PubMed Preprints Rose from 1,482 in 2020 to 20,965 in 2025, but Crossed with the COVID-19 MeSH Term They Were 0 in Every Year — Preprints Are Not Indexed for MEDLINE and Carry No Index Terms, So Counting COVID-19 as a Title/Abstract Word Finds 1,151 in 2020 Alone (0.777 of That Year's Preprints) [M119]

Records found by index terms = only records that were indexed —— What separates them is whether you count by index terms or by title/abstract words

Yuuki YAMAGISHI (山岸 勇輝) ／ Independent Researcher

ORCID 0009-0006-8587-3450

September 2026 ／ CC BY 4.0

Abstract

Using counts from PubMed (U.S. National Library of Medicine) retrieved through E-utilities (27 September 2026), this paper counted records with the publication type Preprint (PubMed preprints) by publication year and month, and compared counts of COVID-19 by index term (MeSH) and by title/abstract word. Whether preprints are good or bad, and research quality, are not discussed. No new mathematical theorem and no new law are claimed.

First, **PubMed preprints were 0 for publication years through 2019, then 1,482 in 2020, 822 in 2022, 13,200 in 2023, and 20,965 in 2025, and 13,935 more in 2026 up to the retrieval date. The count without years, 67,864, equaled exactly the sum of the yearly counts for 2020–2026** (Section 4).

Second, and this is the core of this paper, **crossed with the COVID-19 MeSH term, PubMed preprints were 0 in every year from 2016 to 2025. But counted by the title/abstract word (covid-19), they were 1,151 in 2020 (0.777 of that year's preprints) and 1,012 in 2021 (0.722), and 3,732 in total over 2020–2025** (Section 5).

Third, **the reason is written in the PubMed User Guide. Preprints are records not indexed for MEDLINE, and MeSH terms are assigned during MEDLINE indexing. Indeed, only 1 preprint was in MEDLINE over 2016–2025** (Section 5).

Fourth, **PubMed preprints are not all preprints. According to the User Guide, they are citations to preprints of NIH-funded research included in PMC from June 2020, and 0.9997 of the preprints were in PMC** (Section 4).

Fifth, **even for all records, counting COVID-19 by title/abstract word gave 1.256 times (2020) to 1.450 times (2022) the count by MeSH. Counting by index terms does not find records that were not indexed** (Section 5).

Sixth, what separates them is **whether you count by index terms or by title/abstract words**. When you search PubMed with a disease's MeSH term, records that are not indexed, such as preprints, do not come up. To reduce what you miss, also search by title/abstract words.

Scope note: **This paper claims no new mathematical theorem and no new law** — the counts are public in PubMed, and this paper only counted and compared them. **This paper is a structural interpretation of the shape of numbers in public aggregate data; it is not medical advice, not a diagnosis, and not a treatment recommendation. The author is not a physician. Decisions about individual health should be discussed with a physician or a medical institution. This is a preprint that has not been peer reviewed. Whether preprints are good or bad, and the quality or value of research, are not discussed. Whether preprint results can be trusted is not discussed either. No specific paper, journal, publisher, author, or preprint server is named.** The counts in this paper are "preprints in PubMed," not all preprints. Counts from preprint servers were not used because their terms of use could not be confirmed. Having the word covid-19 in the title or abstract is not the same as the paper being about COVID-19. The counts are as of retrieval and change daily. **About the data**: Courtesy of the U.S. National Library of Medicine. The counts are as of 27 September 2026 and do not reflect the most current or accurate data from NLM. NLM does not endorse this paper. What this paper adds: **showing with counts that PubMed preprints fall entirely outside index-term searches, and showing how many there are when counted by title/abstract words**.

