

医学分野のプレプリント論文とその情報流通

「科学技術」グループ

2022年3月14日

【若者の緊急提言】 コロナ禍で見た日本の課題と解決策
第6期ジュニア・アカデミア政策提言発表会

主催 日本アカデミア

「科学技術」グループ研究報告書

テーマ

医学分野のプレプリント論文とその情報流通

1. 解決すべき課題

信頼できる学術知を確立する一方でその活用をより広く、そして素早く行うというのは古くて新しい課題である。とくに新型コロナウイルスの流行とそれに向けた対応の中で、この問題はより注目を集めることになったと思われる。パンデミックで人命が失われるという未曾有の危機に際して、学術コミュニティや出版社側は平常時の知識創出の枠組みから、より知識共有の迅速性を重視する方向へ舵をきることで対応してきた。例えば、著名な医学雑誌として知られる New England Journal of Medicine は rapid review と呼ばれる制度¹をフルに活用することで1~2週間での論文の掲載決定を可能にした。プレプリント論文、つまり正式な査読を経ないで公開される論文は以前からも理学や数学の分野で存在していたが、迅速かつ低コストで知の共有ができるという点で議論の対象とされることが多くなり、プレプリントサーバーへの論文の投稿数も早いペースで増加している。下のグラフ²は生物学分野のプレプリント論文が投稿される bioRxiv の投稿数を示しているが、ここ数年で急激に伸びていることが確認できる。そのような中、プレプリント論文を最終版だと宣言する学者が出てくるなど、学術知の生産についての根本的な問い直しの声も上がってきている。しかし、プレプリント論文について述べられていることの多くは期待を寄せつつも信頼性に注意が必要であると述べるにとどまっており³、それがどのような貢献をなし、そしてなしうるものなのか、なにができないのかについて考察し、適切な役割を果たせるような仕組みを作る試みは不足しているように思われる。

¹ New England Journal of medicine, Editorial policies

[Editorial Policies | About NEJM](#) retrieved on 28 February

² 科学技術・学術政策研究所「bioRxiv に注目したプレプリントの分析」(2021) p7

³ あくまで一例であるが、[新型コロナ: 査読前論文、コロナで注目 信頼性確保がカギ: 日本経済新聞 \(nikkei.com\)](#) (2022年2月28日取得) は一般的な見方をまとめている。

