publications

MDPI

Article

Papers in and Papers out of the Spotlight: Comparative Bibliometric and Altmetrics Analysis of Biomedical Reports with and without News Media Stories

Artemis Chaleplioglou

Department of Archival, Library & Information Studies, University of West Attica, 12243 Athens, Greece; artemischal@uniwa.gr

Abstract: For decades, the discoverability and visibility of a paper relied on the readership of the academic journal where the publication was issued. As public interest in biomedicine has grown, the news media have taken on an important role in spreading scientific findings. This investigation explores the potential impact of news media stories on the citations and altmetrics of a paper. A total of 2020 open-access biomedical research papers, all published in the same year, 2015, and in journals with an impact factor between 10 and 14, were investigated. The papers were split into two groups based on the sole criterion of receiving or not receiving news media coverage. Papers with news media coverage accounted for 44% of the total. They received, on average, 60% more citations, 104% more blogs, 150% more X posts, 106% more Facebook reports, 40% more Wikipedia references, 85% more videos, and 51% more Mendeley readers than papers without news media coverage. The correlation between news media outlets and increased citations and altmetrics is evident. However, the broader societal impact of news media coverage, in terms of bringing scientific matters or discoveries to the public eye, appears to be more robust when compared to the reactions of the scientific community.

Keywords: news media; science communication; selective dissemination; citations; blogs; tweets; Facebook; Wikipedia; videos; Mendeley

1. Introduction

Recognising a scientific contribution by peers, as recorded by citations, is essential for a scientific author to support career development, increase status in the scientific community, and access financial support for research. International academic editions provide the means to disseminate research findings to the scientific community of specialists in the same topic or domain. However, in the vicinity of thousands or millions of papers, discoverability in the first place becomes crucial. The journal's title, where a scientific paper is published, its reputation, and its impact on the relevant research field are all determinants of citation counts [1]. Furthermore, the bibliometrics of a journal, as the number of articles published and their references and citations measured per year or other chronological period, represent the communication process and ranking among academic editions [2,3]. Of course, the quality of the work, content, importance, novelty, rationale, methodologies, and findings are the essence of an original scientific research paper. But if it is not searchable, discoverable, retrievable, or accessible to many peers, it will not be easily visible and ultimately citable [4].

Disruptive innovations in current awareness services for scientists, such as Current Contents [5], Chemical Abstracts [6], Medline [7], and the introduction of new publication system technologies, namely, electronic [8] and open-access [9] ones, have not only fundamentally changed the searchability and discoverability of papers but also increased the output of worldwide scientific productivity. In biomedicine and health sciences alone, more than seventeen million research papers have been published over the past decade. Together

with the thirty million preceding ones, they compete for citations when the yearly average citation output of the decade is more than fifty million [10]. The citation distribution in biomedicine is highly skewed [11], with almost 35% of the total biomedical papers without citations [10]. The percentage of documents per year without citation has increased for recent publications, with 42% for the year before last, 60% for last year, and 88% for this year [10]. This indicates that the intensity of competition for citations for recent paper publications represents a trend of increasing pressure.

The increasing enthusiasm of the public in the US for scientific advances at the turn of the 20th century and the insufficiency of journalism reporting on research results of this era led to the formation of the first scientific news media organisation, the publisher of Science News [12]. From this point forward, scientific journalism has expanded worldwide in news media agencies. Scientific sections enrich the content of newspapers, magazines, radio, and television. The Internet era brought news outlets and their stories online, accompanied by blogs and social media reposts [13]. The consumption of biomedical, clinical, biological, and health sciences breakthroughs through news media outlets is expected but is not free from exaggerations or distortions in the stories or the follow-up comments and people's views [14]. The media outlets' scientific coverage of biomedical sciences peaked during the COVID-19 pandemic when a wave of information was described as infodemic [15]. The haphazard accumulation of contradicting information derived from scientific communications through scholarly peer-reviewed papers and unreviewed preprints confused the public but, at the same time, exhibited the influential role of news media stories for the people [16]. However, is the impact of news media coverage of the same magnitude to the scientific community and the public?

Previous reports have applied different methodologies to assess press relations to scientific publications and their implication in scholarly communication and society. One approach is to focus on news media channels and their content. By analysing their content and their relations to scientific journal reports and topics of interest, conclusions can be extracted from their mutual interactions [17]. Another approach is topic-oriented, when news stories about a particular subject are analysed to track their source [18,19]. A third approach is to use social media to track their relationships with news and credible informational sources [15] or to check the scientific news background and quality by text analysis, keywords, or phrases in scientific literature [20]. This paper's scientific literature-driven approach was preferred as the starting point for data collection and analysis.

This work investigates whether the bibliometrics and altmetrics of biomedical papers in the news media outlets' spotlight significantly differ from those bypassed. This work is not focused on specific news media outlets or their types, commercial or non-profit, nor their scopes, general, political, business, scientific, educational, health, entertainment or lifestyle-oriented, fact-checking, or biases, but rather on the reference material, namely, the scientific papers reported. A decade-old aged and middle-to-top impact factor-matched, all open-access, bibliographical portfolio of biomedical papers was generated and served as a sample. The portfolio was divided into two subsets based on news media stories. In this paper, the identities or types of news media outlets were not considered because their complete, detailed information was unavailable in the database. The bibliometric and altmetrics of the two groups were compared to show that papers with news stories received more citations, blog reports, X posts, Facebook mentions, Wikipedia references, video references, and more Mendeley readers than their matching counterparts without news media attention. The magnitude of differences was substantially higher for all altmetrics variables compared to citations or Mendeley readers. This indicates that news media stories were inflated by the social media responses of the public and that the papers out of the news media spotlight were not negligible enough to be neglected by the media and left outside of public awareness. The invisibility of a scientific work from the public eye can affect public understanding and opinions. A merely informed society may face the same risks as a misinformed one. In both cases, it would be challenging to make wise decisions for policy

formulations in crucial public health issues [21]. Therefore, the journalist's newsworthiness criteria for a scientific report should be carefully considered and evaluated.

