

Fang, F. C., Steen, R. G., & Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. *Proceedings of the National Academy of Sciences of the United States of America, 109*(42), 17028–17033. https://doi.org/10.1073/pnas.1212247109.

Furman, J. L., Jensen, K., & Murray, F. (2012). Governing knowledge in the scientific community: Exploring the role of retractions in biomedicine. *Research Policy, 41*(2), 276–290. https://doi.org/10.1016/J.RESPOL.2011.11.001.

Goldstein, L., & Eastwood, J. M. (1966). On the primary site of nuclear RNA synthesis. A retraction. *The Journal of Cell Biology, 31*(1), 195. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/5971970.

Grieneisen, M. L., & Zhang, M. (2012). A Comprehensive Survey of Retracted Articles from the Scholarly Literature. *PLoS ONE, 7*(10), e44118. https://doi.org/10.1371/journal.pone.0044118.

Hesselmann, F., Graf, V., Schmidt, M., & Reinhart, M. (2017). The visibility of scientific misconduct: A review of the literature on retracted journal articles. *Current Sociology. La Sociologie Contemporaine, 65*(6), 814–845. https://doi.org/10.1177/0011392116663807.

Lemke, S., Mehrazar, M., Peters, I., Beucke, D., Gottschling, M., Krausz, A., ... Zagorova, O. (2017). Exploring the Meaning and Perception of Altmetrics. https://doi.org/10.5281/ZENODO.1037146.

Moylan, E. C., & Kowalczuk, M. K. (2016). Why articles are retracted: a retrospective cross-sectional study of retraction notices at BioMed Central. *BMJ Open, 6*(11), e012047. https://doi.org/10.1136/bmjopen-2016-012047.

Nath, S. B., Marcus, S. C., & Druss, B. G. (2006). Retractions in the research literature: misconduct or mistakes? *The Medical Journal of Australia, 185*(3), 152–154. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16893357 [8.3.2019].

Oransky, I, Marcus, A. (2010). Retraction Watch.

Oransky, I. (2012). The first-ever English language retraction (1756)? – Retraction Watch. Retrieved March 21, 2018, from https://retractionwatch.com/2012/02/27/the-first-ever-english-language-retraction-1756/.

Palus, S. (2016). Diabetes researcher logged 1 retraction, 3 correx, after PubPeer comments. Retrieved from https://retractionwatch.com/2016/07/04/diabetes-researcher-logged-1-retraction-3-correx-after-pubpeer-comments/[8.3.2019].

Ram, K. (2017). rAltmetrics. Retrieved from https://cran.r-project.org/web/packages/rAltmetric/README.html.

Ribeiro, M. D., & Vasconcelos, S. M. R. (2018). Retractions covered by Retraction Watch in the 2013–2015 period: prevalence for the most productive countries. *Scientometrics, 114*(2), 719–734. https://doi.org/10.1007/s11192-017-2621-6.

Shema, H., Bar-Ilan, J., & Thelwall, M. (2014). Do blog citations correlate with a higher number of future citations? Research blogs as a potential source for alternative metrics. *Journal of the Association for Information Science and Technology, 65*(5). https://doi.org/10.1002/asi.23037.

Steen, R. G., Casadevall, A., & Fang, F. C. (2013). Why has the number of scientific retractions increased?. *PloS one, 8*(7), e68397.

Sugawara, Y., Tanimoto, T., Miyagawa, S., Murakami, M., Tsuya, A., Tanaka, A., ... Narimatsu, H. (2017). Scientific Misconduct and Social Media: Role of Twitter in the Stimulus Triggered Acquisition of Pluripotency Cells Scandal. *Journal of Medical Internet Research, 19*(2). https://doi.org/10.2196/jmir.6706.

The donut and Altmetric Attention Score – Altmetric. (n.d.). Retrieved August 16, 2018, from https://www.altmetric.com/about-our-data/the-donut-and-score/[8.3.2019].

Thelwall, M., & Nevill, T. (2018). Could scientists use Altmetric.com scores to predict longer term citation counts? *Journal of Informetrics, 12*(1), 237–248. https://doi.org/10.1016/J.JOI.2018.01.008.

Van Noorden, R. (2014). Online collaboration: Scientists and the social network. *Nature, 512*(7513), 126–129. https://doi.org/10.1038/512126a.

Wager, E., Barbour, V., Yentis, S., & Kleinert, S. (2009). Retractions: Guidance from the Committee on Publication Ethics (COPE). *Maturitas*. https://doi.org/10.1016/j.maturitas.2009.09.018.

Wager, E., & Williams, P. (2011). Why and how do journals retract articles? An analysis of Medline retractions 1988-2008. *Journal of Medical Ethics, 37*(9), 567–570. https://doi.org/10.1136/jme.2010.040964.

Why has the Altmetric Attention Score for my .paper gone down?: Altmetric Support. (n.d.). Retrieved August 16, 2018, from https://help.altmetric.com/support/solutions/articles/6000108191-why-has-the-altmetric-attention-score-for-my-paper-gone-down- [8.3.2019].

Dr. Hadas Shema
ZBW – Leibniz-Informationszentrum Wirtschaft
Forschungsgruppe Web Science
Düsternbrooker Weg 120
24105 Kiel
h.shema@zbw.eu

2014 promovierte Hadas Shema in Information Science an der Bar-Ilan Universität. Zurzeit ist sie Postdoktorandin mit einem Alexander von Humboldt-Forschungsstipendium an der ZBW–Leibniz-Informationszentrum Wirtschaft.

Oliver Hahn
ZBW – Leibniz-Informationszentrum
Wirtschaft
Forschungsgruppe Web Science
Düsternbrooker Weg 120
24105 Kiel
O.Hahn@zbw.eu

Oliver Hahn ist wissenschaftlicher Mitarbeiter am ZBW Leibniz-Informationszentrum Wirtschaft. 2016 erlangte er den B.Sc. in Volkswirtschaftslehre an der Universität Duisburg-Essen und befindet sich zurzeit im Masterstudiengang Quantitative Economics an der Christian-Albrechts-Universität zu Kiel.

Dr. Athanasios Mazarakis
ZBW – Leibniz-Informationszentrum
Wirtschaft
Düsternbrooker Weg 120
24105 Kiel
und
Forschungsgruppe Web Science
Christian-Albrechts-Universität zu Kiel
Christian-Albrechts-Platz 4
24118 Kiel
a.mazarakis@zbw.eu

Athanasios Mazarakis ist wissenschaftlicher Mitarbeiter und Post-Doktorand für Web Science am ZBW Leibniz-Informationszentrum Wirtschaft und der Christian-Albrechts-Universität zu Kiel. In seiner Forschung befasst er sich mit Gamification, nutzergenerierten Inhalten, Open Access, Altmetrics, Science 2.0 und Open Science.

Prof. Dr. Isabella Peters
ZBW – Leibniz-Informationszentrum
Wirtschaft
Düsternbrooker Weg 120
24105 Kiel
und
Forschungsgruppe Web Science
Christian-Albrechts-Universität zu Kiel
Christian-Albrechts-Platz 4
24118 Kiel
I.Peters@zbw.eu

Isabella Peters ist Professorin für Web Science am ZBW Leibniz-Informationszentrum Wirtschaft und der Christian-Albrechts-Universität zu Kiel. In ihrer Forschung befasst sie sich mit nutzergenerierten Inhalten und wissenschaftlicher Kommunikation, Altmetrics, Science 2.0 und Open Science.