The Impact of Preprints on the Citations of Journal Articles Related to COVID-19

Hiroyuki Tsunoda[1][0000-0002-6897-6611], Yuan Sun[2], Masaki Nishizawa[2], Xiaomin Liu[3], Kou Amano[2][0000-0002-8079-4941] and Rie Kominami[4][0000-0002-5085-3277]

[1] Chuo University, Hachioji-shi, Tokyo 1920393, Japan
[2] National Institute of Informatics, Tokyo 1018430, Japan
[3] National Science Library, Chinese Academy of Sciences, Beijing 100190, China
[4] Tsurumi University, Yokohama Kanagawa 2308501, Japan
tsunoda-h@tsurumi-u.ac.jp

Abstract. To investigate the impact of preprints on the citation counts of COVID-19-related papers, this study compares the number of citations received by drafts initially distributed as preprints and later published in journals with those received by papers directly submitted to journals. The difference in the median number of citations between COVID-19 preprint-distributed papers and COVID-19 directly submitted papers published in 184 journals was tested using the Mann-Whitney U test. The results showed that 129 journals had a statistically significant higher median citation count for COVID-19 preprint-distributed papers compared to directly submitted papers, with a p-value of less than 0.05. In contrast, no journals had a statistically significant higher median citation count for COVID-19 directly submitted papers. This indicates that 70.11% of the journals that published preprint-distributed papers experienced a significant increase in citations. We also identified that among the 184 journals, 13 journals garnered a substantial number of citations. Among the 74,037 COVID-19 papers, preprint-distributed papers (9,028) accounted for only 12.19%. However, among the 2,015,997 citations received by COVID-19 papers, preprint-distributed papers garnered 542,715 citations, representing a substantial 26.92%. These results suggest that distributing preprints prior to formal publication may help COVID-19 research reach a wider audience, potentially leading to increased readership and citations.

Keywords: Preprints, COVID-19, Citation, Bibliometrics.

1 Introduction

Preprints are openly accessible scientific papers distributed from an archive before they are submitted to a journal and undergo peer review. The value and importance of preprints have increased since they contributed to the public health emergency of the COVID-19 pandemic, and their role as a new means of academic information dissemination is now anticipated [1, 2, 3]. However, although preprints are increasing rapidly, they remain confined to certain regions, fields, and subjects, and have not yet spread

throughout the academic community [4]. One reason authors are hesitant to submit to preprints is the concern about publicly sharing research that has not undergone the peer review process. Moreover, many researchers are not even aware of preprints [5]. Some researchers have raised concerns about the quality of preprints since they have not undergone peer review. Recent studies have shown that the content of preprints generally matches that of the papers once they are submitted to journals, peer-reviewed, and published [6, 7, 8, 9]. Additionally, many preprints are published in high-impact journals included in Clarivate Analytics' Web of Science Core Collection, indicating that preprints are not uniformly of low quality [10]. Studies on the peer review period have shown that research papers initially distributed as preprints tend to undergo a shorter peer review process when submitted to journals, leading to faster publication [11, 12]. Preprints indicate that papers related to COVID-19 are published to journals at a higher frequency than those unrelated to COVID-19 [13, 14] and undergo peer review more expeditiously [15, 16]. Studies on dissemination have shown that preprints are extensively covered by the media [17, 18] and are widely disseminated on social media [19]. The wide dissemination of preprints may be a factor in the subsequent citation of the papers after peer review. Furthermore, in the medical field, authors tend to submit COVID-19-related papers, initially distributed as preprints, to journals with high impact factors [20], and these papers receive numerous citations. For example, as of June 2024, in citations of articles indexed in PubMed and assigned PubMed IDs, the paper by Cummings et al. first opened on medRxiv on April 20, 2020, and later published in The Lancet on May 19, 2020, received 1,165 citations [21,22]. The paper by Guan et al. first opened on medRxiv on February 9, 2020, and published in the New England Journal of Medicine on April 30, 2020, received 14,683 citations [23,24]. The manuscript by Wang et al., first opened on medRxiv on March 6, 2020, was revised with Pan as the first author and a new title for journal submission, and published in JAMA on April 10, 2020, received 853 citations [25,26]. Whether the high number of citations is due to individual preprints or the characteristics of preprints themselves remains unclear. The purpose of this study is to compare the number of citations received by drafts initially distributed as preprints that were later published in journals with those received by papers directly submitted to journals, in order to elucidate the impact of preprints on the citation of papers.

2 Data and Method

2.1 Preprints

A preprint is a draft that is made publicly available in an archive before being submitted to a journal. Preprints have not yet undergone peer review. In this study, we will use preprints from two archives, medRxiv and bioRxiv, which focus on medical and biological topics, to investigate papers related to COVID-19. On January 30, 2020, the World Health Organization (WHO) declared the novel coronavirus infection a "Public Health Emergency of International Concern (PHEIC)." Subsequently, considering the global spread and severity of the infection, the WHO declared it a pandemic on March 11, 2020.

2.2 Journal Articles Related to COVID-19

PubMed is a free resource supporting the search and retrieval of biomedical and life sciences literature with the aim of improving global and individual health. The PubMed contains over 37 million citations and abstracts of biomedical literature. Available to the public online since 1996, PubMed was developed and maintained by the National Center for Biotechnology Information (NCBI) at the U.S. National Library of Medicine (NLM), which is part of the National Institutes of Health (NIH) [27]. In this study, to comprehensively collect papers related to COVID-19, we gathered papers published on PubMed. For collecting the papers, we used the Entrez Programming Utilities (E-utilities), a publicly accessible Application Programming Interface (API) for the NCBI Entrez system, and the program was written in Python. The determination of the subjects of the papers was made using the Medical Subject Headings (MeSH) thesaurus created by the National Library of Medicine. Papers recorded with "COVID-19" in MeSH were defined as COVID-19 related papers.