2. Materials and Methods

2.1. Journal and Articles Collection

To generate a homogenous bibliographical portfolio of biomedical papers of closely similar publication attributes by eliminating possible factors that could affect citations and altmetrics, it was decided: (a) to seek journals of the same impact factor rank by the time of their publication; (b) to look for a publication year, a time point old enough to follow-up citations and altmetrics [22] but within the altmetrics era [23], and not severely affected by the COVID-19 observed infodemic [15]; (c) to consider only All Open-Access publications to avoid possible distortions because of paywall accessibility restrictions; and (d) to consider only research papers, not reviews, news, views, or editorials. A preliminary analysis of news media stories' attention for publications in highly ranked scientific journals suggested the most appropriate segregating impact factor values from 10 to 14. The content of these middle-to-top-ranked journals received a 50:50 news media outlet's attention, according to their various journalism scopes. The publication era selected was 2015. The Clarivate Journal Citation Reports (JCR) Science Citation Index Expanded (SCIE) 2015 edition® was used as a reference. The indexed journals for the following 6 Web of Science Categories: Medicine, General and Internal; Medicine, Research and Experimental; Physiology; Biochemistry and Molecular Biology; Genetics and Heredity; and Neuroscience, were considered, and 26 journal titles were found within the pre-specified impact factor limits, as depicted in Figure 1. These titles were the following: *Neuron, Molecular Cell, Molecular Biology and Evolution, Journal of Clinical Investigation, Brain, Journal of Experimental Medicine, Nature Chemical Biology, Genes & Development, American Journal of Human Genetics, Nature Structural &Molecular Biology, Genome Research, PLOS Medicine, Molecular Psychiatry, Progress in Neurobiology, Trends in Biochemical Sciences, Trends in Neurosciences, Annual Review of Genetics, Acta Neuropathologica, Genome Biology, Progress in Lipid Research, Biological Psychiatry, Natural Product Reports, Annual Review of Medicine, Molecular Aspects of Medicine, Biological Reviews,* and *Molecular Systems Biology*. These journals were applied as Source Title to the Digital Science Dimensions® search query to retrieve all papers issued by them in 2015 with the additional pre-specified inclusion criteria of only research papers and only All Open-Access publications. The outcome included 18 source titles, namely, *Acta Neuropathologica, American Journal of Human Genetics, Biological Psychiatry, Brain, Genes & Development, Genome Biology, Journal of Clinical Investigation, Journal of Experimental Medicine, Molecular Aspects of Medicine, Molecular Cell, Molecular Psychiatry, Nature Structural & Molecular Biology, Neuron, PLOS Medicine, Progress in Lipid Research, Progress in Neurobiology, Trends in Biochemical Sciences,* and *Trends in Neurosciences*, with 2020 publications. All journals included in the study publish papers in the English language.

2.2. Data Extraction

The outcome of the bibliographic investigation was retrieved through the export results function of Digital Science Dimensions® by downloading the export full record comma-separated values and the export for the bibliometric mapping file. The bibliometric information for the number of citations received by a publication and its field citation ratio (FCR) and relative citation ratio (RCR) were retrieved with the export of the entire record file. FCR is the comparative ratio of the citations received by a publication to the average citations received by similarly aged publications of the same field. RCR indicates the relative citation performance of a publication to other publications in the same area of research. However, the altmetrics indicators were not directly extractable from the database. Therefore, the collective export full record extracted files were merged into a single spreadsheet file with additional altimetric columns as follows: Altmetric Attention Score®, as the weighted count of all online altmetrics calculated by Digital Science Dimensions®, News outlets, referred to news stories online, blogs references,

references in policy sources, X posts (former tweets), patent citations, mentions by peer review sites, Facebook mentions, Sina Weibo mentions, a Chinese microblogging service, Wikipedia, Google+, a social network that ceased operations in 2019, Reddit, videos from YouTube creators, research highlight platforms, such as faculty opinions, Mendeley readers, and CiteULike bookmarks. These data were directly accessed through the Altmetric® doughnut link accompanied by each publication record from Digital Science Dimensions® and manually copied to the spreadsheet file next to each paper title.
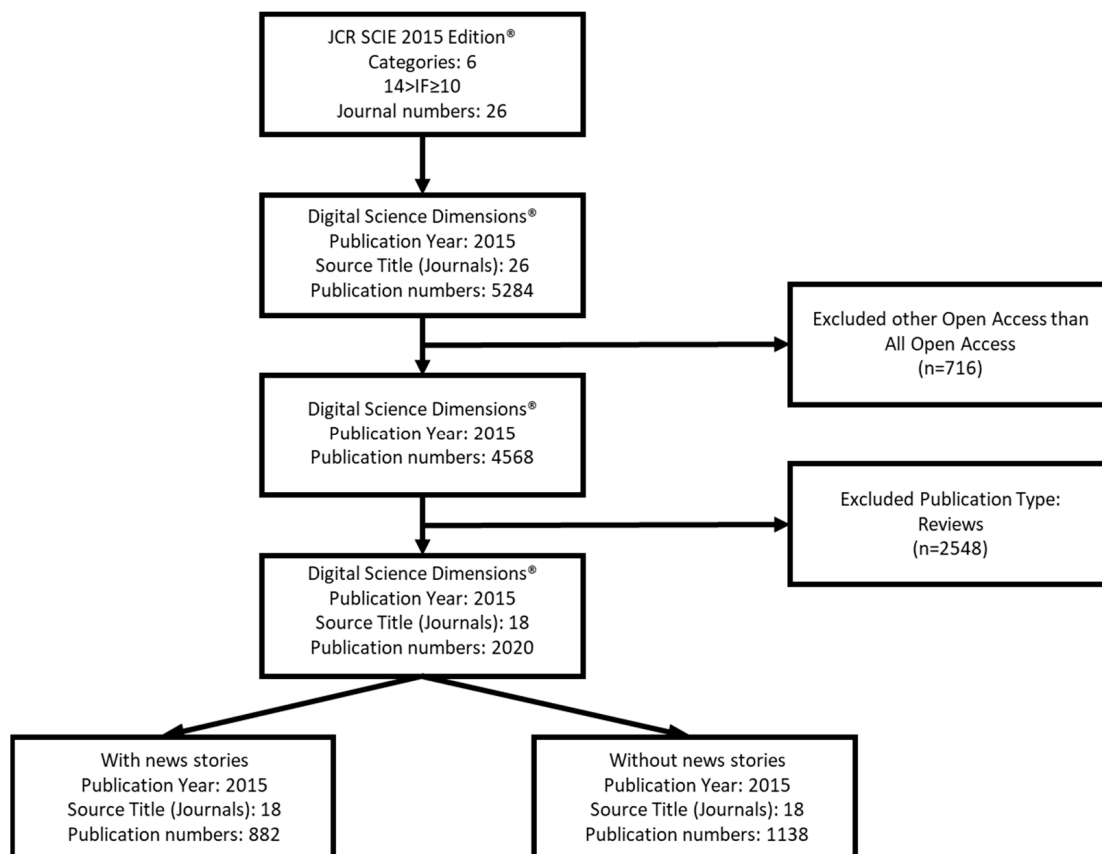
Figure 1. Inclusion criteria and numbers of included and excluded articles. JCR SCIE, Journal Citation Reports Science Citation Index Expanded.

2.3. Statistical Analysis

Descriptive statistics, box and whisker plots, agreement statistics on dichotomous ratings, and Pearson correlation tests were performed with Microsoft Excel® Version 2408 (Build 17928.20114). Likelihood ratio chi-square tests were performed to calculate odds ratios and their 95% confidence interval. Data are presented as mean $\pm$ standard deviation. A *p*-value of <0.05 was considered statistically significant. Power calculations for proportions were performed before considering biomedical literature associations with news stories and the strategic design of this observational study. For a difference of 5% of observed proportions of compared datasets of a sample size of 100 to 1000 at the 0.05 significant level with the Yates correction factor, the statistical power increased from 0.187 to 0.986. Therefore, subgrouping considerations that reduce the comparable sample size <100 should be considered carefully.