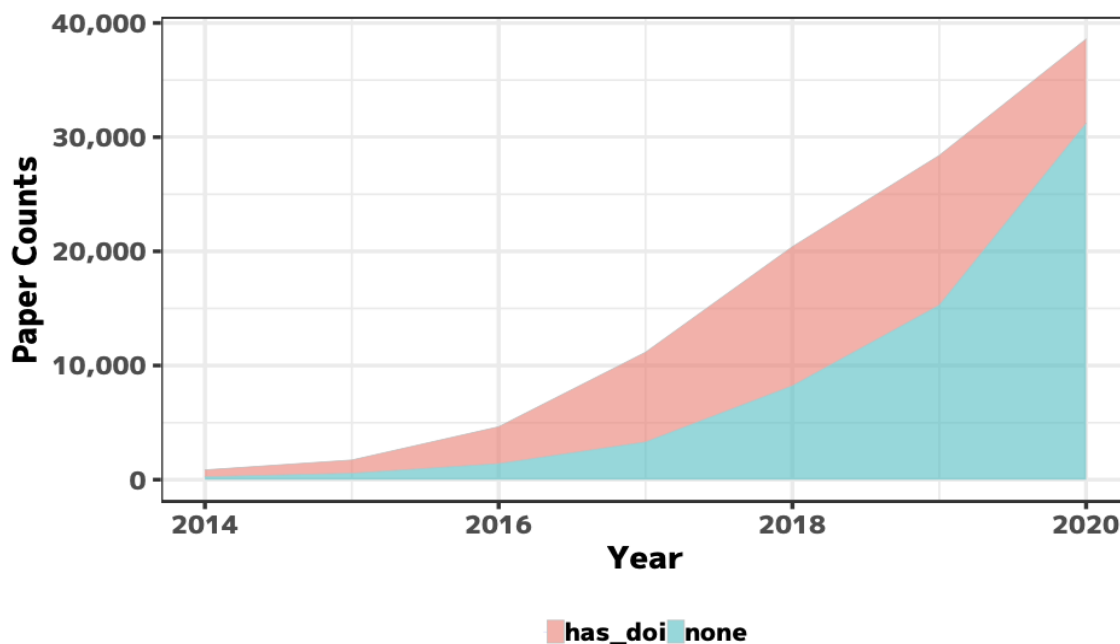


図2 年・ジャーナル DOI 有無別の投稿数推移

しかし、遍く存在する学術分野のプレプリント論文に関してそのような考察をすることは時間、能力の点で不可能であることから、本研究では課題の範囲を医学分野に絞り、とくにコロナウイルス関係の論文を手掛かりとして考察を進めることとする。医学分野における学術知は場合によっては人命を左右しうる問題であることから、リスクコミュニケーション的な観点にも注意を払いつつ検討していきたい。

2. 問題意識

(1) プレプリント概観

そもそもプレプリント論文とはなにか。その名前が示すようにプレプリント論文は学会等による正式な査読を経るまえの段階の論文のことをさす。一般的にはプレプリントサーバーとよばれるサーバーに投稿されることが多い。最近登場したもののようにも思われるが、数学や物理学等の分野でarXivというプレプリントサーバーが開設されたのは1970年代のことであり、その後ほかの分野に拡大していったという歴史がある。一般論としてプレプリント論文には以下のようなメリット、デメリットがある。

まずメリットとしては①知識が公開されるまでの速さと簡素さ、②DOI付与による研究先取り権の確定、③開放性がある。①について、査読は場合によっては数か月かかることもあるが、プレプリントはそれを経ないため当然素

早く公開することができるし、まただれでも公開することができる。これは若手の研究者に成果を世に出す機会が与えられるという点でも重要である。②については、DOI にタイムスタンプもついてくるため、誰が先に研究成果を出したかということが一義的に明確になるということである。③はプレプリントサーバーにアクセスするのに査読雑誌におけるようなコストはかからないため、より広く論文に書かれた知識へのアクセシビリティを確保することができる⁴。一方でデメリットとしては当然ながら査読がないため質が低いものが混ざりがちであるという点があげられる⁵。たとえば 2020 年にコロナウイルスと HIV の遺伝的類似性を指摘する論文が bioRxiv に投稿されたが、理論的厳密性を大きく欠き、様々な学者からの批判を受けて数日後に撤回されるということがあった⁶。

(2) 医学分野のプレプリントの動向について

では、医学分野のコロナウイルス関係のプレプリント論文の動向はどのようなものであったか。新型コロナウイルス流行前後では投稿数は劇的に変化した。2019 年 1 年間の medRxiv (医学のプレプリントサーバー) の投稿数は 789 本だったのに対し、2020 年 1 月から 2022 年 1 月の 2 年間で medRxiv に投稿された新型コロナに関する論文は 16,607 本となっている⁷。比較のため既存の雑誌でのコロナウイルス関係の論文、論考の数を調べてみると、New England Journal of medicine は 2020 年 4 月から 2022 年 1 月までで 257 本⁸、Nature とその姉妹雑誌については 2020 年 2 月から 2022 年 2 月までで 6359 本⁹、Lancet については 2020 年 1 月から 2021 年 3 月までで 534 本である¹⁰。このように情報量の観点からプレプリントの数が圧倒的に勝っていることがわ

⁴ 科学技術・学術審議会情報委員会「プレプリントをめぐる近年の動向及び今後の科学技術行政への示唆」(2020) p6 [プレプリントをめぐる近年の動向及び今後の科学技術行政への示唆 \(mext.go.jp\)](#) (2022 年 2 月 28 日取得)

⁵ 前掲注 4、p7

⁶ 実際に撤回された論文は [Uncanny similarity of unique inserts in the 2019-nCoV spike protein to HIV-1 gp120 and Gag | bioRxiv](#) で確認できる。

⁷ medRxiv “OVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv”から執筆者がカウント (<https://connect.medrxiv.org/relate/content/181>)

⁸ [The New England Journal of Medicine: Research & Review Articles on Disease & Clinical Practice \(nejm.org\)](#) から執筆者がカウント

⁹ [Nature](#) から執筆者がカウント

¹⁰ Johns Hopkins University, [2019 Novel Coronavirus Research Compendium \(NCRC\) \(jhsph.edu\)](#) retrieved on 28 February

