PubMed のプレプリントは 2020 年の 1,482 件から 2025 年の 20,965 件に増えたが、COVID-19 の MeSH と重ねるとどの年も 0 件だった——プレプリントは MEDLINE の索引を受けず索引の語が付かないので、COVID-19 の語を題名・要旨で数えると、2020 年だけで 1,151 件（その年のプレプリントの 0.777）ある [M119]

索引の語で見つかる記録 ＝ 索引を受けた記録だけ ── 分けるのは、索引の語で数えるか、題名・要旨の語で数えるか

山岸 勇輝（Yuuki YAMAGISHI）／ 独立研究者
ORCID 0009-0006-8587-3450
2026 年 9 月 ／ CC BY 4.0

要旨

PubMed（U.S. National Library of Medicine）の件数を E-utilities で取り（2026 年 9 月 27 日）、出版形式 Preprint の付いた記録（PubMed のプレプリント）を出版年・出版月ごとに数え、COVID-19 を索引の語（MeSH）で数えた数と、題名・要旨の語で数えた数を比べた。プレプリントの良し悪しや、研究の質は論じない。新しい数学的定理も新しい法則も主張しない。

第一に、**PubMed のプレプリントは 2019 年までの出版年では 0 件で、2020 年 1,482 件・2022 年 822 件・2023 年 13,200 件・2025 年 20,965 件と増え、2026 年も取得日までに 13,935 件あった。年を切らない数 67,864 件は、2020〜2026 年の年ごとの数の和とちょうど等しかった**（第 4 節）。

第二に、これが本稿の芯である。**COVID-19 の MeSH と重ねると、PubMed のプレプリントは 2016〜2025 年のどの年も 0 件だった。ところが題名・要旨の語（covid-19）で数えると、2020**

年 **1,151** 件（その年のプレプリントの **0.777**）・**2021** 年 **1,012** 件（**0.722**）で、**2020〜2025** 年の和は **3,732** 件あった（第 5 節）。

第三に、**理由は PubMed User Guide に書かれている。プレプリントは MEDLINE の索引を受けない記録で、MeSH は MEDLINE の索引で付く。実際、プレプリントのうち MEDLINE の中にあったのは 2016〜2025 年で 1 件だけだった**（第 5 節）。

第四に、**PubMed のプレプリントはプレプリントの全体ではない。User Guide によれば、2020 年 6 月から PMC に入った NIH の資金による研究のプレプリントの引用で、プレプリントの 0.9997 が PMC の中にあった**（第 4 節）。

第五に、**記録の全部でも、COVID-19 を題名・要旨の語で数えた数は MeSH で数えた数の 1.256 倍（2020 年）〜1.450 倍（2022 年）だった。索引の語で数えると、索引を受けていない記録は見つからない**（第 5 節）。

第六に、分けるのは **索引の語で数えるか、題名・要旨の語で数えるか** である。PubMed で病名の MeSH を使って調べると、プレプリントのような索引を受けない記録は出てこない。探し漏れを減らしたいときは、題名・要旨の語でも調べる。

射程注記：本稿は新しい数学的定理も新しい法則も主張しない——件数は PubMed で公開されており、本稿は数えて比べただけである。**本稿は公開の集計データの数の形を述べる構造的解釈であり、医学的な助言・診断・治療の推奨ではない。著者は医師ではない。個人の健康に関する判断は医師・医療機関に相談すること。査読を経ていないプレプリントである。 プレプリントの良し悪し、研究の質や価値を論じない。プレプリントの結果を信じてよいかどうかも論じない。特定の論文・雑誌・出版社・著者・プレプリントの置き場を名指さない。** 本稿の数は「PubMed に入っているプレプリント」の数で、プレプリント全体の数ではない。プレプリントの置き場の件数は、利用条件を確かめられなかったので使っていない。題名・要旨に covid-19 の語があることは、その論文が COVID-19 を主題にしていることと同じではない。件数は取得した時点のもので、日々変わる。**データについて**：Courtesy of the U.S. National Library of Medicine。件数は 2026 年 9 月 27 日に取得した時点のもので、NLM の最新・最も正確なデータを反映していない。NLM は本稿を推奨していない。本稿が加えるのは、**PubMed のプレプリントが索引の語の検索から丸ごと外れることを数で示し、題名・要旨の語で数えるとどれだけあるかを示したこと**である。