3. Results

To investigate the effects of news stories of scientific paper publications on their citations and altmetrics, a bibliographical portfolio of common attributes was generated, as depicted in the flowchart presented in Figure 1. The common qualities applied to achieve

as much uniformity as possible were for all works to be published in a cohort of biomedical journals with a journal impact factor between 10 and 14 by 2015, the year of their publication, and to be original research papers, all open access. The information about bibliometrics and altmetrics was manually collected from Digital Science Dimensions®, for each report that met the inclusion criteria. This 2020 original research biomedical report portfolio was divided into two groups of 882 published papers (44% of the total) and 1138 without (56% of the total) news stories. The publications with and without news stories are presented in Figure 2. The percentage of papers receiving news attention per journal varied from 30% (*American Journal of Human Genetics*) to 52% (*Acta Neuropathologica*). The documents with news stories received 5121 mentions by news outlets, on average 5.8 ± 9.9, with a lower quartile (25% percentile) of 1, a median of 3, and an upper quartile (75% percentile) of 7 mentions. The distribution of news stories per journal studied is presented in Figure 3. Except for journals with less than five publications with new stories, the rest exhibited similar news story distributions. Noteworthy, outliers receive significantly higher news stories far above the upper quartile of the journals' paper distribution. This suggests that these papers' unique individual characteristics, topic, title, authors, and content strongly contribute to the increased attraction of news outlets, which, for some, is accompanied by increased altmetrics.
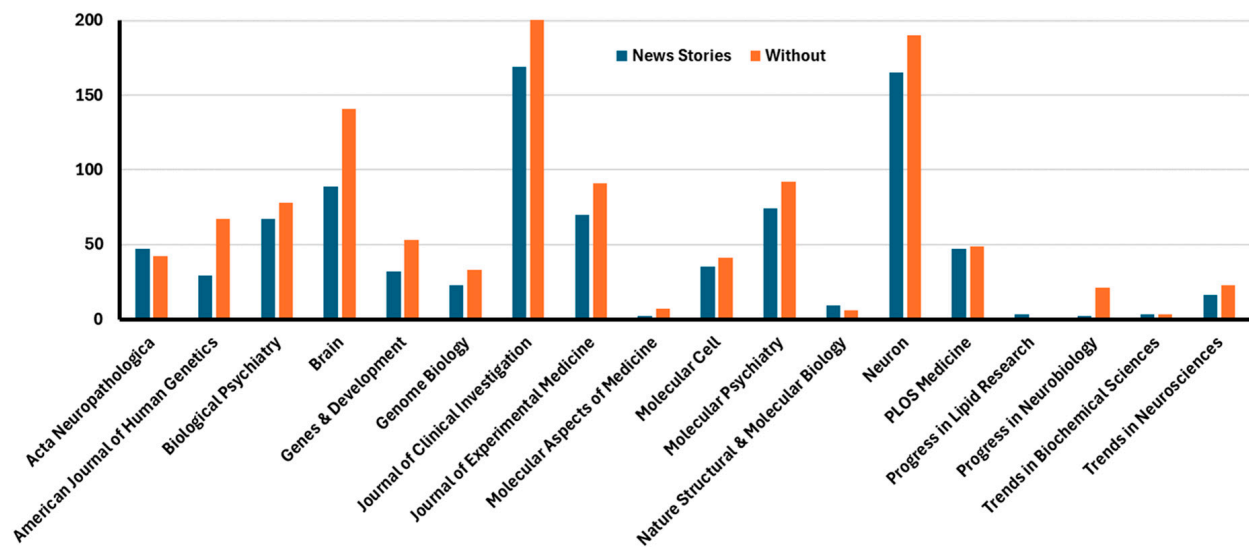
Figure 2. Original research, open-access papers with (blue) and without (orange) new stories per biomedical journal with an impact factor between 10 and 14 when published in 2015.

The number of papers with citations, FCR, RCR, Altmetric Score, blogs, policy sources, X posts, patent citations, peer review sites, Facebook, Weibo, Wikipedia, mentions in Q & A, Google+, Reddit, videos, faculty opinions, Mendeley, and CiteULike per group with and without news stories are depicted in Table 1. Nearly all papers in both groups with and without news stories received at least one citation, one X post, or one Mendeley reader, but other altmetrics variables differed significantly with 3.6-fold more papers with blog mentions, 3.7-fold more with policy sources, twice as many with Google+ mentions, 5-fold more with Reddit mentions, 2.6-fold more papers with videos in the news stories group, whilst the rest, namely, patent citations, Facebook, Weibo, Wikipedia, faculty opinions, and CiteULike, were pretty similar.
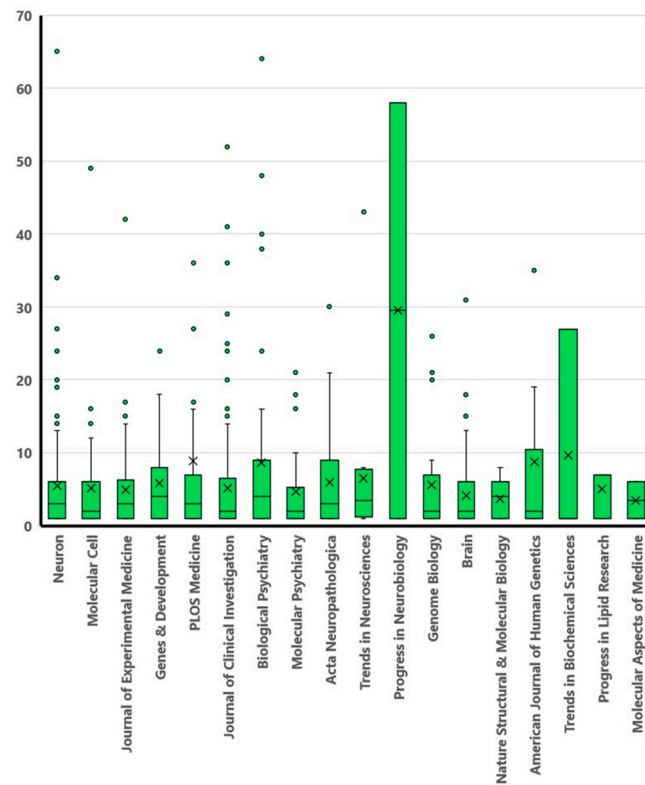
Figure 3. The distribution of news stories per journal. Circles (○) indicate outliers, and Xs indicate the mean markers. *Progress in Neurobiology*, *Trends in Biochemical Sciences*, *Progress in Lipid Research,* and *Molecular Aspects of Medicine* have less than five papers in the news stories group.

Table 1. Papers with at least one citation or altmetrics mention.