2.3 Collection Period and Types of Documents

Since papers on topics related to COVID-19 had already been published in journals as early as 2019, the paper collection period was set from 2019 to 2023. Journals publish articles of various types. Here, we defined the document types as journal articles and reviews to measure impact using citation counts.

2.4 Citation Counts

In citation analysis, the citation counts often used are the "Times Cited" produced by Clarivate and provided by WoS. Measuring impact requires investigating a large number of papers, but Clarivate imposes restrictions on downloading papers programmatically. Therefore, in this study, we utilized PubMed, which allows citation downloads via API. When a paper published on PubMed is cited by another paper on PubMed, the citing paper is displayed in the "Cited by" section. The citation count of a paper is determined by counting the number of citing papers.

2.5 Analysis and Examination of Citation Counts

There are several methods for assigning citation counts to a collection of papers gathered by media such as journals or by subject. The total citation count of the papers that constitute the elements of the collection, the average obtained by dividing the total citation count by the number of elements, and the median citation count of the papers are potential candidates for representing the citation count of the collection. The total citation count is not adopted because the number of elements in the collections varies significantly. The distribution of citation counts in the paper collections is not normal; there are many papers with low citation counts and few papers with high citation counts, indicating a non-normal distribution. The average citation count is strongly influenced by a few highly-cited papers. On the other hand, the median is not affected by these

outliers and represents the relative position of the entire collection. Therefore, the median is adopted. For testing the difference in citation counts, we employ the Mann-Whitney U test, a non-parametric test used for collections with significantly different sizes and non-normal distributions. A p-value of less than 0.05 is considered statistically significant.

2.6 Journal Selection

First, preprints that had been submitted to journals, accepted, and assigned a publisher's DOI were identified. Next, among the journals that published papers initially disseminated as preprints, those indexed in PubMed were identified. Finally, from these, the journals with a high number of papers initially disseminated as preprints were selected for the investigation.

3 Result

3.1 Status of the Paper

As for the journals selected for investigation, 184 journals that published at least 10 papers initially disseminated as preprints were identified from PubMed. Metadata for 952,973 papers published by these journals between 2019 and 2023 were collected from PubMed. The total number of citations these papers received from other papers indexed in PubMed was 9,393,014

3.2 COVID-19 Papers and Non-COVID-19 Papers

The median number of citations for COVID-19 and non-COVID-19 papers published in 184 journals was tested using the Mann-Whitney U test. Among the COVID-19 papers, 155 journals showed a significant increase in citations with a p-value of less than 0.05. Conversely, for non-COVID-19 papers, only 2 journals showed a significant increase in citations. This means that 84.24% of the journals that published COVID-19 papers experienced an increase in citations. This situation is shown in Table 1.

Table 1. COVID-19 Papers and Non-COVID-19 Journals.

Difference	Significant	(%)	Non-Significant	(%)	Journals
Positive	155	84.24	8	4.35	163
Zero	8	4.35	6	3.26	14
Negative	2	1.09	5	2.72	7
	165		19		184

Of the 952,973 papers, 74,037 were assigned the MeSH term COVID-19 (hereafter referred to as COVID-19 papers), with a total citation count of 2,015,997. On the other hand, 878,936 papers were not assigned this term (hereafter referred to as non-COVID-19 papers), with a total citation count of 7,377,017. Thus, 7.77% of the COVID-19 papers accounted for 21.46% of the citations. This situation is shown in Table 2.

Table 2. COVID-19 Papers and Non-COVID-19 Papers.

	Papers	(%)	Citations	(%)
COVID-19	74,037	7.77	2,015,997	21.46
Non-COVID-19	878,936	92.23	7,377,017	78.54
	952,973		9,393,014	

3.3 Papers initially disseminated as preprints and direct journal submission papers

The difference in the median number of citations between preprint-disseminated papers and direct-submitted papers published in 184 journals was tested using the Mann-Whitney U test. For preprint-disseminated papers, the p-value was less than 0.05, indicating that there were 161 journals with significantly higher citation counts. In contrast, there were only 3 journals with significantly higher citations for direct-submitted papers. This means that 87.50% of the journals that published preprint-disseminated papers experienced an increase in citations. This situation is shown in Table 3.

Table 3. Preprint-disseminated and direct-submitted journals.

Difference	Significant	(%)	Non-Significant	(%)	Journals
Positive	161	87.50	12	6.52	173
Zero	2	1.09	4	2.17	6
Negative	3	1.63	2	1.09	5
	166		18		184

Out of 952,973 papers, the number of papers initially disseminated as preprints（hereafter referred to as preprint-disseminated papers） was 54,909, with a total of 1,293,049 citations. On the other hand, the number of papers submitted directly to journals (hereafter referred to as direct-submitted papers) was 898,064, with a total of 8,099,965 citations. The 5.76% of preprint-disseminated papers garnered 13.77% of the citations. This situation is shown in Table 4.

Table 4. Preprint-disseminated and direct-submitted papers.