---

1. 記号と問い

- **PubMed のプレプリント**：`preprint[pt]`（出版形式 Preprint）。PubMed User Guide の原文は "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 の MeSH**：`"covid-19"[mh]`（索引の語）。**COVID-19 の語**：`covid-19[tiab]`（題名・要旨の語）。
- **MEDLINE**：`medline[sb]`。User Guide は、プレプリントは MEDLINE の索引を受けない記録の例だと書いている（"some citations are not indexed for MEDLINE, such as preprints, …"）。

問い：PubMed のプレプリントは、どれだけ増えたか。COVID-19 を MeSH で数えるのと、題名・要旨の語で数えるのとで、プレプリントの数はどれだけ違うか。

---

2. 下見と設計

正直に書く：在庫の段階では、プレプリントの置き場（medRxiv など）の件数で「プレプリントの伸び」を書くつもりだった。置き場の API の説明頁と説明頁を読んだが、API が返す件数を再利用するときの条件の記載を見つけられなかった。そこで、利用条件を確かめてある PubMed の出版形式 Preprint だけを数えることにした。

	内容
データ	PubMed E-utilities の件数（本取得 97 回・追加取得 15 回）、PubMed User Guide
本取得で気づいたこと	プレプリントかつ COVID-19 の MeSH が、どの年も 0 件だった
そこで足したもの	COVID-19 を題名・要旨の語で数える問い合わせ（プレプリントと記録の全部）
比べたもの	出版年・出版月ごとのプレプリント、PMC と MEDLINE の中にある割合、COVID-19 の MeSH と語

---

3. データと確かめ

計算 3.1 取得した件数の JSON 2 本・取得の道具 2 本・User Guide の写しの 5 本を SHA-256 で固定し、次を確かめた。検算は保存した JSON だけから計算し、PubMed には問い合わせない。

- 本取得 97 回（2026 年 9 月 27 日 02:46:07〜02:47:38）と追加取得 15 回（02:48:12〜02:48:26）がそろう。

- 読み替えに、出版形式（"preprint"[Publication Type]）・MeSH（"covid-19"[MeSH Terms]）・題名と要旨（"covid-19"[Title/Abstract]）・PMC・MEDLINE が種類どおりに入る。
- 警告は件数 0 の問い合わせの「No items found」だけ（43 回）で、式の読み違いを示す警告は 0 回。
- User Guide の写しに、プレプリントが NIH の資金による研究の引用であることと、MEDLINE の索引を受けないことの原文がある。
- 部分集合の向きがすべて成り立つ。

---

4. PubMed のプレプリントの数

計算 4.1 出版年ごとの PubMed のプレプリント

出版年	記録	プレプリント	PMC の中	MEDLINE の中
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

観察 4.2：PubMed のプレプリントは 2019 年までの出版年では 0 件で、2020 年 1,482 件から 2022 年 822 件に減り、2023 年 13,200 件・2025 年 20,965 件と増えた。2026 年も取得日までに 13,935 件あった。年を切らない数 67,864 件は、2020〜2026 年の年ごとの和とちょうど等しく、年のあいだの重なりは無かった。 2022→2023 年の増え方の理由は、件数からは分からない。

観察 4.3：プレプリントのうち PMC の中にある割合は 2016〜2025 年で 0.9997、MEDLINE の中にあるのは 1 件だった。 User Guide によれば、PMC がプレプリントを収め始めたのは 2020 年 6 月だが、出版日で数えると 2020 年 1 月の 5 件から、4 月には 104 件あった。出版日はプレプリントの日付で、PubMed に入った日ではない。

---

5. 芯 ── 索引の語で数えるか、題名・要旨の語で数えるか

計算 5.1 COVID-19 を MeSH で数えた数と、題名・要旨の語で数えた数

出版年	プレプリント：MeSH	プレプリント：語（プレプリントの割合）	記録の全部：MeSH	記録の全部：語（語 ÷ 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）