	Papers with News Stories	Papers without News Stories
Citations	875 (99.2% of the total)	1120 (98.4% of the total)
FCR	800 (90.7%)	978 (85.9%)
RCR	882 (100%)	1137 (99.9%)
Altmetrics Score	882 (100%)	1077 (94.6%)
Blogged	464 (53.7%)	171 (15.0%)
Policy sources	76 (8.6%)	26 (2.3%)
X posts	854 (96.8%)	1033 (90.8%)
Patent Citations	267 (30.2%)	238 (20.9%)
Peer review sites	16 (1.8%)	11 (1.0%)
Facebook	611 (69.3%)	837 (73.6%)
Weibo	27 (3.1%)	37 (3.3%)
Wikipedia	158 (17.9%)	200 (17.6%)
Mentions in Q & A	5 (0.6%)	3 (0.3%)
Google+	155 (17.6%)	96 (8.4%)
Reddit	76 (8.6%)	18 (1.6%)
Videos	57 (6.5%)	29 (2.5%)
Faculty Opinions	86 (9.7%)	111 (9.8%)
Mendeley	881 (99.9%)	1078 (94.7%)
CiteULike	247 (28%)	274 (24.0%)

The total number of these variables metrics per group is presented in Table 2. Papers with news media coverage exhibited a 60% increase in the average number of citations per paper, 70% higher average FCR, and 69% higher RCR. However, regarding the altmetrics these differences are much more significant with a 9-fold increase in the average altmetrics score for papers with news stories when compared to those without, 7-fold more blogs, 6-fold more policy sources, 2.7-fold more X posts, 50% more patent citations, 3.4-fold more

Facebook mentions, 5-fold more mentions in Weibo, twice more mentions in Wikipedia, 7-fold more mentions in Google+, 6-fold more mentions in Reddit, 6.5-fold more videos, and 60% more of the average Mendeley readers. The only variables with similar average values were peer review sites, mentions in Q & A, faculty opinions and CiteULike. These positive correlations between news stories, bibliometrics and altmetrics are further increased when the number of news stories per paper is considered. Papers that received news outlets' attention of the upper quartile and above, with at least seven news stories, received on average twice as many citations, blogs, X posts, Facebook mentions, Wikipedia references, videos, and Mendeley readers than papers of the lower quartile.

Table 2. A total number of citations or altmetrics mentions with descriptive statistics (mean ± standard deviation).
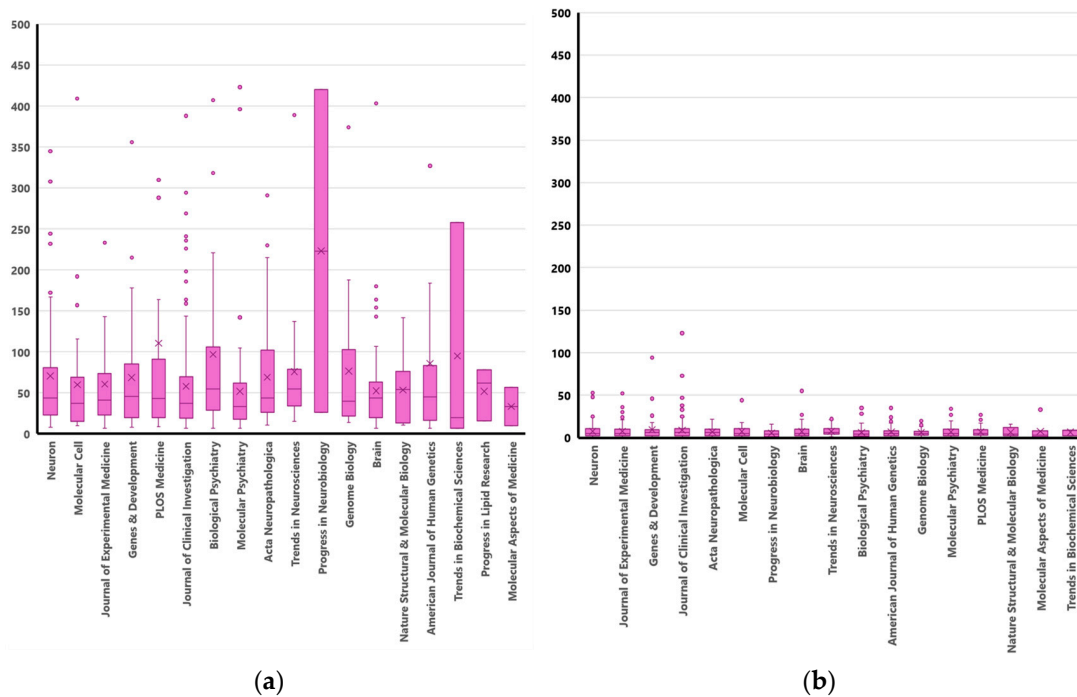
	Papers with News Stories	Papers without News Stories
Citations	135,184 (153.3 ± 188.3)	109,246 (96.0 ± 97.2)
FCR	20,174.3 (22.8 ± 31.8)	15,180 (13.3 ± 15.7)
RCR	4326.9 (4.9 ± 6.1)	3270.4 (2.9 ± 3.1)
Altmetrics Score	60,403 (68.5 ± 106.5)	8443 (7.4 ± 8.7)
Blogged	1260 (1.4 ± 2.2)	171 (0.2 ± 0.5)
Policy sources	104 (0.1 ± 0.4)	26 (0.0 ± 0.1)
X posts	18,423 (20.9 ± 70.7)	8904 (7.8 ± 11.4)
Patent Citations	1030 (1.2 ± 3.7)	905 (0.8 ± 3.0)
Peer review sites	16 (0.0 ± 0.1)	11 (0.0 ± 0.1)
Facebook	2150 (2.4 ± 4.9)	837 (0.7 ± 1.4)
Weibo	72 (0.1 ± 0.8)	37 (0.0 ± 0.5)
Wikipedia	368 (0.4 ± 1.8)	200 (0.2 ± 0.7)
Mentions in Q & A	5 (0.0 ± 0.1)	3 (0.0 ± 0.1)
Google+	577 (0.7 ± 4.6)	96 (0.1 ± 0.4)
Reddit	92 (0.1 ± 0.4)	18 (0.0 ± 0.1)
Videos	113 (0.1 ± 0.7)	29 (0.0 ± 0.2)
Faculty Opinions	86 (0.1 ± 0.3)	111 (0.1 ± 0.3)
Mendeley	196,970 (223.3 ± 236.5)	159,501 (140.2 ± 135.4)
CiteULike	390 (0.4 ± 0.9)	428 (0.4 ± 0.8)

The distributions of citations per journal with and without news stories are depicted in Figure 4. News stories affect the odds of receiving more citations for publications in the same journal and year, as indicated by the increase in the lower, median, and upper quartiles. The effect of news media stories on altmetrics scores is far more prominent as it is presented in the comparative depiction of the distributions of altmetrics scores per journal with and without news stories in Figure 5. Orders of magnitude increase in the lower, median, and upper quartiles of the distribution of altmetrics are evident for all journals for the papers that attract news outlets' attention compared to those without.