	Papers	(%)	Citations	(%)
Preprint-disseminated	54,909	5.76	1,293,049	13.77
direct-submitted	898,064	94.24	8,099,965	86.23
	952,973		9,393,014	

3.4 Papers initially disseminated as preprints and direct journal submission papers related to COVID-19

The difference in the median number of citations between COVID-19 preprint distribution papers and COVID-19 directly submitted papers published in 184 journals was tested using the Mann-Whitney U test. For COVID-19 preprint distribution papers, the p-value was less than 0.05, indicating that there were 129 journals with significantly higher citation counts. In contrast, there were no COVID-19 directly submitted papers

with significantly higher citations. This means that 70.11% of the journals that published preprint distribution papers experienced an increase in citations. This situation is shown in Table 5.

Table 5. COVID-19 Preprint-disseminated and direct-submitted papers.

Difference	Significant	(%)	Non-Significant	(%)	Journals
Positive	129	70.11	44	23.91	173
Zero	0	0	3	1.63	3
Negative	0	0	8	4.35	8
	129		55		184

Out of 74,037 COVID-19 papers, the number of preprint distribution papers (referred to as COVID-19 preprint distribution papers) was 9,028, with a total of 542,715 citations. On the other hand, the number of papers directly submitted to journals (referred to as COVID-19 directly submitted papers) was 65,009, with a total of 1,473,282 citations. Thus, 12.19% of the preprint distribution papers accounted for 26.92% of the citations. This situation is shown in Table 6.

Table 6. COVID-19 Preprint-disseminated and direct-submitted papers.

	Papers	(%)	Citations	(%)
Preprint-disseminated	9,028	12.19	542,715	26.92
direct-submitted	65,009	87.81	1,473,282	73.08
	74,037		2,015,997	

Table 7 shows the number of COVID-19 preprint-disseminated papers and direct-submitted papers in journals, the total number of citations, the median number of citations, the p-values, the significance, and the differences.

Table 7. COVID-19 preprint-disseminated and direct-submitted papers.

	COVID-19 preprint-disseminated			COVID-19 direct-submitted					
Journal	Papers	Citations	Median	Papers	Citations	Median	P-value	Sig	Diff
PLoS One	1,094	24,519	9	4,885	44,273	3	0.0000	**	6
Sci Rep	507	10,223	10	2,925	25,686	3	0.0000	**	7
Nat Commun	437	27,514	35	597	21,340	12	0.0000	**	23
BMJ Open	325	6,153	8	1,798	11,673	2	0.0000	**	6
Clin Infect Dis	211	11,606	20	865	38,559	13	0.0000	**	7
Front Immunol	190	5,340	11	2,065	27,884	5	0.0000	**	6
Elife	189	5,801	13	67	2,708	17	0.4921		-4
Viruses	172	3,483	9.5	1,651	15,407	3	0.0000	**	6.5
Vaccine	135	3,311	8	1,034	10,225	3	0.0000	**	5
Front Public Health	134	2,621	5	3,677	20,832	2	0.0000	**	3
Int J Infect Dis	129	4,713	14	1,020	23,952	9	0.0010	**	5
Proc Natl Acad Sci U S A	128	9,158	39.5	430	18,381	12	0.0000	**	27.5
J Med Virol	126	3,582	12	1,499	42,951	7	0.0000	**	5
J Infect Dis	121	4,610	19	423	8,533	6	0.0000	**	13
BMC Public Health	116	1,604	8	1,382	11,561	3	0.0000	**	5
PLoS Pathog	113	3,623	16	142	3,407	8	0.0005	**	8
PLoS Comput Biol	112	1,983	8	90	571	3	0.0000	**	5
Int J Environ Res Public Health	111	2,141	9	7,125	68,095	3	0.0000	**	6
Microbiol Spectr	103	1,005	6	324	1,804	3	0.0000	**	3
JAMA Netw Open	103	5,025	22	1,123	28,960	7	0.0000	**	15
Science	102	37,396	201.5	199	26,314	39	0.0000	**	162.5
BMC Infect Dis	101	1,573	9	791	7,138	4	0.0000	**	5
Nature	96	57,982	373.5	345	40,857	19	0.0000	**	354.5
Sci Total Environ	94	4,657	17	951	27,664	8	0.0000	**	9
Cell Rep	91	4,660	26	135	3,773	13	0.0000	**	13
BMC Med	88	3,347	16.5	131	2,413	9	0.0003	**	7.5
Emerg Infect Dis	87	4,898	20	550	14,133	7	0.0000	**	13
Euro Surveill	86	5,983	25.5	336	11,535	10	0.0000	**	15.5
EBioMedicine	84	4,086	25	250	4,639	8	0.0000	**	17
Epidemics	79	693	5	78	321	2	0.0000	**	3
Epidemiol Infect	77	1,067	6	324	2,938	4	0.0123	*	2
J Clin Microbiol	76	4,727	39	129	4,107	14	0.0000	**	25
Cell	76	35,659	168.5	147	26,695	70	0.0000	**	98.5
J Med Internet Res	72	2,185	15	883	16,379	7	0.0001	**	8
Nat Med	71	24,065	189	207	26,472	41	0.0000	**	148
mBio	67	2,318	19	155	2,546	8	0.0000	**	11
J Clin Virol	64	2,483	19.5	284	8,283	9.5	0.0007	**	10
Cell Rep Med	59	2,454	20	95	1,622	10	0.0002	**	10
J Virol	58	2,339	19	117	3,863	6	0.0000	**	13
Clin Microbiol Infect	57	2,411	21	331	11,858	11	0.0002	**	10
BMJ	56	12,876	97	700	18,444	2	0.0000	**	95
JCI Insight	55	2,781	21	103	2,723	10	0.0013	**	11
BMJ Glob Health	54	2,520	20.5	452	5,460	5	0.0000	**	15.5
Emerg Microbes Infect	51	2,311	27	257	11,610	10	0.0004	**	17
Sci Transl Med	50	4,199	52.5	78	3,708	26	0.0003	**	26.5
PLoS Med	49	2,538	28	93	2,073	8	0.0000	**	20
Sci Adv	49	4,292	32	99	2,677	10	0.0001	**	22
J Clin Invest	48	5,956	48	133	5,969	16	0.0001	**	32
JMIR Public Health Surveill	48	939	8	386	6,114	5.5	0.0452	*	2.5
Lancet Infect Dis	46	8,386	93	392	28,183	14	0.0000	**	79
N Engl J Med	39	37,780	459	344	103,665	40.5	0.0000	**	418.5
J R Soc Interface	39	546	10	44	282	4	0.0116	*	6
Commun Biol	37	1,011	12	100	1,162	4	0.0004	**	8
Public Health	34	982	11.5	377	3,107	3	0.0000	**	8.5
J Virol Methods	33	469	7	170	1,268	4	0.0002	**	3
J Infect	32	1,023	21	158	14,585	24	0.3847		-3
Cell Host Microbe	31	3,893	73	85	8,776	37	0.0068	**	36
Int J Mol Sci	31	606	11	1,150	11,643	4	0.0001	**	7
mSphere	31	622	12	66	1,181	7	0.0487	*	5
Infect Genet Evol	30	729	10.5	120	2,228	6	0.0506		4.5
Nat Microbiol	30	10,260	70	43	2,295	20	0.0002	**	50
Sci Immunol	30	4,325	109	68	5,225	41.5	0.0012	**	67.5
J Theor Biol	29	434	6	51	224	3	0.0242	*	3
Front Cell Infect Microbiol	28	439	6.5	392	4,333	4	0.0153	*	2.5
Influenza Other Respir Viruses	28	233	5	168	1,321	3	0.0348	*	2
Int J Epidemiol	28	1,100	14.5	71	1,565	9	0.1125		5.5
Med	27	1,207	27	57	1,109	9	0.0001	**	18
Am J Epidemiol	26	487	10	84	1,006	3	0.0023	**	7
Ann Intern Med	25	4,204	34	281	10,133	5	0.0000	**	29

