



論文や研究データに対する研究者の 態度変化と質の担保

文部科学省 科学技術・学術政策研究所 (NISTEP)

データ解析政策研究室

室長 林 和弘

研究官 尾崎 翔美



研究成果

科学のベンチマーク

大学のベンチマーク



研究人材

博士人材追跡調査



研究活動

NISTEP定点調査



イノベーション

イノベーション調査



産業界

民研調査

科技予測調査

科学技術指標

その他

- オープンサイエンス等の新たな研究の潮流をとらえた調査研究
 - ◆ （発表1）「研究データの公開と研究データ管理に関する実態調査 2024：データ公開の拡大とリポジトリ活用の深化」
→ 研究データに対する研究者の態度の変化を報告
- データサイエンス、AI関連技術等を用いた、新たな調査研究・データ解析手法の開発
 - ◆ （発表2）「撤回論文の動向調査：OpenAlexとRetraction Watch Databaseを用いた分析」
→ 研究論文の質に関する調査を報告
- オープンサイエンスが指向する科学、社会、科学と社会の変容を促す対話と共創の場の構築

（STI Horizon誌の編集・発行）



研究データの公開と研究データ管理に関する実態調査2024： データ公開の拡大とリポジトリ活用の深化

文部科学省 科学技術学術政策研究所 (NISTEP)
データ解析政策研究室
池内有為 (客員研究官)
林 和弘 (室長)



1. 調査の概要



研究者

2016

Open Data

OA

2018

Open Data

OA

2020

Open Data

OA

2022

Open Data

OA

Preprints

2024

Open Data

OA

Preprints



専門家NW

第5期科学技術基本計画
第6期科学技術・イノベーション基本計画
統合イノベーション戦略

Preprints



研究機関

RDM
Data Policy

JPCOAR/AXIES

RDM
Data Policy

JPCOAR/AXIES

RDM
Data Policy

JPCOAR/AXIES

対象 科学技術専門家ネットワーク(大学・公的機関・団体、企業)

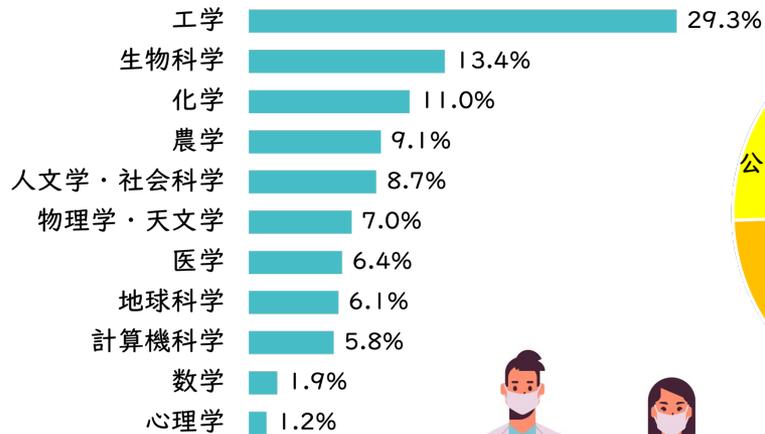
方法 オンライン質問紙調査 (Cuenote)

期間 2024年11月5日~2024年12月25日

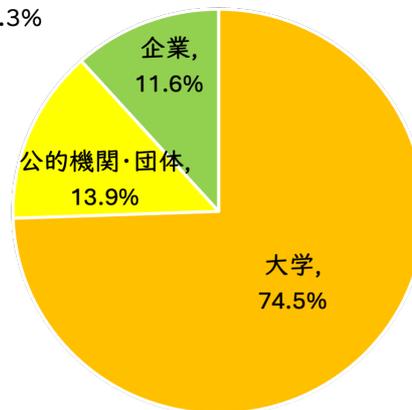
有効回答 1,237 (回答率 73.9%)



研究分野



所属機関



年齢層

年代	回答者数	比率
30代以下	180	14.6%
40代	613	49.6%
50代	353	28.5%
60代以上	91	7.4%
合計	1,237	100.0%