Notably, the odds ratio for increased citations, FCR, altmetrics score, blogs, policy sources, X posts, Facebook mentions, Weibo, Wikipedia, Google+, Reddit, videos and Mendeley readers of original biomedical research, all open-access reports with news outlets attention when compared to age- and attributes-matched reports, all published in the same cohort of journals with similar median to high impact factors, found to be statistically significant different as depicted in the agreement statistical analysis in Figure 6. The highest odds ratios were obtained for altmetrics score and Mendeley readers, with OR values of nearly 50 (95% CI, 7–357), followed by blogs with an OR 9 (95% CI, 7–11), and Reddit with an OR 6 (95% CI, 3.5–10). This observation suggests that the impact of news stories is more substantial for altmetrics when compared to bibliometrics.
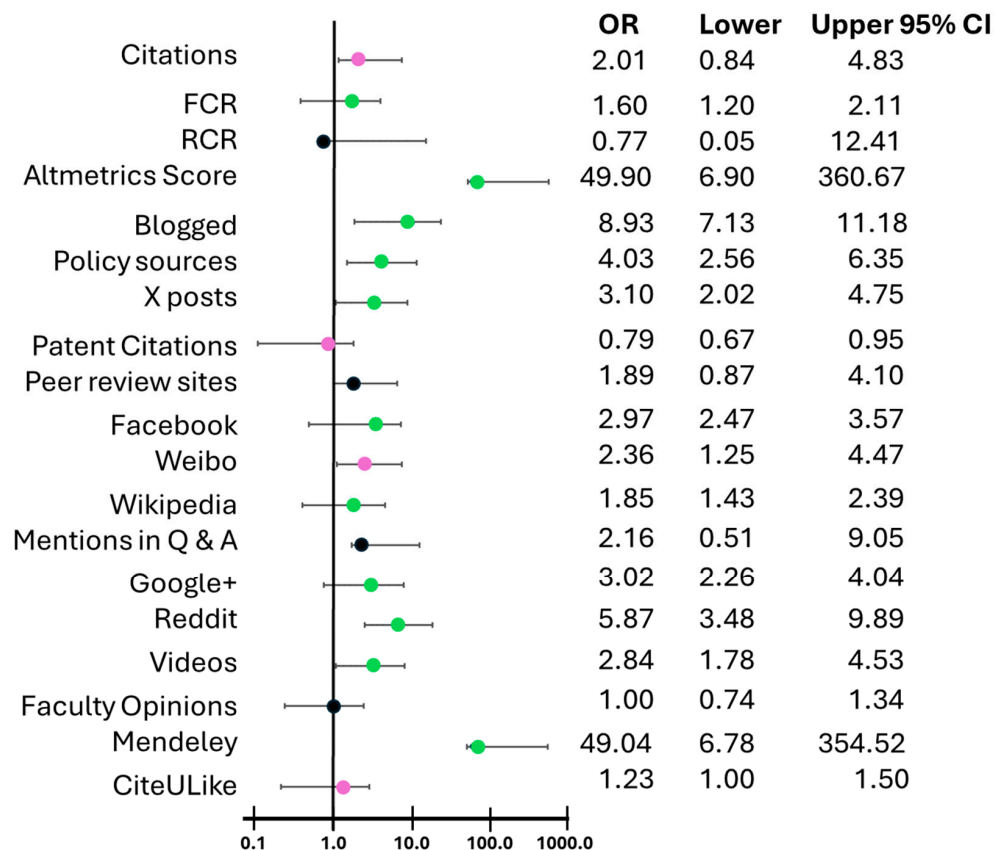
(**a**)

(**b**)
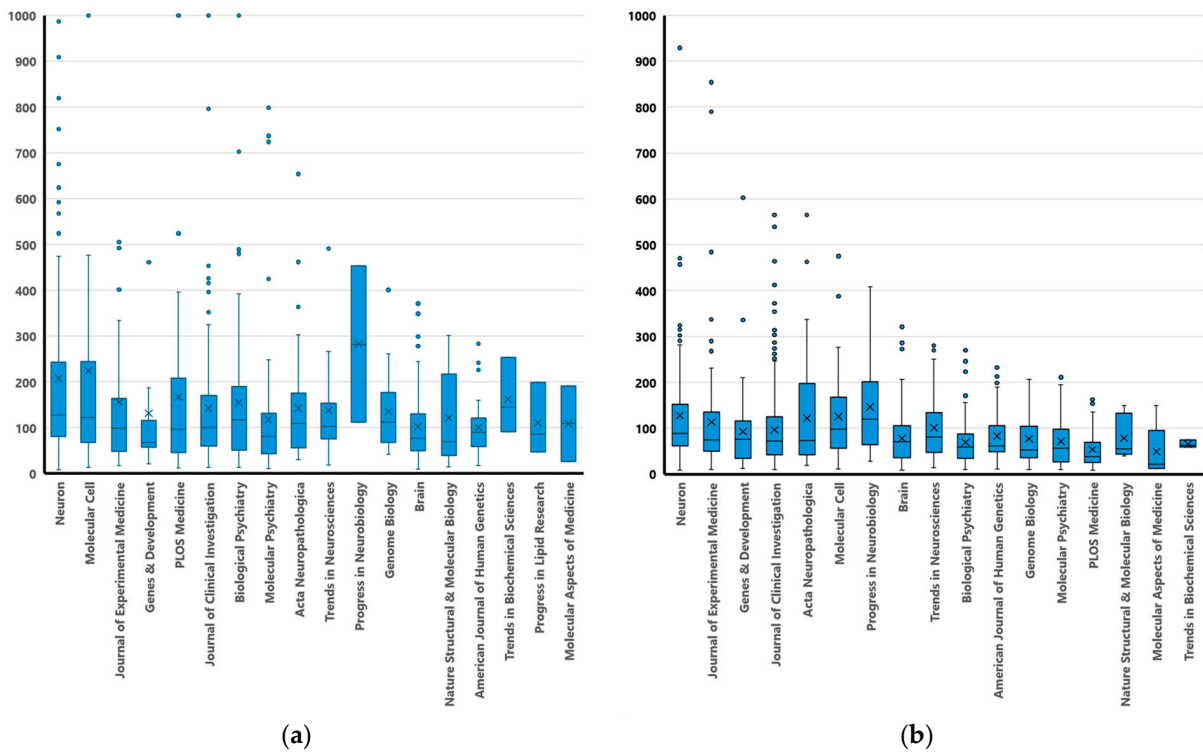
Figure 4. Distributions of citations per journal: (**a**) Papers with news stories; (**b**) Papers without news stories. Circles (○) indicate outliers, and Xs indicate the mean markers. *Progress in Neurobiology*, *Trends in Biochemical Sciences*, *Progress in Lipid Research,* and *Molecular Aspects of Medicine* have less than five papers in the news stories group, and *Progress in Lipid Research* and *Trends in Biochemical Sciences* in the control group.

(**a**)

(**b**)

Figure 5. Distributions of altmetrics scores per journal: (**a**) Papers with news stories; (**b**) Papers without news stories. Circles (○) indicate outliers, and Xs indicate the mean markers. *Progress in Neurobiology*, *Trends in Biochemical Sciences*, *Progress in Lipid Research,* and *Molecular Aspects of Medicine* have less than five papers in the news stories group, and *Progress in Lipid Research* and *Trends in Biochemical Sciences* in the control group.

|		OR	Lower	Upper 95% CI
Citations		2.01	0.84	4.83
FCR		1.60	1.20	2.11
RCR		0.77	0.05	12.41
Altmetrics Score		49.90	6.90	360.67
Blogged		8.93	7.13	11.18
Policy sources		4.03	2.56	6.35
X posts		3.10	2.02	4.75
Patent Citations		0.79	0.67	0.95
Peer review sites		1.89	0.87	4.10
Facebook		2.97	2.47	3.57
Weibo		2.36	1.25	4.47
Wikipedia		1.85	1.43	2.39
Mentions in Q & A		2.16	0.51	9.05
Google+		3.02	2.26	4.04
Reddit		5.87	3.48	9.89
Videos		2.84	1.78	4.53
Faculty Opinions		1.00	0.74	1.34
Mendeley		49.04	6.78	354.52
CiteULike		1.23	1.00	1.50

Figure 6. Odds ratios (OR) with a 95% confidence interval (lower and upper 95% CI) of receiving at least one citation or altmetrics mention when at least one new story has been received by a research paper (logarithmic scale). With violet statistical significance p-value$_{\text{(likelihood ratio chi-square)}} < 0.05$, with green, p-value$_{\text{(likelihood ratio chi-square)}} < 0.001$, with black, p-value$_{\text{(likelihood ratio chi-square)}} \geq 0.05$.