J Biol Chem	25	806	22	52	920	7	0.0066	**	15
J Chem Inf Model	25	311	8	93	1,361	4	0.2597		4
J Exp Med	25	3,243	46	47	2,403	12	0.0003	**	34
PLoS Biol	25	1,421	40	51	1,058	8	0.0000	**	32
Am J Trop Med Hyg	24	448	9.5	352	4,835	7	0.2061		2.5
Bioinformatics	24	420	4	54	505	3	0.2197		1
Eur J Epidemiol	24	856	11.5	62	1,624	7.5	0.3327		4
Nucleic Acids Res	24	562	15.5	79	11,927	23	0.2090		-7.5
Water Res	24	1,398	25	72	1,539	3	0.0002	**	22
Brief Bioinform	23	289	9	147	1,836	7	0.9964		2
Hum Vaccin Immunother	23	661	9	751	7,336	4	0.0120	*	5
Virus Res	23	465	13	108	2,593	6	0.4270		7
Environ Res	22	782	22.5	384	6,787	8	0.0066	**	14.5
Lancet Public Health	22	4,289	69.5	86	4,486	19	0.0001	**	50.5
Nat Immunol	22	3,642	90	92	3,716	7	0.0000	**	83
Diagn Microbiol Infect Dis	21	150	3	152	1,144	4	0.4829		-1
Genome Med	21	731	18	23	303	7	0.0028	**	11
J Immunol	21	511	11	72	1,246	7.5	0.0797		3.5
J Epidemiol Community Health	20	758	13.5	60	1,683	7	0.0631		6.5
BMC Health Serv Res	20	577	7	545	2,307	2	0.0001	**	5
J Public Health (Oxf)	20	408	7	339	2,770	2	0.0007	**	5
Comput Biol Med	20	325	6.5	362	5,061	6	0.6515		0.5
Eur Respir J	19	3,184	78	73	4,360	35	0.1051		43
J Hosp Infect	19	728	12	242	4,749	4.5	0.0003	**	7.5
Lancet Microbe	19	1,740	28	101	3,975	11	0.0041	**	17
Anal Chem	19	280	11	215	2,300	3	0.0075	**	8
Immunity	19	1,936	114	93	9,506	52	0.1267		62
Lancet Digit Health	19	1,090	25	80	2,336	13	0.0160	*	12
Nutrients	19	1,551	14	713	13,342	5	0.0005	**	9
PLoS Negl Trop Dis	19	388	15	92	1,412	7	0.0543		8
J Infect Chemother	18	136	5.5	230	1,067	2	0.0190	*	3.5
Age Ageing	18	261	11.5	91	1,784	6	0.6619		5.5
Life Sci Alliance	18	239	11	28	163	3	0.0018	**	8
Medicine (Baltimore)	18	243	9	869	3,602	1	0.0002	**	8
Signal Transduct Target Ther	17	1,893	52	211	7,316	14	0.0044	**	38
J Proteome Res	17	670	11	81	1,316	10	0.1353		1
J Travel Med	17	452	20	218	6,329	6	0.0018	**	14
J Infect Public Health	17	186	6	442	6,089	4	0.1466		2
Mol Biol Evol	17	375	16	13	344	14	0.8669		2
Nat Genet	17	978	50	8	224	10.5	0.1225		39.5
Philos Trans R Soc Lond B Biol Sci	17	743	18	10	81	5	0.0023	**	13
Thorax	17	912	33	70	2,514	14	0.0469	*	19
Transfusion	17	293	15	143	1,118	4	0.0003	**	11
BMC Med Res Methodol	16	173	7.5	86	482	2	0.0005	**	5.5
ACS Infect Dis	16	496	21	66	1,166	6	0.0254	*	15
Antiviral Res	16	570	14.5	92	3,584	3	0.0355	*	11.5
Biosens Bioelectron	16	734	31	246	5,441	9	0.0014	**	22
J Affect Disord	16	987	11	466	12,493	7	0.0301	*	4
J Occup Environ Med	16	102	5	234	1,285	1	0.0071	**	4
Math Biosci Eng	16	184	4	210	852	1	0.0027	**	3
Brain Behav Immun	15	1,223	15	103	9,919	17	0.5879		-2
Infect Dis (Lond)	15	241	11	136	1,400	3	0.0113	*	8
Eur J Clin Microbiol Infect Dis	15	287	17	98	2,898	8	0.1674		9
BMC Bioinformatics	15	116	5	45	171	2	0.1291		3
Biochem Biophys Res Commun	15	969	21	90	1,927	6	0.0117	*	15
J Mol Diagn	15	138	7	28	283	6	0.6731		1
Nat Hum Behav	15	3,248	137	96	6,043	11.5	0.0000	**	125.5
Lancet Respir Med	14	2,270	112	340	20,884	11	0.0000	**	101
Disaster Med Public Health Prep	14	109	3	520	2,384	1	0.0809		2
Infect Control Hosp Epidemiol	14	258	15.5	551	4,297	3	0.0001	**	12.5
Ann Epidemiol	14	861	6.5	84	1,028	3.5	0.0685		3
Cells	14	194	10	288	5,473	9	0.8642		1
Eur J Public Health	14	221	7.5	125	1,000	3	0.0366	*	4.5
Infection	14	414	6	238	4,004	7	0.1235		-1
Int J Public Health	14	296	6	234	923	1	0.0052	**	5
J Med Microbiol	14	286	13	46	235	3	0.0018	**	10
J Transl Med	14	766	15	110	3,589	11	0.6213		4
Transbound Emerg Dis	14	289	11	114	2,174	9	0.5111		2
Ann Rheum Dis	13	583	31	90	4,314	17.5	0.1152		13.5
Clin Chem	13	578	36	45	1,424	8	0.0038	**	28
JAMA	13	2,323	134	766	97,429	17	0.0006	**	117