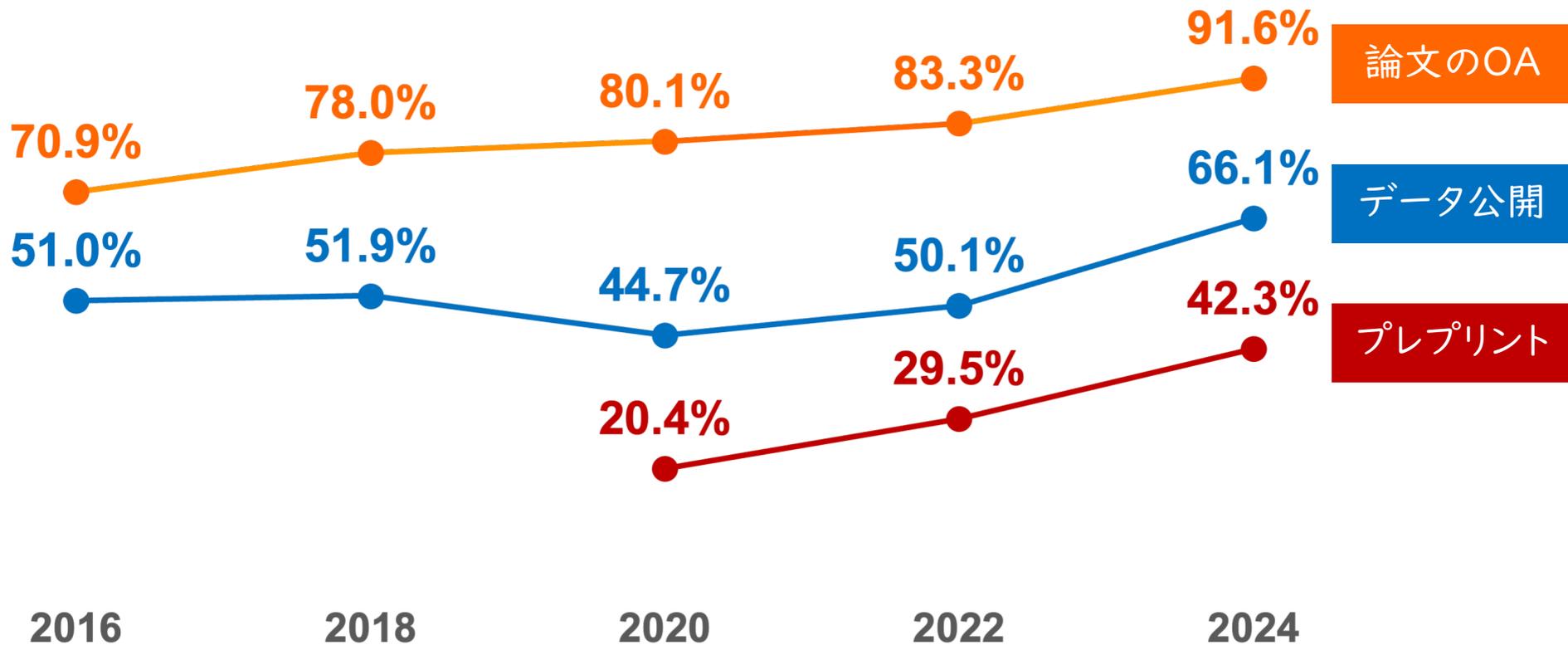
質問項目

章	カテゴリ	問題数
1	回答者に関する質問	2
2	研究データの提供(共有)状況	2
3	公開データの入手状況	6
4	データマネジメントプラン(DMP)の作成状況	4
5	研究データの公開状況	7
6	研究データの整備・公開・保存(RDM)について	4
7	データ公開に対する評価	2
8	研究領域・方法について	1
9	自由回答	1
	合計	29