Pearson correlation analysis of paper citations, FCR, altmetrics score, news stories, X posts, Facebook, Mendeley, and patents citations for the complete bibliographic portfolio of 2020 reports revealed the correlations between these variables as indicated in Table 3 and Figure 7a. These correlations reflect linear associations between these parameters, as Figure 7b depicts. Citations correlate well with FCR and Mendeley but not with news stories, whilst news stories correlate with altmetrics score and, to a lesser extent, with X posts and Facebook mentions. The scatter plots and linear regressions of news stories exhibited slopes of 1.146 for citations, 0.653 for FCR, 10.015 for altmetrics score, 3.284 for X posts, 0.236 for Facebook, 6.180 for Mendeley, but 0.003 for patents citations. This is an additional indication of the impact of news stories on altmetrics when compared to bibliometrics.

Table 3. Correlation matrix (Pearson).

Variables	Citations	FCR	Altmetric	News	X posts	Facebook	Mendeley	Patents
Citations	1	0.900 *	0.222 *	0.204 *	0.099 *	0.112 *	0.827 *	0.257 *
FCR	0.900 *	1	0.214 *	0.191 *	0.109 *	0.146 *	0.801 *	0.244 *
Altmetric	0.222 *	0.214 *	1	0.935 *	0.695 *	0.600 *	0.281 *	0.027
News	0.204 *	0.191 *	0.935 *	1	0.492 *	0.484 *	0.232 *	0.007
X posts	0.099 *	0.109 *	0.695 *	0.492 *	1	0.470 *	0.168 *	−0.008
Facebook	0.112 *	0.146 *	0.600 *	0.484 *	0.470 *	1	0.142 *	0.002
Mendeley	0.827 *	0.801 *	0.281 *	0.232 *	0.168 *	0.142 *	1	0.194 *
Patents	0.257 *	0.244 *	0.027	0.007	−0.008	0.002	0.194 *	1

* Values are different form 0 with a significance level *alpha* = 0.05 (p-value$_{\text{(Pearson)}} < 10^{-4}$).
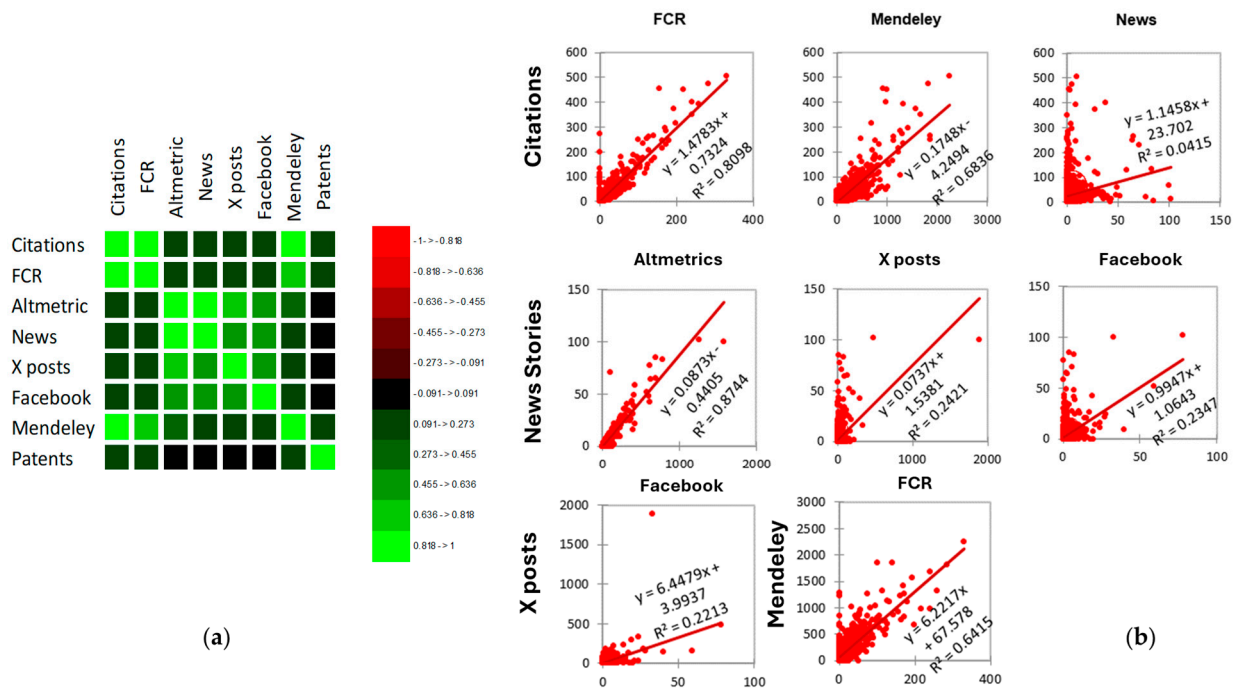
Figure 7. Pearson correlation and scatter plots with linear regression analysis of variables: (**a**) Pearson correlation analysis square of citations, FCR, altmetrics score, news stories, X posts, Facebook, Mendeley, and patents citations, with light green positive, almost linear correlation, dark green positive but less well-fitted to linear correlation, with black no correlation, with dark red negative but less well-fitted to linear correlation, and with red negative, almost linear correlation, according to the pallet of colour coding and Pearson correlation coefficient values presented aside; (**b**) Scatter plots with linear regressions of the papers examined from top to bottom, citations as a function of FCR (left), Mendeley readers (middle), and news stories (right), news stories as a function of altmetrics score (left), X posts (middle), and Facebook (right), and X posts as a function of Facebook (left) as well as Mendeley as a function of FCR (right) (n = 2020). The linear trendline is depicted in dark red along with the linear equation, and the R-squared goodness of fit calculated by regression analysis is presented in the scatter plots.

Collectively, these data suggest that news outlets' stories on original research paper publications are an independent factor correlated well with enhanced bibliometrics parameters but strongly associated with increased altmetrics variables.

4. Discussion

Quantifying multivariable-dependent trends in cross-sectional studies, such as the correlation of news stories with bibliometrics and altmetrics, can be particularly challenging. Therefore, this study focused on generating homogenous and well-matched groups of scientific publications to compare. A total of 2020 original biomedical research articles were investigated, and all open-access articles published within the same year in 18 journals had impact factors between 10 and 14. By selecting 2015 as the publication year, the recent effects of COVID-19 infodemics were also avoided [15]. This bibliographic portfolio was split into two groups of articles: those with and those without news stories. By controlling for the research field, year of publication, journal impact factor, and accessibility to readers, the effects of news stories were accessible.