Am J Infect Control	13	202	3	367	4,719	3	0.2247		0
Antimicrob Agents Chemother	13	896	14	35	940	12	0.0774		2
BMJ Open Respir Res	13	344	23	94	901	4.5	0.0008	**	18.5
Biophys J	13	266	11	22	123	4	0.1405		7
CMAJ	13	639	25	228	3,659	5	0.0001	**	20
CMAJ Open	13	130	7	92	637	3	0.0102	*	4
Epidemiology	13	262	9	41	169	2	0.0001	**	7
Eur J Immunol	13	374	20	60	745	7	0.0054	**	13
Genome Biol	13	159	10	3	15	6	0.2511		4
J Am Med Inform Assoc	13	174	10	150	3,839	6	0.1720		4
J Gen Virol	13	283	13	25	495	4	0.0864		9
PeerJ	13	40	3	157	273	1	0.0025	**	2
Proc Biol Sci	13	153	5	9	98	6	1.0000		-1
Swiss Med Wkly	13	302	11	146	1,525	4	0.0108	*	7
Virology	13	262	19	59	666	6	0.0221	*	13
Lancet	12	4,907	368.5	613	124,046	14	0.0000	**	354.5
Psychiatry Res	12	2,628	26	351	16,022	8	0.0564		18
J Biomol Struct Dyn	12	91	5	442	7,462	4	0.9687		1
Emerg Med J	12	121	7.5	83	558	3	0.1362		4.5
Sci Data	12	313	8.5	90	1,288	3	0.0285	*	5.5
Math Biosci	12	368	9.5	41	452	3	0.0857		6.5
J Gen Intern Med	11	318	26	261	3,069	3	0.0061	**	23
Biomed Res Int	11	189	11	179	1,370	3	0.0042	**	8
Cell Rep Methods	11	61	4	16	232	4	1.0000		0
Chem Senses	11	555	16	8	60	4	0.0892		12
Clin Chem Lab Med	11	367	31	120	3,436	5	0.0005	**	26
EMBO Mol Med	11	432	22	42	1,181	11.5	0.3452		10.5
J Biomed Inform	11	70	4	71	442	4	0.9563		0
J Mol Biol	11	435	15	20	664	3	0.2291		12
J Occup Health	11	43	3	36	397	4	0.6491		-1
Mol Cell	11	816	74	13	1,639	42	0.3241		32
Occup Environ Med	11	219	5	30	328	2.5	0.2611		2.5
Lancet Glob Health	10	2,344	117.5	164	5,989	11	0.0001	**	106.5
Pediatr Infect Dis J	10	153	11	307	3,520	3	0.0348	*	8
Nat Biotechnol	10	1,203	59	33	3,243	29	0.1765		30
Am J Pathol	10	529	24.5	21	931	17	0.3002		7.5
BMJ Paediatr Open	10	68	4	72	280	2	0.0209	*	2
Br J Gen Pract	10	197	7	99	879	1	0.0054	**	6
Environ Sci Technol	10	312	17.5	85	983	2	0.0024	**	15.5
Int J Med Inform	10	92	5.5	89	626	3	0.0276	*	2.5
J Glob Health	10	240	28	440	3,261	3	0.0004	**	25
Lancet Healthy Longev	10	534	48	58	619	3.5	0.0001	**	44.5
Microb Genom	10	102	8	22	113	1	0.0013	**	7
Trans R Soc Trop Med Hyg	10	129	14	80	498	3	0.0013	**	11