---

1. Notation and question

- **PubMed preprints**: `preprint[pt]` (publication type Preprint). The original text of the PubMed User Guide: "As of June 2020, PubMed Central (PMC) includes preprints that report NIH-funded research results. Citations to these preprints are deposited in PubMed."
- **COVID-19 MeSH**: `"covid-19"[mh]` (index term). **COVID-19 word**: `covid-19[tiab]` (title/abstract word).
- **MEDLINE**: `medline[sb]`. The User Guide gives preprints as an example of records not indexed for MEDLINE ("some citations are not indexed for MEDLINE, such as preprints, …").

Question: How much did PubMed preprints increase? How different is the number of preprints when COVID-19 is counted by MeSH and by title/abstract word?

---

2. Preview and design

Written honestly: at the stock stage, the plan was to write about "the growth of preprints" using counts from preprint servers (such as medRxiv). The servers' API description page and about page were read, but no statement of the conditions for reusing the counts returned by the API could be found. So this paper counted only the PubMed publication type Preprint, whose terms of use have been confirmed.

	Content
Data	PubMed E-utilities counts (97 main queries and 15 additional queries) and the PubMed User Guide
Noticed in the main retrieval	Preprints crossed with the COVID-19 MeSH term were 0 in every year
Added	Queries counting COVID-19 by title/abstract word (for preprints and for all records)
Compared	Preprints by publication year and month, shares in PMC and in MEDLINE, COVID-19 by MeSH and by word

---

3. Data and checks

Calculation 3.1 Five files — the two count JSON files, the two retrieval scripts, and the copy of the User Guide — were fixed by SHA-256, and the following were checked. The checker computes only from the saved JSON files and does not query PubMed.

- The 97 main queries (27 September 2026, 02:46:07–02:47:38) and 15 additional queries (02:48:12–02:48:26) are complete.

- The query translations contain, as intended, the publication type ("preprint"[Publication Type]), MeSH ("covid-19"[MeSH Terms]), title/abstract ("covid-19"[Title/Abstract]), PMC, and MEDLINE.
- The only warnings were "No items found" for queries with a count of 0 (43); warnings indicating a misread query were 0.
- The copy of the User Guide contains the original text stating that preprints are citations to NIH-funded research and are not indexed for MEDLINE.
- All subset relations hold.

---

4. The number of PubMed preprints

Calculation 4.1 PubMed preprints by publication year

Publication year	Records	Preprints	In PMC	In MEDLINE
2019	1,418,997	0	0	0
2020	1,641,213	1,482	1,481	0
2021	1,787,579	1,401	1,401	0
2022	1,783,557	822	822	0
2023	1,697,588	13,200	13,188	1
2024	1,739,848	16,059	16,059	0
2025	1,881,905	20,965	20,964	0

Observation 4.2: **PubMed preprints were 0 for publication years through 2019, fell from 1,482 in 2020 to 822 in 2022, and rose to 13,200 in 2023 and 20,965 in 2025. There were 13,935 more in 2026 up to the retrieval date. The count without years, 67,864, equaled exactly the sum of the yearly counts for 2020–2026, so there was no overlap between years.** The reason for the rise from 2022 to 2023 cannot be told from the counts.

Observation 4.3: **Of preprints, 0.9997 were in PMC over 2016–2025, and 1 was in MEDLINE.** According to the User Guide, PMC began including preprints in June 2020, but counting by publication date there were 5 in January 2020 and 104 by April. The publication date is the preprint's date, not the date it entered PubMed.

---

5. The core — counting by index terms or by title/abstract words

Calculation 5.1 COVID-19 counted by MeSH and by title/abstract word

Publication year	Preprints: MeSH	Preprints: word (share of preprints)	All records: MeSH	All records: word (word ÷ MeSH)
2020	0	1,151 (0.777)	64,256	80,713 (1.256)
2021	0	1,012 (0.722)	90,378	124,435 (1.377)
2022	0	547 (0.665)	80,134	116,167 (1.450)
2023	0	452 (0.034)	55,306	78,854 (1.426)
2024	0	308 (0.019)	36,506	50,286 (1.377)
2025	0	262 (0.012)	25,622	35,567 (1.388)

Observation 5.2: **Crossed with the COVID-19 MeSH term, PubMed preprints were 0 in every year from 2016 to 2025. Counted by title/abstract word, they were 1,151 in 2020 (0.777 of that year's preprints), 1,012 in 2021 (0.722), and 547 in 2022 (0.665), and 3,732 in total over 2020–2025.** Preprints are not indexed for MEDLINE (User Guide), so they carry no MeSH index terms.