The distribution of news stories per journal is similar except for journals with less than five papers with news stories. The outliers above the upper quartile of these distributions indicate journals with unique characteristics that explain the increased attention by news outlets and other altmetrics variables that reflect public opinion, reactions, or discussions on them. As recently postulated, a scientific publication communicating to different audiences,

peers, or the public may produce different responses [24]. This study sheds light on the differential dynamics of the impact of biomedical research on specialists and non-specialist audiences. News stories correlate with more citations, blogs, X posts, Facebook reports, Wikipedia references, videos, and Mendeley readers in biomedicine. However, their impact on papers' altmetrics is several times stronger than bibliometrics, as indicated by two lines of evidence: the collective data of a total number of citations or altmetrics mentions with their descriptive statistics and the box and whisker plots of paper distribution per journal according to citations or altmetrics score. Whilst the outliers indicate papers with unique characteristics that explain their high citability or altmetrics attraction state, the distributions per journal attest that grouping by news stories reflects actual differences in papers published in the same journal and year in bibliometrics and altmetrics. However, the predictive ability of news stories for citations or altmetrics variables differs significantly, as indicated by agreement statistics. According to this analysis, the predictive ability of at least one news story is good for overall altmetrics score, blogs, policy sources, X posts, Facebook, Wikipedia, Google+, Reddit, videos, and Mendeley readers. Therefore, the impact of news stories on social media appears to be significantly more potent than on the scientific audience. Although it has been shown that early Mendeley readers correlate well with later citation counts [25], the odds ratio of papers with news stories versus papers without suggests that the impact on citations is moderate compared to a 25-times more substantial effect on Mendeley readers. When considering that the publications explored are already ten years old, so there was enough time to receive citations, it may be concluded that the Mendeley readers may cite only a few from all the reports they accessed and read; in addition, the Mendeley readers population is separated by few expert scientific authors when compared to many young investigators that still do not produce their peer-review works.

Pearson correlations accompanied by linear regression analyses were performed to detect linear relationships between the variables examined. These analyses were not performed in groups but in the whole bibliographic portfolio. They showed that the bibliometric variables correlate with each other and Mendeley, and altmetrics variables correlate well but with lesser linear relationships rather than modifier factors. News stories correlate linearly with the overall altmetrics score, with a relationship of 1 news story to an increase of 10 in the altmetrics score. This finding suggests an amplification of news stories by altmetrics resources. There are also partial linear correlations between news stories with X posts and Facebook mentions but not with citations. This finding indicates that even though news stories are associated with more citations, this relationship is not linear, and the effect of outliers with individual characteristics that attract citations may be significant.

In the literature, the interplay between scientific papers and news media stories has been considered under different contexts and niches, but, in most cases, it involves the Internet and social media [26]. The contagion-like diffusion of information in social media can be attributed to network users or content derived from external sources like news outlets [27]. However, as superspreaders of information, penetrating and impactful newspapers or news channels may deliver erroneous messages due to citation bias, false consistency, lack of clarity and transparency, overgeneralisation, exaggeration, insulation, or noncredible sources [28]. Recently, it was shown that there is a positive correlation between the external popularity of research outside the scientific community by noteworthy media coverage and the number of scientific reports published after media coverage [29]. This investigation aligned the Altmetric® Mainstream-Media-Scores (MMS) with thematically similar articles from the PubMed database to obtain increased scientific publications on the same topic as a research article that received widespread news media coverage after this incident. This report did not examine the citations, bibliometrics, or other altmetrics but showed that news stories may influence scientific authors' research investigations. However, the trigger of scientific community attention was the extensive, exceeding a hundred news stories per paper, news media attention on a single paper. The data presented here for the bibliometrics of scientific articles with and without news stories agree with

this observation, but by considering a single news story of a paper. The average number of citations received after a single news story mention was found to be increased by 60% compared to a matching paper without a news story. This finding is important because only a tiny fraction, in this investigation, 0.1%, of the published scientific literature achieves over a hundred news media stories.

In a recent bibliometric and altmetrics investigation, Digital Science Dimensions® was used to generate a bibliographic portfolio of COVID-19 reports derived from Jordan and published between 2019 and 2022, which were subgrouped according to Altmetric Explorer® or Semantic Scholar® Highly Influential Citations (HICs) and compared [30]. The Semantic Scholar® HICs represent an AI-generated classification of citations regarding the context of the citations in a paper. There was a lack of correlation between Semantic Scholar® HICs and Altmetric® Attention Scores. This finding agrees with the Pearson correlation accompanied by linear regression analyses presented here, as citations and the Altmetric® Attention Score Pearson correlation coefficient value was 0.05. This is explained by the different motivations and criteria that define the citation, and the specific altmetrics-mentioned behaviour.

The data presented here concerning biomedicine suggest that news media are potent influencers of the scientific community and society and may generate trends in research investigations, funding flow, and public opinions and beliefs. This is important for all the stakeholders involved in scientific research, investigators, publishers, funders, policymakers, news media outlets, journalists, and the public, for introspection and careful consideration of these stimuli. News stories may deliver novel and groundbreaking science, but there is a risk of exaggeration, misinterpretation, or misleadingness. A critical perspective and cross-evaluation are necessary to extract conclusions and decide actions.

This study has some limitations that should be mentioned. Firstly, obtaining the flow of altmetrics variables in time series was impossible due to the platform restrictions. Secondly, selecting a publication year to achieve uniformity of citation and altmetrics information may mask longitudinal trends. Thirdly, this study did not consider the individual characteristics of each publication examined, such as topic, title, authors, affiliated institutions or countries, abstracts, or keywords. Fourthly, the criteria for journalists to select newsworthy research papers is out of the scope of this work. Fifth, this report could not obtain the full text of the news stories about the documents examined to comment on the manner of scientific information delivery or explanation by the news outlets. Sixth, the generalizability of these findings may be limited because of the enrolment of highly ranked biomedical scientific journals.

This is the first cross-sectional investigation of news stories' effects on bibliometrics and altmetrics in biomedicine, with uniformity of age, journal impact factor, and access-matched groups of papers. Future prospective studies would shed light on the timing of bibliometrics or altmetrics responses on a scientific paper upon news story release. Specific altmetrics indicators exhibit individual patterns of interest, converging or diverging from bibliometrics. Retrospective studies would demonstrate the socioeconomic determinants underlying these trends, expert opinions or stakeholders' perspectives against scientific investigations. These tendencies suggest social, economic, and political driving forces over science that may directly or indirectly influence scientific topic trends or research funding. Seeking specific thematic patterns or lemmas in scientific papers' titles, abstracts, or keywords that attract news media interest would be important to deliver. Its quantified observations provide a better understanding of the relationships between published research reports' audiences, scientific experts, and the public.