観察 5.2：COVID-19 の MeSH と重ねると、PubMed のプレプリントは 2016〜2025 年のどの年も 0 件だった。題名・要旨の語で数えると、2020 年 1,151 件（その年のプレプリントの 0.777）・2021 年 1,012 件（0.722）・2022 年 547 件（0.665）で、2020〜2025 年の和は 3,732 件だった。 プレプリントは MEDLINE の索引を受けない（User Guide）ので、索引の語である MeSH が付かない。

観察 5.3：記録の全部でも、COVID-19 を題名・要旨の語で数えた数は、MeSH で数えた数の 1.256 倍（2020 年）から 1.450 倍（2022 年）だった。 索引の語で数えると、索引を受けていない記録は見つからない（User Guide は、そうした記録の例として、プレプリントのほか、Online Ahead of Print の記録や、MEDLINE に索引されない雑誌に載った米国政府の資金による

研究を挙げている)。ただし、題名・要旨に語があることは、その論文の主題であることと同じではない。

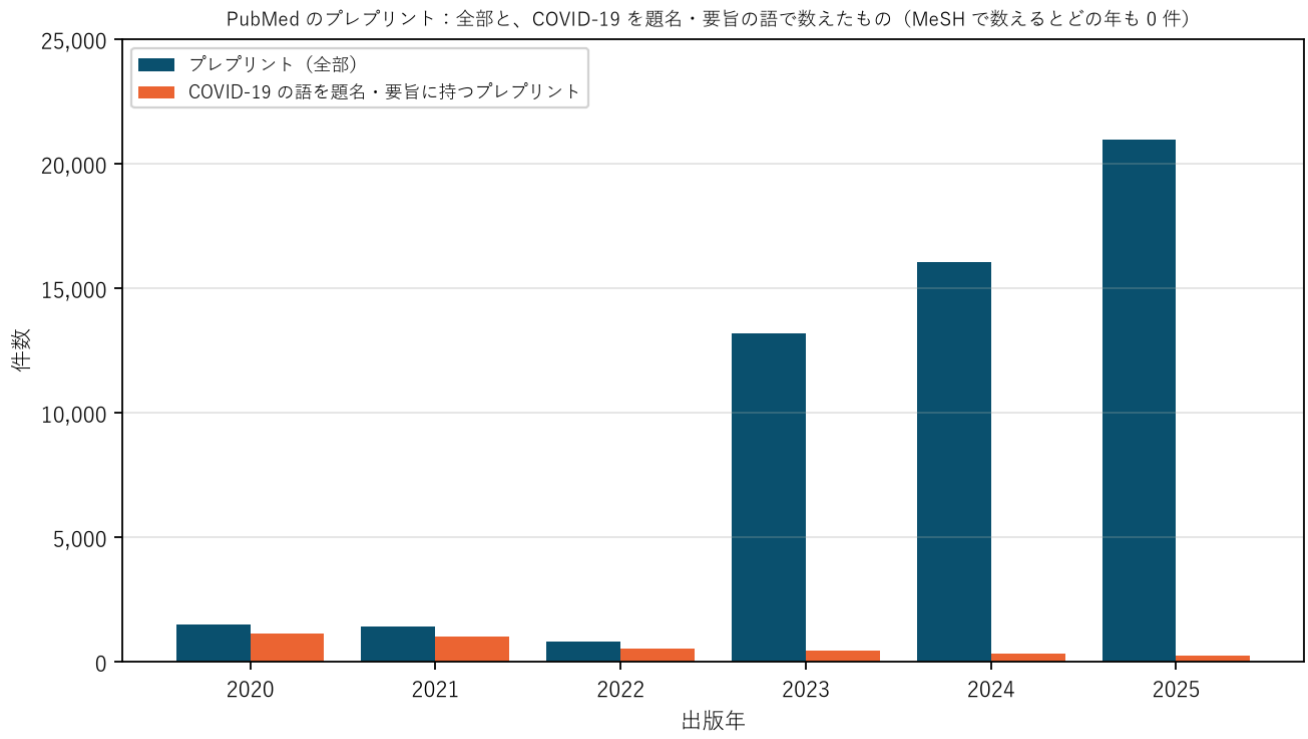


図1：PubMed のプレプリント（紺）と、そのうち COVID-19 の語を題名・要旨に持つもの（橙）。COVID-19 の MeSH と重ねると、どの年も 0 件。