[**: p-value < 0.01, *: p-value < 0.05]

4 Analysis and Discussion

4.1 Distribution of the median number of citations by journal

Figure 1 shows the distribution of the median number of citations for COVID-19 papers and non-COVID-19 papers, preprint-disseminated and direct-submitted papers, and COVID-19 preprint-disseminated and direct-submitted papers. Figure 1 shows that preprint-distributed papers have higher citation counts than COVID-19 papers, and that COVID-19 preprint-distributed papers tend to have even higher citation counts.

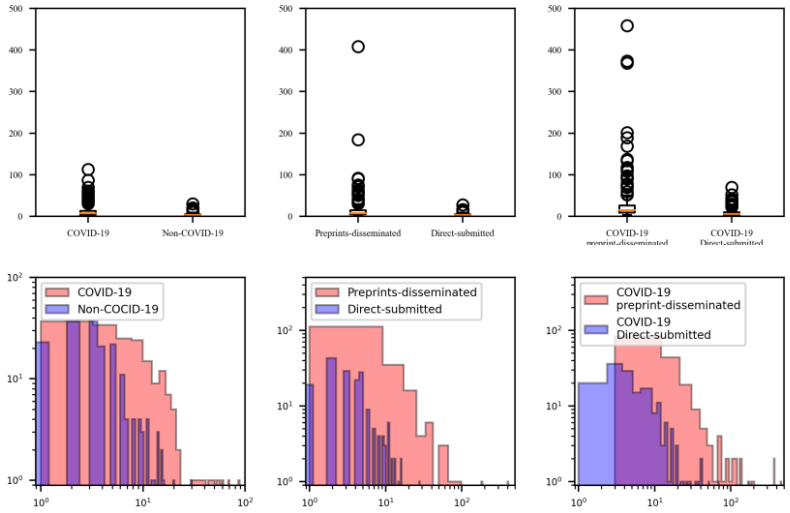

Fig. 1. Distribution of the median number of citations.

4.2 Relative position of journals based on the distribution of median citation counts

In journals with a higher median citation count for preprint distribution papers compared to papers directly submitted to journals—i.e., journals with a positive difference—clustering with a citation threshold of 20 resulted in seven clusters. The first cluster consisted of a single journal, *The New England Journal of Medicine*. The second cluster included two journals, *The Lancet* and *Nature*. The third cluster comprised one journal, *Cell*. The fourth cluster included two journals, *Science* and *Nature Medicine*. The fifth cluster consisted of seven journals: *Nature Immunology*, *The Lancet Infectious Diseases*, *BMJ*, *The Lancet Global Health*, *The Lancet Respiratory Medicine*, *JAMA*, and *Nature Human Behaviour*. The seventh cluster included two journals, *Immunity* and *Science Immunology*. Finally, the sixth cluster included the remaining 158 journals. The 13 journals classified into clusters 1 through 5 were found to have garnered an especially large number of citations. This cluster is illustrated in the dendrogram shown in Figure 2.

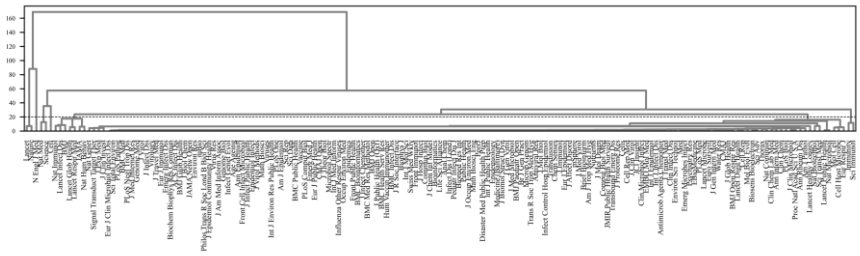

Fig. 2. Positive median difference in citation counts of journal articles on COVID-19.

Figure 3 shows the diagram obtained by adding two clusters with zero and negative differences to the seven clusters with positive differences shown in Figure 2. Clearly, among the seven positive clusters, clusters 1 through 5 are located in the upper-left part of the figure 3, visually confirming that distributing the draft as a preprint significantly increased the number of citations.