Observation 5.3: **Even for all records, counting COVID-19 by title/abstract word gave 1.256 times (2020) to 1.450 times (2022) the count by MeSH.** Counting by index terms does not find records that were not indexed (the User Guide gives, as examples of such records, preprints, Online Ahead of Print citations, and U.S. government-funded research published in journals not indexed for MEDLINE). However, having the word in the title or abstract is not the same as it being the paper's subject.

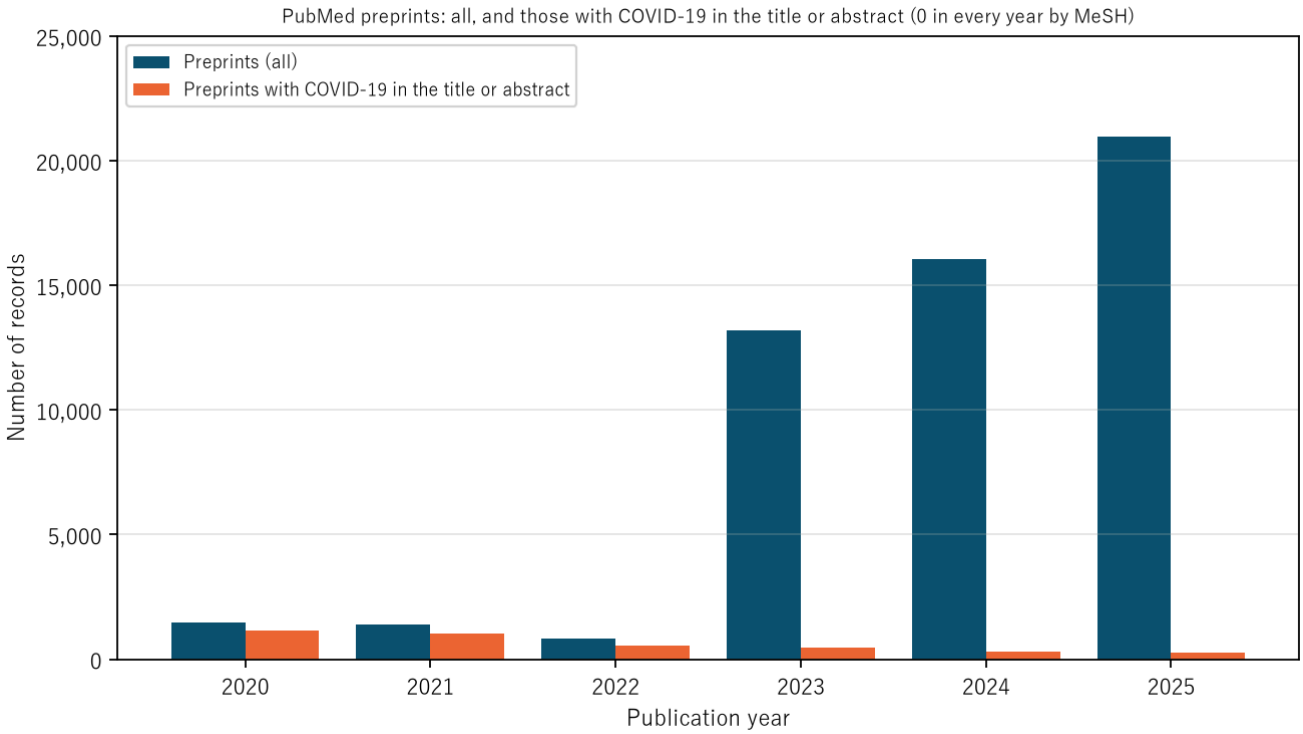


Figure 1: PubMed preprints (navy) and those with the word COVID-19 in the title or abstract (orange). Crossed with the COVID-19 MeSH term, they are 0 in every year.