かる。

これらのデータに加え、ジョンズ・ホプキンス大学の科学者が Norvel coronavirus research compendium という取り組み¹¹において 2019 年に新型コロナに関連する論文を厳選した結果約 216 本しか残らなかったことを踏まえると、プレプリント論文が既存の学術雑誌と比較してより大量の情報を発信し流通させているものの、価値がある論文は非常に少ないと推論することができる。実際、医学臨床分野の専門家からは、自身はプレプリント論文を活用しておらず、かつ臨床分野においてプレプリント論文が活用された事例は聞いたことがないと伺った。一方、基礎研究（ウイルスの構造解析など）においては先取権の確保という観点でプレプリントサーバーは一定の役割を担うと考えられる。

(3) プレプリント論文のもたらした混乱

プレプリント論文は迅速な情報発信が可能である。しかし最近では各報道機関や SNS におけるプラットフォーム上でも多くのプレプリント論文が取り上げられ、それらの中には玉石混交の石も多く含まれ結果として不確かな情報を流通させることになる。大手新聞機関ではプレプリントを報じる際に現在ではプレプリント論文について真偽に注意すべき旨を示しているが、SNS 上ではプレプリントである指摘はなく不確実なまま流通している。例えば 1 で挙げた事例ではあるが、インド・ニューデリーの科学者らは、新型コロナウイルスと HIV が「不気味なほど」似ているとする研究結果を bioRxiv に投稿した¹²。世界中の科学者から批判を浴びて速やかに撤回されたが、筆頭著者がツイッターで宣伝し、加えてハーバードの医療経済学者もリツイートしていたことによって、既に 1 万 7 0 0 0 件以上のツイートに引用され、25 のニュースメディアが取り上げていた。このツイートには、論文がプレプリントであるため情報に不確実性に注意すべきである等の注意書きは見られない。

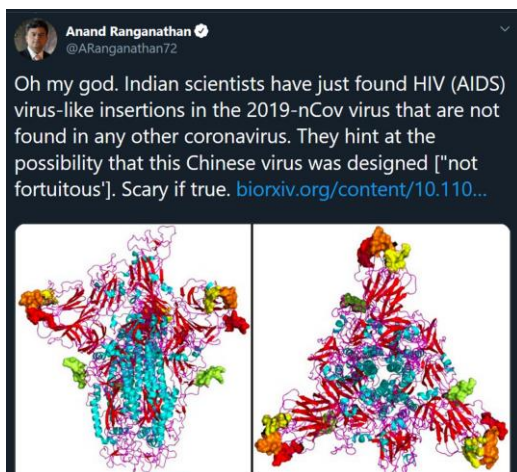
¹¹

¹² Is It Fake News or Is It Open Science? Science Communication in the COVID-19 Pandemic

<https://journals.sagepub.com/doi/full/10.1177/1050651920958506>

Uncanny similarity of unique inserts in the 2019-nCoV spike protein to HIV-1 gp120 and Gag

<https://www.biorxiv.org/content/10.1101/2020.01.30.927871v1>



その他にもプレプリントがもたらした不確実な情報があたかも真実であるかのように流布された事案は多く存在する。イベルメクチンのコクランレビュー¹⁴はプレプリントを対象としている。また、COVID-19の治療法として、ヒドロキシクロロキンやクロロキンを含むマラリア薬が有効であるという情報が流布され、一部の当局により調査結果は却下された¹⁵ものの、Professor Didier Raoult が個人的に行った研究に関するプレゼンテーションのYouTubeビデオを投稿し、140万回以上再生されている¹⁶。対してこの調査に関する査読付き論文は8万回ほどしかダウンロードされていない¹⁷。このようにプレプリントによる情報は不確実なものが多いにもかかわらず、その情報はSNSやYouTube等で「不確実性の指摘なく」拡散されてしまう。その上特にSNSの大手プラットフォーム上の各ポリシーではプレプリント論文の掲載に関する規定は見当たらない。SNSが私たちの生活に深く浸透し、迅速かつ簡単にアクセスできる点に鑑みると、今後各SNSのプラットフォームにおけるポリシー規定をプレプリント論文の勃興に対応した形での改定が必要になってくる。

¹³ Don't believe the conspiracy theories you hear about coronavirus and HIV

<https://massivesci.com/notes/wuhan-coronavirus-ncov-sars-mers-hiv-human-immunodeficiency-virus/>

¹⁴ コクランは、治療と予防に関する医療情報を検証している機関である。

<https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD015017.pub2/full>

¹⁵ COVID-19: Could Hydroxychloroquine Really Be An Answer?