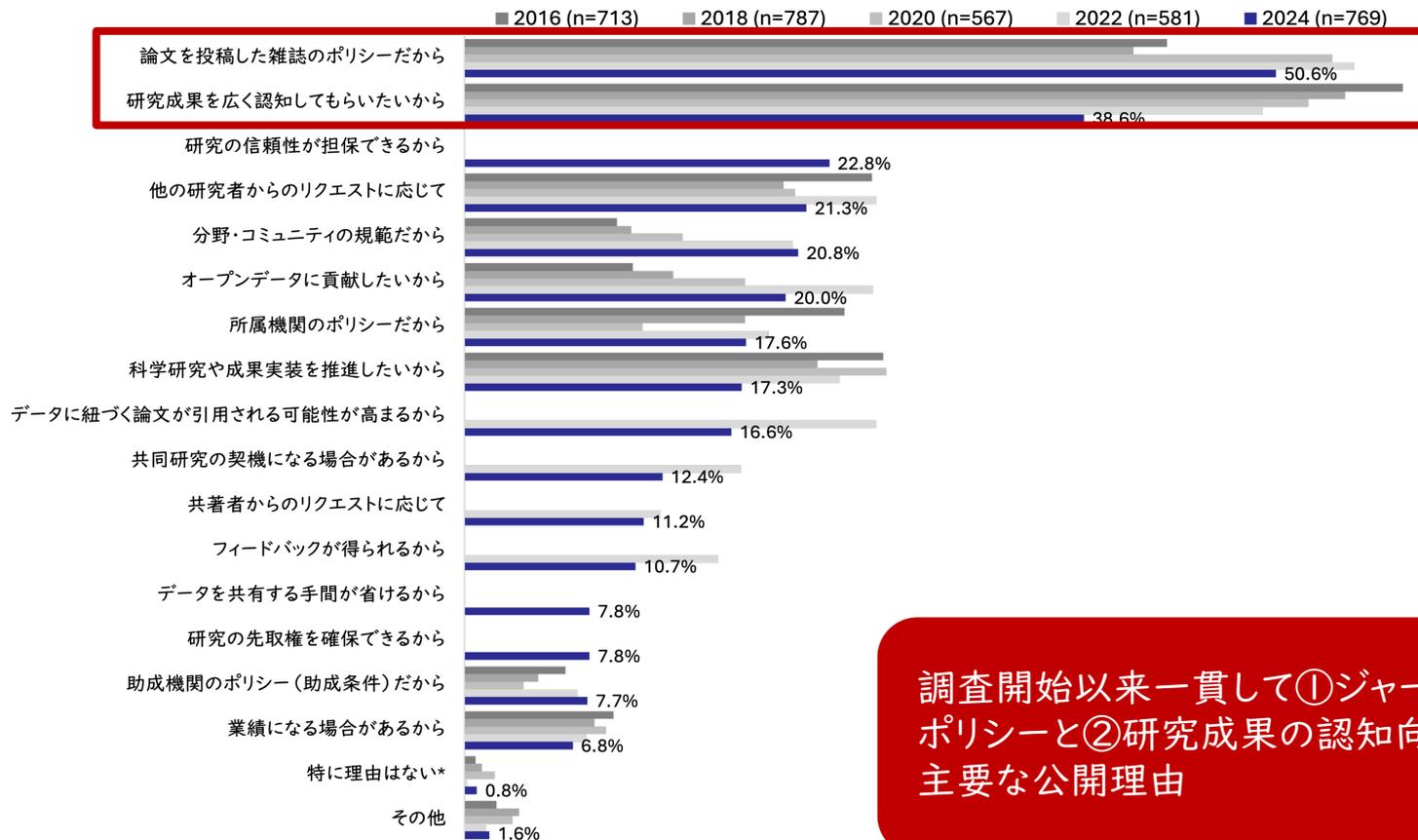
2. 主要な結果

データ公開経験の経年変化 (2016-24)