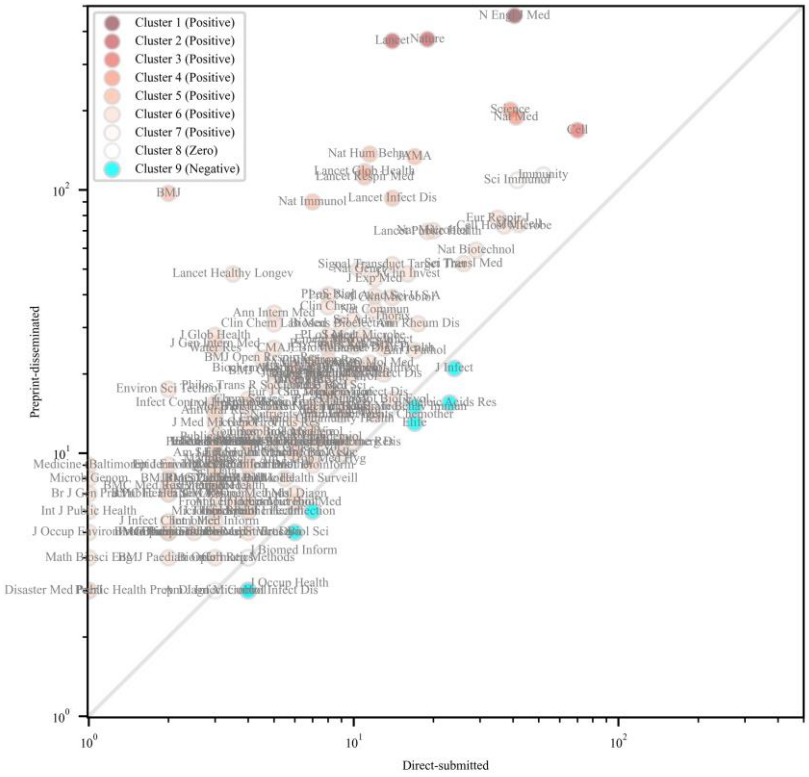

Fig. 3. Median citation count of journal articles on COVID-19.

4.3 Impact of Preprints on the Citations of Journal Articles Related to COVID-19

In most journals, COVID-19 papers received more citations than non-COVID-19 papers. Furthermore, papers first distributed as preprints received more citations than those directly submitted to journals by the authors. It can be inferred that the prior dissemination of preprints allowed the research to reach a wider audience, thereby gaining more readers and citations.

5 Conclusion

In this study, we analyzed the number of times PubMed articles were cited by other PubMed articles to elucidate the impact of preprints on COVID-19 research. The results revealed that, among papers on COVID-19, preprint-distributed papers received significantly more citations than direct journal submission papers, as demonstrated by the analysis of the median difference in journal citation counts. This indicates that distributing preprints exerts a strong influence on the subsequent citation counts of papers published in journals. For novel topics like COVID-19, it is crucial for researchers to rapidly disseminate their findings. Therefore, researchers are encouraged to utilize preprints to share their research outcomes promptly.

Acknowledgments. This work was supported by JSPS KAKENHI Grant Numbers JP19K12707, JP20K12569, JP22K12737, and ROIS NII Open Collaborative Research 2023(23FS01)

Disclosure of Interests. The authors have no conflicts of interest directly relevant to the content of this article.

References

1. Gianola S, Jesus TS, Bargeri S, Castellini G. Characteristics of academic publications, preprints, and registered clinical trials on the COVID-19 pandemic. PLOS ONE **15**(10), e0240123 (2020)
2. Blatch-Jones A.J., Recio Saucedo A., Giddins B.: The use and acceptability of preprints in health and social care settings: A scoping review. PLOS ONE **18**(9), e0291627 (2023)
3. Zeng L.: Changes in health communication in the age of COVID-19: A study on the dissemination of preprints to the public. Frontiers in Public Health **11**, 1078115 (2023)
4. Wang P, Tian D.: Bibliometric analysis of global scientific research on COVID-19. Journal of Biosafety and Biosecurity **3**(1), 4–9 (2021)
5. Fraser, N., Mayr, P., Peters, I.: Motivations, concerns and selection biases when posting preprints: A survey of bioRxiv authors. PLOS ONE **17**(11), e0274441 (2022)
6. Brierley, L., Nanni, F., Polka, J.K., Dey, G., Pálfy, M., Fraser, N., et al.: Tracking changes between preprint posting and journal publication during a pandemic. PLOS Biology **20**(2), e3001285 (2022)
7. Janda, G., Khetpal, V., Shi, X., Ross, J.S., Wallach, J.D.: Comparison of clinical study results reported in medRxiv preprints vs peer-reviewed journal articles. JAMA Network Open **5**(12), e2245847 (2022)
8. Nelson, L., Ye, H., Schwenn, A., Lee, S., Arabi, S., Hutchins, B.I.: Robustness of evidence reported in preprints during peer review. Lancet Global Health **10**(11), e1684–e1687 (2022)
9. Davidson, M., Evrenoglou, T., Graña, C., Chaimani, A., Boutron, I.: Comparison of effect estimates between preprints and peer-reviewed journal articles of COVID-19 trials. BMC Medical Research Methodology **24**(1), 9 (2024)
10. Tsunoda, H., Sun, Y., Nishizawa, M., Liu, X., Amano, K.: An analysis of published journals for papers posted on bioRXiv. Proceedings of the Association for Information Science and Technology **56**, 783-784 (2019)