<https://www.medscape.com/viewarticle/927033>

¹⁶ Coronavirus : diagnostiquons et traitons ! Premiers résultats pour la chloroquine

[Coronavirus : diagnostiquons et traitons ! Premiers résultats pour la chloroquine](https://www.youtube.com/watch?v=Ugk1Uj1t8p4)

¹⁷ In Vitro Antiviral Activity and Projection of Optimized Dosing Design of Hydroxychloroquine for the Treatment of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)

<https://academic.oup.com/cid/article/71/15/732/5801998?login=false>

(4) 公開性の問題

さらに根本的な観点としては、medRxiv や bioRxiv をはじめとしたプレプリントサーバーがオープンアクセスの理念のもとに無料で誰でも簡単にアクセスできるものになっていることについても、もちろん科学をより開かれたものにしていくという観点では価値のあるものであるが、一方で課題も存在すると考えられる。先に議論したようにプレプリントの主要な価値は基礎研究における先取権の確保に見出される。一方で、薬品の効能のような広く関心をもたれやすい応用分野におけるプレプリントについては貢献というよりはむしろ害をもたらした側面が大きい。よって、プレプリントの役割を基礎研究分野における先取権の確定に特化させ、潜在的な危険を最小化するためにプレプリントサーバーのプレプリントを誰がみることができるのかということについて検討していく必要があると思われる。

そこで本提言ではプレプリント論文を考慮した各 SNS のポリシーとより根本的な課題としてプレプリント論文の公開範囲について提言を行う。

3. 現在みられる課題解決策とその問題点

(1) SNS のポリシーについて

a. Google、YouTube

まず Google は新型コロナウイルス感染症発生当初から新型コロナウイルスに関する広告掲載のポリシー策定とその不正に関して策を講じてきた。特にパンデミックに乗じて利益を得ようとするコンテンツや世界的な公衆衛生の危機に対して配慮を欠いたコンテンツを禁止している¹⁸。Google が管理している YouTube も新型コロナウイルスに関する深刻な危害を及ぼすコンテンツの共有を禁止している。

b. Yahoo

Google と同様に Yahoo もミスリードを招いたり他の利用者に危害を加えるコンテンツは禁止されている¹⁹。

c. Twitter

Twitter は新型コロナウイルスに関する意図的な陰謀を想起させる情報や健康被害につながり得る承認されていない治療法や予防法に関する虚偽の主張や

¹⁸ <https://support.google.com/google-ads/answer/9811449?hl=ja>

¹⁹ <https://news.yahoo.co.jp/newshack/policy/publishing/>

誤解を招く主張や情報の共有を禁止している²⁰。特に Twitter は利用者に対してポリシーに違反し、深刻な健康被害をもたらすことが判明したツイートを削除するように要請するとしている。さらにこのような情報を発信したアカウントを一時的にロックし警告としてカウントされる。こうしたことに加えて新型コロナウイルスに関する誤解を招く情報がツイートに含まれている場合、正しい情報が掲載されたラベルを当該ツイートに適用することがある。

d. Facebook

また Facebook についても新型コロナウイルス関連のコンテンツの削除に取り組んでおり、例えば新型コロナウイルスに関する存在や深尺度、確実な治療法や予防方法に関する偽情報や正しい衛生習慣を阻害する情報は削除される²¹。

e. TikTok

TikTok に関しては、ガイドラインの詳細について詳しくは一般に公開していないが（引用：新型コロナウイルス（COVID-19）やワクチンに関連する誤った情報や、反ワクチンの情報を含む虚偽または誤解を招くようなコンテンツを禁止）、虚偽情報や誤解を招くような情報を拡散するコンテンツに関しては、特別なトレーニングを受けたチームがアカウントを特定し、削除している²²。ユーザーが誤情報を見つけた場合には、アプリ内で報告することが可能であり、また Politifact、Lead Stories、SciVerify、AFP などのファクトチェックパートナーも存在している。事実、2021 年 4 月から 6 月の間、新型コロナウイルスに関する誤情報を発信した 27,518 本の動画を削除し、この内 71% の動画が一度もユーザーに視聴されることなく削除されている。