---

6. 既刊との関係

M005 は、病名の言い方の違いが件数に効くのは引用符を付けたときであることを示した。本稿は、索引の語（MeSH）と題名・要旨の語で、プレプリントという記録の一群が丸ごと入るか外れるかが変わることを示した。M118 は、出版形式の札が MEDLINE の中にあるかどうかで割合の向きが変わることを示した。本稿のプレプリントは、MEDLINE の外にある札の一例である。

---

7. 限界

1. **プレプリントの良し悪しや、プレプリントの結果を信じてよいかを論じない。**
2. 本稿の数は PubMed に入っているプレプリント（NIH の資金による研究）の数で、プレプリント全体ではない。プレプリントの置き場の件数は使っていない。
3. 題名・要旨の語で数えた数は、その語を含む記録の数で、主題の数ではない。
4. 2022→2023 年の増え方の理由は確かめていない。
5. 件数は 2026 年 9 月 27 日の時点のもので、日々変わる。

結び

PubMed のプレプリントは、2020 年から現れ、2023 年から大きく増えた。けれども COVID-19 の MeSH で数えると、プレプリントはどの年も 0 件で、題名・要旨の語で数えると 2020 年だけで 1,151 件あった。プレプリントは MEDLINE の索引を受けないので、索引の語では見つからない。

分けるのは一つ——索引の語で数えるか、題名・要旨の語で数えるかである。

PubMed で病名の MeSH を使って調べたら、**そこにはプレプリントが入っていないことを覚えておく。** 探し漏れを減らしたいときは、**題名・要旨の語でも調べる。** そしてプレプリントは、査読を経る前の報告である。

データと再現

データ	PubMed（U.S. National Library of Medicine）の E-utilities（esearch）で取った件数と読み替えの式（2026 年 9 月 27 日 02:46:07〜に取得・Courtesy of the U.S. National Library of Medicine）。PubMed User Guide
利用条件	NLM Terms and Conditions（利用は無料・出典を明記・NLM の推奨を示唆しない・最新のデータでないことを明記）
範囲	PubMed の出版形式 Preprint（出版年 2016〜2026・出版月 2019〜2021）、COVID-19 の MeSH と題名・要旨の語
検算	`検算/M119.py`（`--zu` で図も描く）。本文の数はすべて `検算/M119_output.txt` から取った
生データ	`data/pubmed_m119_20260927.json`・`data/pubmed_m119_extra_20260927.json`、取得の道具 `data/fetch_pubmed.py`・`data/fetch_pubmed_extra.py`、`data/pubmed_userguide_20260927.html`、`data/SOURCES.md`（利用条件・SHA-256・原文の抜き書き・プレプリントの置き場の件数を使わなかった理由）
図	`図/M119_zu1`（png・svg・英語版）

参考文献

- U.S. National Library of Medicine. PubMed（E-utilities, esearch）. Courtesy of the U.S. National Library of Medicine（2026年9月27日に取得）. https://eutils.ncbi.nlm.nih.gov/entrez/eutils/
- U.S. National Library of Medicine. PubMed User Guide（2026年9月27日に利用）. https://pubmed.ncbi.nlm.nih.gov/help/
- 山岸勇輝 (2026). 病名の言い方の違いが効くのは、引用符を付けたときである [M005]. Zenodo. https://doi.org/10.5281/zenodo.22963601
- 山岸勇輝 (2026). ランダム化比較試験の出版形式が PubMed の全記録に占める割合は 1990 年も 2025 年も 0.0164 で変わらないが、MEDLINE の索引が済んだ記録の中で割ると 0.0174 から 0.0282 に上がる——その札の 0.9997 は MEDLINE の中にあり、分母の MEDLINE の割合だけが

0.944 から 0.580 に下がった。症例報告は札の 0.4511 しか MEDLINE の中になく、下がり方が分母で変わる [M118]. Zenodo（未公開）.