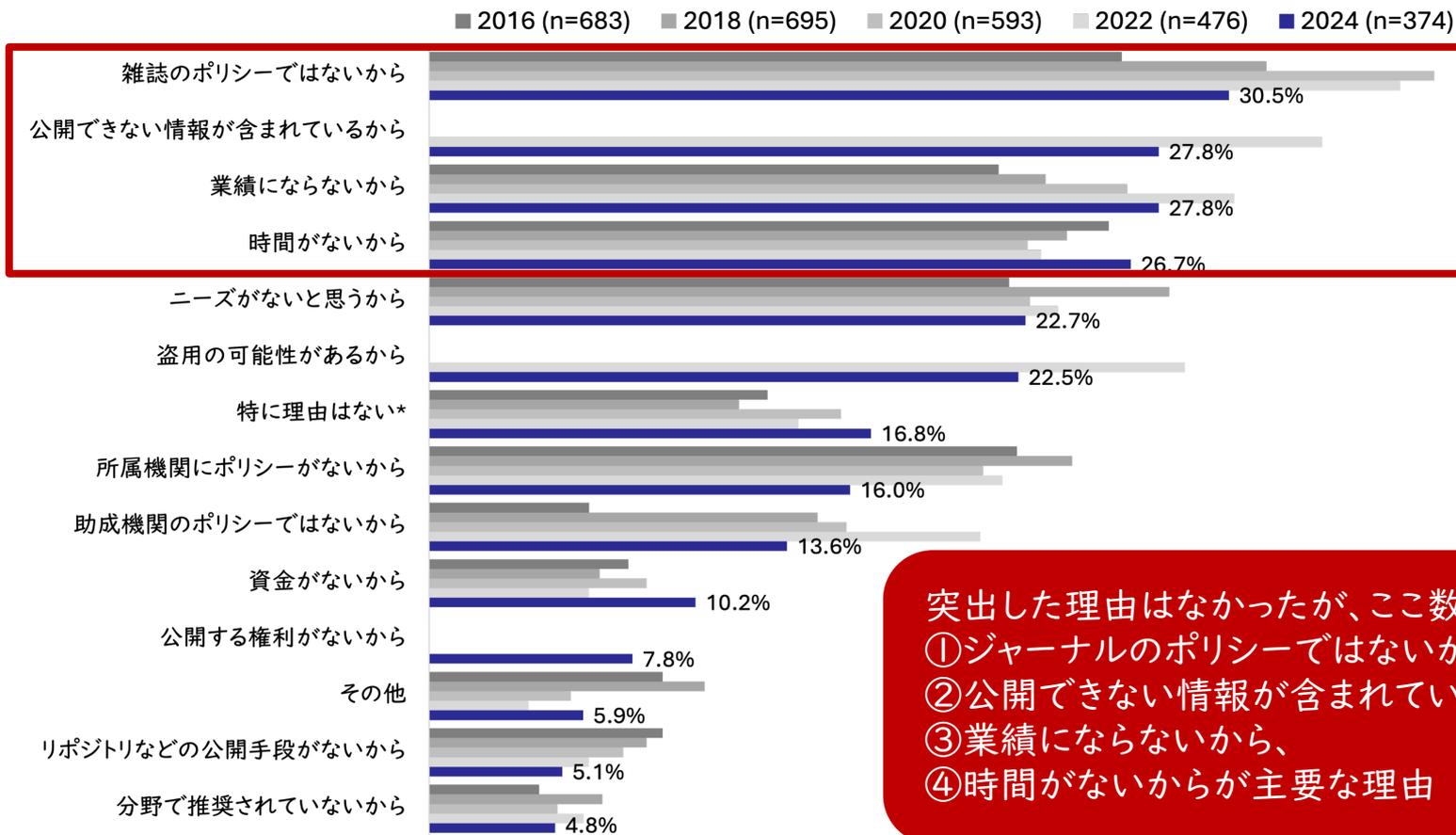
*論文のOAはGold OAに限らず、Green OA（機関リポジトリ、セルフアーカイブ）やSNSでの公開も含む