(2) プレプリントサーバーによる事前のチェック

プレプリントサーバーは査読を経る前の論文を掲載するものではあるものの、本当になんでも自由に投稿できるというわけではない。各サーバーにより違いこそあるものの、一定程度事前のチェックを働かせることが一般的である。ここでは、医学領域に関係するプレプリントサーバーとして medRxiv と bioRxiv について見ていきたい。

a. medRxiv

ピアレビュー並みに厳密なものではないが、剽窃や非科学的な内容、不適

²⁰ <https://help.twitter.com/ja/rules-and-policies/medical-misinformation-policy>

²¹ <https://www.facebook.com/help/230764881494641>

²² <https://www.tiktok.com/safety/ja-jp/covid-19/>

切な記事のスタイル、公衆の健康をそこなう恐れのある内容についてスクリーニングがかけられるとしている。²³

b. bioRxiv

medRxiv と同じく、剽窃や不適切表現、非科学的な内容などの項目について簡単に事前チェックがされる。²⁴

(3) 残っている問題点

a. SNS ポリシーについて

上述した通り各 SNS ポリシーでは新型コロナウイルスに関する誤った情報や誤解を招く情報の共有を可能な限り防ぐための取り組みを行っているが、本提言で調査した SNS のプラットフォームのいずれにもプレプリント論文に関する規定は存在しなかった。現に Yahoo は新型コロナウイルスに罹患した患者の嗅覚の喪失に関するプレプリント論文の研究²⁵を注意書きなどなしに掲載し、そのままの状態となっていた。

新型コロナウイルスやフェイクニュース問題への注目を受けて各種 SNS の誤解を招く情報には策が講じられているものの、コロナ禍により急増したプレプリント論文のもたらす真偽の不確かな情報に対応できるものとはなっていないことがわかる。

b. サーバーによる事前チェックについて

上記の事前チェックは当然簡易なものである。なぜなら査読並みに厳密に事前審査をおこなってしまえばそもそもそれは査読雑誌とやっていることが変わらなくなり、プレプリントサーバーとしての存在意義を失ってしまうことになるからである。そしてこのチェックはコロナにおけるプレプリントの勃興においてあまり役割を果たすことができなかつたと言わざるを得ない。2 の問題意識のところでもみたように、実際には公衆にとってリスクのあるようなプレプリントの情報が拡散されてしまった。サーバーによる事前チェックはリスクコントロールの仕組みとしては不十分であり、より実効性のある対策が望まれるということがわかる。

²³ [Frequently Asked Questions \(FAQ\) | medRxiv](#)

²⁴ [Frequently Asked Questions \(FAQ\) | bioRxiv](#)

²⁵ <https://www.yahoo.com/lifestyle/50-percent-people-covid-long-145549538.html>

50 Percent of People With COVID Get This Long-Term Symptom, Study Says

4. グループとして考える課題解決策と、課題解決までの道筋

(1)プラットフォームのポリシー改定

プラットフォーム事業者に対し、ユーザーがプレプリントの内容に関するツイート等の情報発信をする場合は、その情報が不確かであることを言及しなければならないなどのガイドラインを策定するように求めることが一つの策として考えられる。確かに EU(フランスやドイツ)ではフェイクニュースへの対策が行われているが、プレプリントは不確実性を有しているという点でフェイクニュースとは一概には断定できないためそういったグレーな事象への対応は未成熟である。第一にユーザーに対してプレプリント論文そのもの、もしくはそれにつながるリンクを掲載する場合に、「この情報は査読を経していないプレプリント段階での情報であり、今後、内容が修正される可能性がある(日付)」と付記しなければならないということを義務付ける。これに違反したユーザーに対してはプラットフォーム事業者がアカウントへ罰則(行為の繰り返しに対する警告等の軽微な罰則が望ましい)を科すこととし、加えて元のツイートの「このツイートはプレプリント段階の情報である」等のマークを付けることとする。このマークを実際のプレプリント論文にも義務的に付けるものとし、全体的なプレプリントの標準マークにしていくことも将来的には考えられる。第二にプレプリント由来の情報リストを事業者が作成し公開することで、どの情報が不確かなのかを一般的に周知することも必要である(その情報が検証された結果どうなったかも継続して記載する必要がある)。

また、ユーザーがプレプリント論文そのもの、もしくはそれにつながるリンク等を貼らずに拡散する場合も考えられる。しかしこの発信が問題となるときには、すでに上記のリストが作成済みであると考えられる。このため、問題となる情報の重要単語を含むツイートを全てピックアップするアルゴリズムを事業者が作成し(例: COVID-19とHIV)、そのツイートに対し上記のリストに飛ぶための警告文をつけることで、一般のユーザーに情報共有を図ることができる。