---

6. Relation to earlier papers

M005 showed that the difference in how a disease is named affects the counts only when quotation marks are used. This paper showed that index terms (MeSH) and title/abstract words decide whether a whole group of records, preprints, is included or left out. M118 showed that the direction of a share changes with whether the labels lie within MEDLINE. The preprints in this paper are one example of labels outside MEDLINE.

---

7. Limitations

1. **Whether preprints are good or bad, and whether preprint results can be trusted, are not discussed.**

2. The counts in this paper are of preprints in PubMed (NIH-funded research), not all preprints. Counts from preprint servers were not used.
3. Counts by title/abstract word are counts of records containing the word, not counts of subjects.
4. The reason for the rise from 2022 to 2023 was not checked.
5. The counts are as of 27 September 2026 and change daily.

Conclusion

PubMed preprints appeared from 2020 and increased greatly from 2023. But counted by the COVID-19 MeSH term, preprints were 0 in every year, while counted by title/abstract word there were 1,151 in 2020 alone. Preprints are not indexed for MEDLINE, so index terms do not find them.

What separates them is one thing — whether you count by index terms or by title/abstract words.

When you search PubMed with a disease's MeSH term, **remember that preprints are not included.** To reduce what you miss, **also search by title/abstract words.** And preprints are reports that have not yet been peer reviewed.

Data and reproduction

Data	Counts and query translations retrieved from PubMed (U.S. National Library of Medicine) E-utilities (esearch), 27 September 2026, from 02:46:07. Courtesy of the U.S. National Library of Medicine. PubMed User Guide
Terms of use	NLM Terms and Conditions (free of charge; credit the source; do not imply NLM endorsement; state that the data are not the most current)
Coverage	PubMed publication type Preprint (publication years 2016–2026; months 2019–2021), COVID-19 by MeSH and by title/abstract word
Verification	`検算/M119.py` (`--zu` also draws the figure). Every number in the text is taken from `検算/M119_output.txt`
Raw data	`data/pubmed_m119_20260927.json` and `data/pubmed_m119_extra_20260927.json`, the retrieval scripts `data/fetch_pubmed.py` and `data/fetch_pubmed_extra.py`, `data/pubmed_userguide_20260927.html`, and `data/SOURCES.md` (terms of use, SHA-256, excerpts of the original text, and why preprint server counts were not used)
Figure	`図/M119_zu1` (png, svg, English version)

References

- U.S. National Library of Medicine. PubMed (E-utilities, esearch). Courtesy of the U.S. National Library of Medicine (retrieved 27 September 2026). https://eutils.ncbi.nlm.nih.gov/entrez/eutils/
- U.S. National Library of Medicine. PubMed User Guide (accessed 27 September 2026). https://pubmed.ncbi.nlm.nih.gov/help/
- Yamagishi, Y. (2026). Which Name You Use for a Disease Matters Only When You Put It in Quotes [M005]. Zenodo. https://doi.org/10.5281/zenodo.22963601
- Yamagishi, Y. (2026). The Share of PubMed Records with the Randomized Controlled Trial Publication Type Is 0.0164 in Both 1990 and 2025, but Divided Within Records Indexed for MEDLINE It Rises from 0.0174 to 0.0282 — 0.9997 of Those Labels Are in MEDLINE, and Only the Denominator's MEDLINE

Share Fell from 0.944 to 0.580. Only 0.4511 of Case Report Labels Are in MEDLINE, and How Much the Share Falls Depends on the Denominator [M118]. Zenodo (unpublished).