データ公開理由（複数選択）

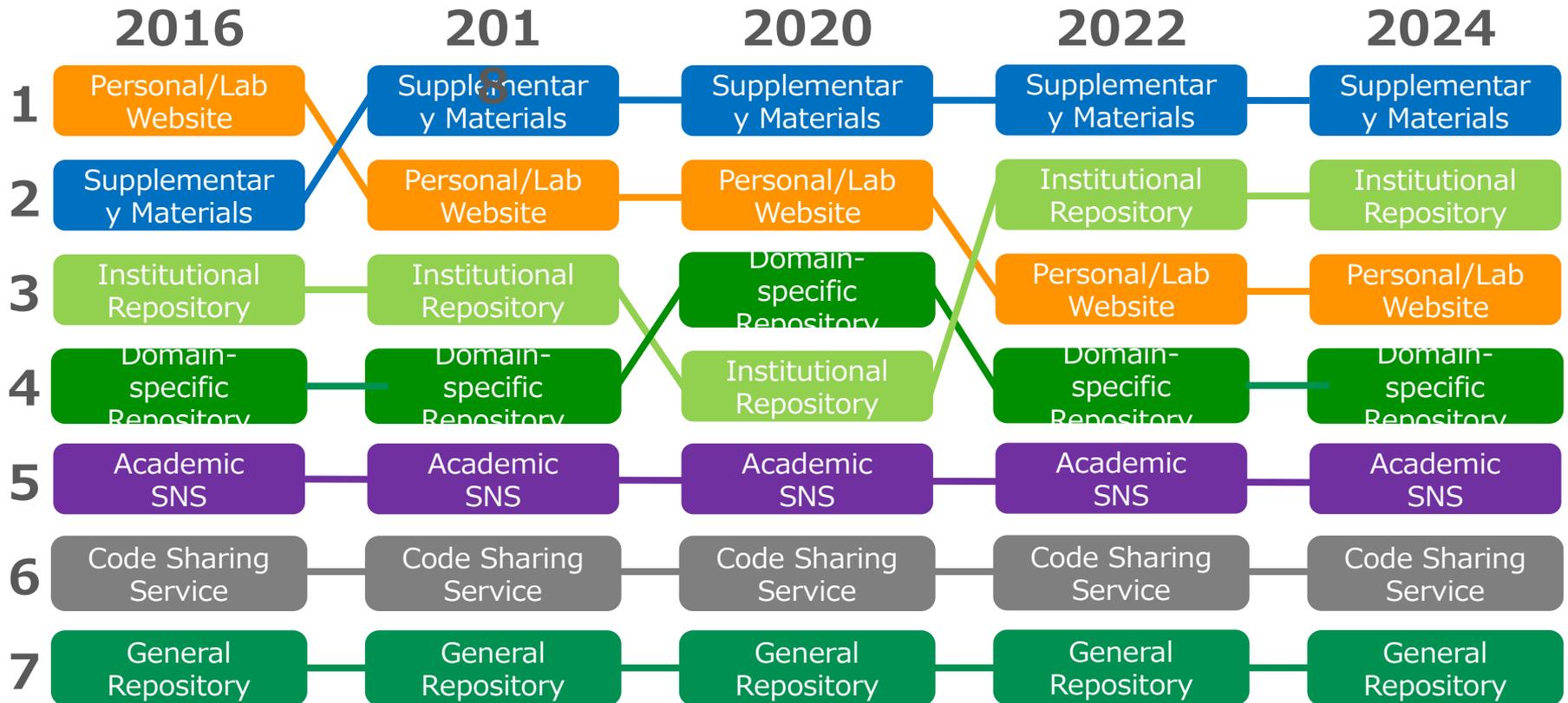


調査開始以来一貫して①ジャーナルのポリシーと②研究成果の認知向上が主要な公開理由

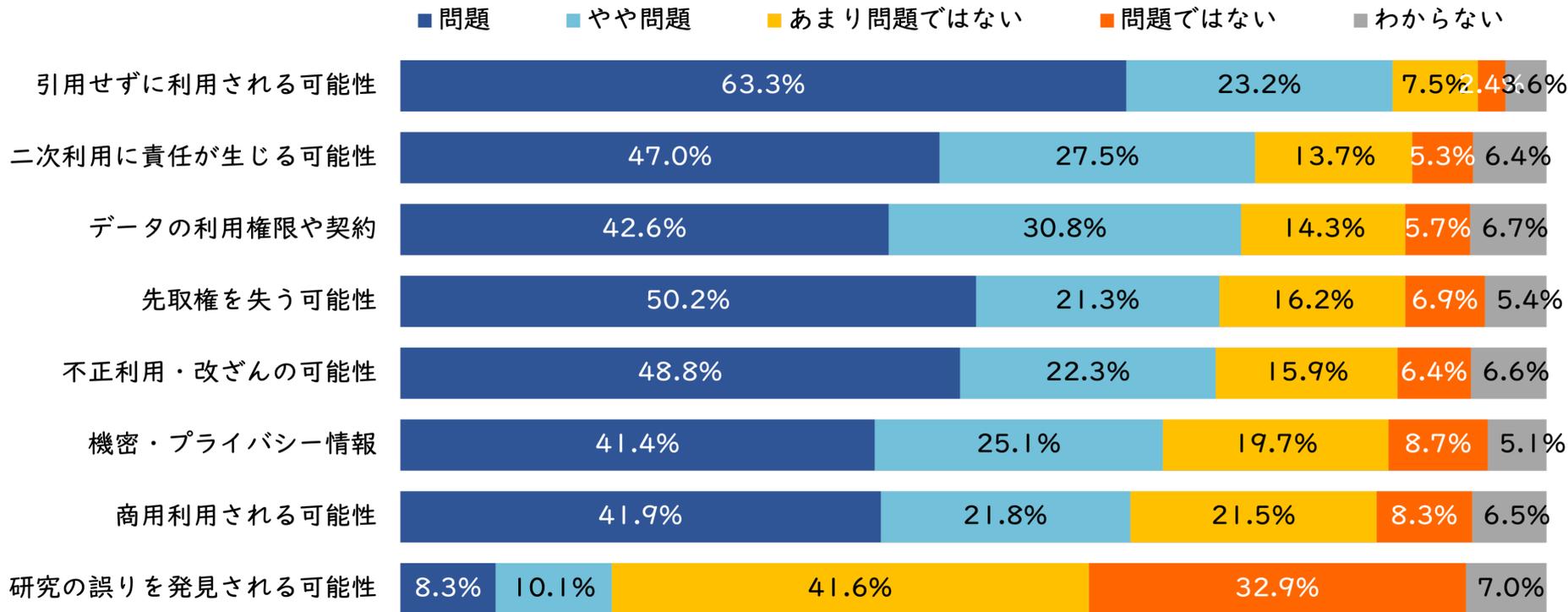
データ未公開理由（複数選択）



突出した理由はなかったが、ここ数年は
 ①ジャーナルのポリシーではないから、
 ②公開できない情報が含まれている、
 ③業績にならないから、