ここで注意すべきは、ツイートに対し削除命令を出すなどの行為をしてしまうと、情報規制である等の反対意見が出る可能性があるということである。プレプリント段階の情報は「不確かであり誤っている可能性がある」だけであって「完全に誤っている」といういわゆるフェイクニュースではないので、削除命令まで出すべきではなくマークを付ける等で充分である。この方法には、すでに拡散されている情報の根源の発信元(最初のツイート主)にマークを付けることで、全てのリツイートの投稿にもマークがつくことになり、結果的に拡散ルートを逆利用した「情報修正の拡散」になるということも意

図している。これは、新型コロナウイルスがHIVと酷似しているという情報が拡散した事例において、情報元の論文やツイートが削除されたにもかかわらず、情報が正しいとして発信され続けた事例を教訓としている。

(2) サーバーへのアクセス制限

プレプリントサーバーによる事前のチェックを超えて、medRxivなどのサーバー側でアクセスのできる主体を一定の範囲に制限するということが二つ目の策として考えられる。アクセス可能な人の範囲としては、そのサーバーが取り扱っている学問領域の学会に加盟している人という切り分け方にするのがよいと考えられる。このようにすることで、プレプリントサーバーに投稿する際、読み手の方もそれなりに知見があるということを想定しなければならないので、功名心に基づいて不確かな論稿を出そうというインセンティブは著しく低下することが想定される。これはもちろんプレプリントのかかげるオープンサイエンスの理念には逆行するものであるが、そもそも医学領域におけるプレプリントサーバーの意義は基礎研究について誰が先にその研究にたどり着いたかということを確認化することであり、学者の内輪をこえて広く関心をもたれやすい応用研究における重要性は低いことから、アクセスの制限はプレプリントサーバーが先取権の明確化という医学領域において果たすべき役割に専念できるようにする措置であるということが出来る。

5. 課題解決策の効果・副作用・残された課題

(1) ポリシー改定

各SNSプラットフォームのポリシーにプレプリント論文の項目を追加することで、プラットフォーム上で流通するプレプリント論文が整理され、混乱を引き起こすことを抑止することが可能となる。フェイクニュースや誹謗中傷対応への文脈において、既にプラットフォーム自身で流通する情報について必要な制限を課すことは社会的に求められている一方で、表現の場として過度の制限にならないよう常に配慮することが求められる。提言ではラベル（注意書き）付与による注意喚起を促しているが、ラベル付与によるユーザーの使いやすさの低減は招きかねないので、工夫が求められる。加えて、アルゴリズムの活用など多くの方法でプレプリント論文の情報を区別できるよう提言しているが、必ずしも全てのプレプリント論文関連の情報を網羅できるとは限らない。

また、上記の施策の基盤としての情報教育も欠かせない。プレプリント論文とはいかなるものなのかを一般の方々が理解していただけるよう、報道で伝え続けたりアプリダウンロード時に明記することも重要である。

(2) サーバーへのアクセス制限

アクセス制限により、研究者以外の知識のない人が生命に関わる不確実な情報に触れ、世間にその情報を流通させる事態を防ぐことが可能となる。アクセスできる研究者がサーバーで得た情報を SNS 等で流すことも考えられるが、SNS でのポリシー改定により対応できるため、プレプリント論文の不確実な情報が世間に流出し、混乱を招く事態を防ぐことが可能となる。一方で、プレプリントサーバーには、既存の査読付きジャーナルの購読料が非常に高額であり誰でもが最新の学術知にアクセスできるわけではないという現在の査読誌の問題の一つの解決策としての側面や、素早い研究データの共有による研究の加速という利益も一部では見込まれていた。アクセス制限によりこれらのオープンアクセスの利益が失われる。しかし、上述したようにコロナ禍でプレプリントを根拠とする治療薬についての誤った情報が流布し健康被害が発生しうる事態が生じたことを踏まえてオープンアクセスの利益と社会への弊害とを比較衡量すれば、医学臨床分野においては生命が優先されるべきである。

6. メンバー（氏名のみ記載）※リーダーに☆、サブリーダーに○を付ける

☆岸原大樹	○村上将隆
羽深英	來山歩美