5. Conclusions

This cross-sectional controlled study demonstrated the strong impact of news media outlets on social media mentions and its close correlation with more citations for the mentioned papers in biomedicine. News media stories may serve as predictors for high altmetrics performance and, to a lesser extent, for bibliometrics. The relationship of news

stories with each social media resource suggests the co-existence of modifying factors possibly associated with each paper's individual characteristics. Outliers represent unique cases of attention attraction by the scientific community, the public or both. The critical appraisal and dissemination criteria of a recently published scientific work remain to be elucidated. Still, overall, the data presented suggest that any news media publicity impacts the bibliometrics and altmetrics of a biomedical original research paper.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the author on request.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Didegah, F.; Thelwall, M. Which factors help authors produce the highest impact research? Collaboration, journal and document properties. *J. Informetr.* **2013**, *7*, 861–873. [CrossRef]
2. Glänzel, W.; Moed, H.F. Journal impact measures in bibliometric research. *Scientometrics* **2002**, *53*, 171–193. [CrossRef]
3. Weingart, P. Impact of bibliometrics upon the science system: Inadvertent consequences? *Scientometrics* **2005**, *62*, 117–131. [CrossRef]
4. Baykoucheva, S. *Driving Science Information Discovery in the Digital Age*; Chandos Publishing: Hull, UK, 2021. [CrossRef]
5. Garfield, E. Current Contents: Its impact on scientific communication. *Interdiscip. Sci. Rev.* **1979**, *4*, 318–323. [CrossRef]
6. Dittmar, P.G.; Stobaugh, R.E.; Watson, C.E. The chemical abstracts service chemical registry system. I. General design. *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 111–121. [CrossRef]
7. Hristovski, D.; Stare, J.; Peterlin, B.; Dzeroski, S. Supporting discovery in medicine by association rule mining in Medline and UMLS. In *MEDINFO*; Series, Studies in Health Technology and Informatics; IOS Press: Clifton, VA, USA, 2001; Volume 84, pp. 1344–1348. [CrossRef]
8. Neavill, G.B. Electronic publishing, libraries, and the survival of information. *Libr. Resour. Tech. Serv.* **1984**, *28*, 76–89. Available online: http://digitalcommons.wayne.edu/slisfrp/47 (accessed on 13 September 2024).
9. Suber, P. Removing the barriers to research: An introduction to open access for librarians. *Coll. Res. Libr. News* **2003**, *64*, 92–94. Available online: http://nrs.harvard.edu/urn-3:HUL.InstRepos:3715477 (accessed on 13 September 2024). [CrossRef]
10. Digital Science & Research Solutions, Inc. Dimensions. Available online: https://app.dimensions.ai/ (accessed on 17 July 2024).
11. Albarrán, P.; Crespo, J.A.; Ortuño, I.; Ruiz-Castillo, J. The skewness of science in 219 sub-fields and a number of aggregates. *Scientometrics* **2011**, *88*, 385–397. [CrossRef]
12. Blum, D. Science journalism grows up. *Science* **2021**, *372*, 323. [CrossRef]
13. Segado-Boj, F.; Noguera-Vivo, J.M.; Hermida, A. Introduction. Users, content and platforms: A multidimensional approach to the research of news sharing. *Commun. Soc.* **2022**, *35*, 201–203. [CrossRef]
14. Brossard, D. New media landscapes and the science information consumer. *Proc. Natl. Acad. Sci. USA* **2013**, *110* (Suppl. S3), 14096–14101. [CrossRef] [PubMed]
15. Gallotti, R.; Valle, F.; Castaldo, N.; Sacco, P.; De Domenico, M. Assessing the risks of 'infodemics' in response to COVID-19 epidemics. *Nat. Hum. Behav.* **2020**, *4*, 1285–1293. [CrossRef] [PubMed]
16. West, J.D.; Bergstrom, C.T. Misinformation in and about science. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e1912444117. [CrossRef]
17. Lewison, G. Biomedical news items and feature articles in New Scientist magazine and the research papers that they cite. In Proceedings of the 12th Meeting of the International Society for Scientometrics and Informetrics, Rio de Janeiro, Brazil, 14–17 July 2009; pp. 719–729. Available online: https://www.issi-society.org/proceedings/issi_2009/ISSI2009-proc-vol2_Aug2009_batch1-paper-21.pdf (accessed on 13 September 2024).
18. Hubner, A. How did we get here? A framing and source analysis of early COVID-19 media coverage. *Commun. Res. Rep.* **2021**, *38*, 112–120. [CrossRef]
19. Fleerackers, A.; Riedlinger, M.; Moorhead, L.; Ahmed, R.; Alperin, J.P. Communicating scientific uncertainty in an age of COVID-19: An investigation into the use of preprints by digital media outlets. *Health Commun.* **2022**, *37*, 726–738. [CrossRef]
20. Smeros, P.; Castillo, C.; Aberer, K. Scilens: Evaluating the quality of scientific news articles using social media and scientific literature indicators. In Proceedings of the World Wide Web Conference, San Francisco, CA, USA, 13–17 May 2019; pp. 1747–1758. [CrossRef]
21. Nirenberg, M.W. Will society be prepared? *Science* **1967**, *157*, 633. [CrossRef]
22. Aksnes, D.W.; Langfeldt, L.; Wouters, P. Citations, citation indicators, and research quality: An overview of basic concepts and theories. *Sage Open* **2019**, *9*, 2158244019829575. [CrossRef]
23. Priem, J.; Taraborelli, D.; Groth, P.; Neylon, C. Altmetrics: A Manifesto. 2011. Available online: https://altmetrics.org/manifesto (accessed on 13 September 2024).

24. Torres-Salinas, D.; Docampo, D.; Arroyo-Machado, W.; Robinson-García, N. The many publics of science: Using altmetrics to identify common communication channels by scientific field. *Scientometrics* **2024**, *129*, 3705–3723. [CrossRef]
25. Thelwall, M. Early Mendeley readers correlate with later citation counts. *Scientometrics* **2018**, *115*, 1231–1240. [CrossRef]
26. Luc, J.G.; Archer, M.A.; Arora, R.C.; Bender, E.M.; Blitz, A.; Cooke, D.T.; Ni Hlci, T.; Kidane, B.; Ouzounian, M.; Varghese, T.K., Jr.; et al. Does tweeting improve citations? One-year results from the TSSMN prospective randomized trial. *Ann. Thorac. Surg.* **2021**, *111*, 296–300. [CrossRef]
27. Pei, S.; Muchnik, L.; Andrade, J.S., Jr.; Zheng, Z.; Makse, H.A. Searching for superspreaders of information in real-world social media. *Sci. Rep.* **2014**, *4*, 5547. [CrossRef] [PubMed]
28. Elson, M.; Ferguson, C.J.; Gregerson, M.; Hogg, J.L.; Ivory, J.; Klisanin, D.; Markey, P.M.; Nichols, D.; Siddiqui, S.; Wilson, J. Do policy statements on media effects faithfully represent the science? *Adv. Methods Pract. Psychol. Sci.* **2019**, *2*, 12–25. [CrossRef]
29. Leidecker-Sandmann, M.; Koppers, L.; Lehmkuhl, M. Correlations between the selection of topics by news media and scientific journals. *PLoS ONE* **2023**, *18*, e0280016. [CrossRef] [PubMed]
30. Dardas, L.A.; Sallam, M.; Woodward, A.; Sweis, N.; Sweis, N.; Sawair, F.A. Evaluating research impact based on Semantic Scholar highly influential citations, total citations, and altmetric attention scores: The quest for refined measures remains illusive. *Publications* **2023**, *11*, 5. [CrossRef]