11. Tsunoda, H., Sun, Y., Nishizawa, M., Liu, X., Amano K.: The influence of bioRχiv on PLOS ONE's peer-review and acceptance time. Proceedings of the Association for Information Science and Technology **57**, e398 (2020)
12. Tsunoda, H., Sun, Y., Nishizawa, M., Liu, X., Amano, K.: How preprint affects the publishing process: Duration of the peer review process between bioRxiv and journal papers. Proceedings of the Association for Information Science and Technology **59**, 505–509 (2022)
13. Otridge, J., Ogden, C.L., Bernstein, K.T., Knuth, M., Fishman, J., Brooks J.T.: Publication and impact of preprints included in the first 100 editions of the CDC COVID-19 Science Update: Content analysis. JMIR Public Health and Surveillance **8**(7), e35276 (2022)
14. Tsunoda, H., Sun, Y., Nishizawa, M., Liu, X., Amano, K., Kominami, R.: Role of preprints in scholarly communication on COVID-19 - A quantitative survey on medRxiv and bioRxiv. Proceedings of ISSI 2023 **3**, 89–90 (2023)
15. Fraser, N., Brierley, L., Dey, G., Polka, J.K., Pálfy, M., Nanni, F., Coates, J.A.: The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape. PLOS Biology **19**(4), e3000959 (2021)
16. Tsunoda, H., Sun, Y., Nishizawa, M., Liu, X., Amano, K., Kominami, R.: The Pivotal Role of Preprint Platforms in Disseminating COVID-19 Research: A Global Investigation of Country-Level Activities. Lecture Notes in Computer Science 14458 (2023)
17. 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 Communication **37**(6), 726–738 (2022)
18. van Schalkwyk, F., Dudek, J.: Reporting preprints in the media during the COVID-19 pandemic. Public Understanding of Science **31**(5), 608–616 (2022)
19. Sarkis-Onofre, R., Girotto, C., Agostini, B.A.: Exploring the use of preprints in dentistry. Journal of Dentistry **136**, 104634 (2023)
20. Tsunoda, H., Sun, Y., Nishizawa, M., Liu, X., Amano, K., Kominami, R.: The impact of preprints on COVID-19 research dissemination: A quantitative analysis of journal publications. Proceedings of the Association for Information Science and Technology (in press)
21. Cummings, M.J., Baldwin, M.R., Abrams, D., Jacobson, S.D., Meyer, B.J., Balough, E.M., Aaron, J.G., Claassen, J., Rabbani, L.E., Hastie, J., Hochman, B.R., Salazar-Schicchi, J., Yip, N.H., Brodie, D., O'Donnell, M.R.: Epidemiology, clinical course, and outcomes of critically ill adults with COVID-19 in New York City: a prospective cohort study. medRxiv (2020)
22. Cummings, M.J., Baldwin, M.R., Abrams, D., Jacobson, S.D., Meyer, B.J., Balough, E.M., Aaron, J.G., Claassen, J., Rabbani, L.E., Hastie, J., Hochman, B.R., Salazar-Schicchi, J., Yip, N.H., Brodie, D., O'Donnell, M.R.: Epidemiology, clinical course, and outcomes of critically ill adults with COVID-19 in New York City: a prospective cohort study. Lancet **395**(10239):1763–1770 (2020)
23. Guan, W.J., Ni, Z.Y., Hu, Y., Liang, W.H., Ou, C.Q., He, J.X., Liu, L., Shan, H., Lei, C.L., Hui, D.S.C., Du, B., Li, L.J., Zeng, G., Yuen, K.Y., Chen, R.C., Tang, C.L., Wang, T., Chen, P.Y., Xiang, J., Li, S.Y., Wang, J.L., Liang, Z.J., Peng, Y.X., Wei, L., Liu, Y., Hu, Y.H., Peng, P., Wang, J.M., Liu, J.Y., Chen, Z., Li, G., Zheng, Z.J., Qiu, S.Q., Luo, J., Ye, C.J., Zhu, S.Y., Zhong, N.S.: Clinical characteristics of 2019 novel coronavirus infection in China: medRxiv (2020)
24. Guan, W.J., Ni, Z.Y., Hu, Y., Liang, W.H., Ou, C.Q., He, J.X., Liu, L., Shan, H., Lei, C.L., Hui, D.S.C., Du, B., Li, L.J., Zeng, G., Yuen, K.Y., Chen, R.C., Tang, C.L., Wang, T., Chen, P.Y., Xiang, J., Li, S.Y., Wang, J.L., Liang, Z.J., Peng, Y.X., Wei, L., Liu, Y., Hu, Y.H., Peng, P., Wang, J.M., Liu, J.Y., Chen, Z., Li, G., Zheng, Z.J., Qiu, S.Q., Luo, J., Ye, C.J.,

Zhu, S.Y., Zhong, N.S.: Clinical characteristics of coronavirus disease 2019 in China. New England Journal of Medicine **382**(18), 1708–1720 (2020)

25. Wang, C., Liu, L., Hao, X., Guo, H., Wang, Q., Huang, J., He, N., Yu, H., Lin, X., Pan, A., Wei, S., Wu, T.: Evolving Epidemiology and Impact of Non-pharmaceutical Interventions on the Outbreak of Coronavirus Disease 2019 in Wuhan, China. medRxiv (2020)
26. Pan, A., Liu, L., Wang, C., Guo, H., Hao, X., Wang, Q., Huang, J., He, N., Yu, H., Lin, X., Wei, S., Wu, T.: Association of public health interventions with the epidemiology of the COVID-19 outbreak in Wuhan, China. JAMA **323**(19), 1915–1923 (2020)
27. PubMed Overview, https://pubmed.ncbi.nlm.nih.gov/about/, last accessed 2024/7/17