 ④時間がないからが主要な理由



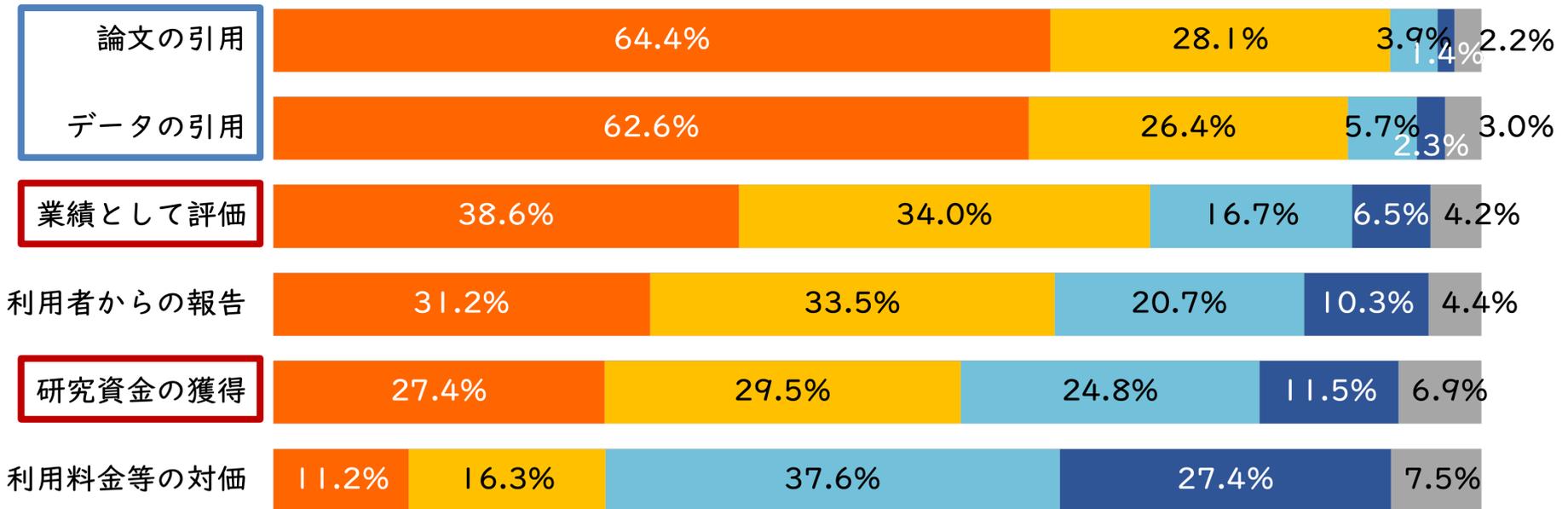
データを公開する場合の懸念 (n=1,167)



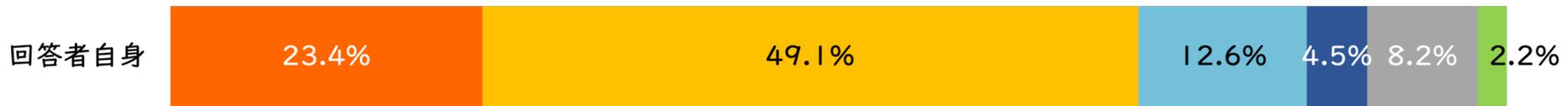
データ公開経験をもつ研究者は増加したものの依然として懸念が強い

データ公開のインセンティブの重要度 (n=1,167)

■ 重要
 ■ やや重要
 ■ あまり重要ではない
 ■ 重要ではない
 ■ わからない



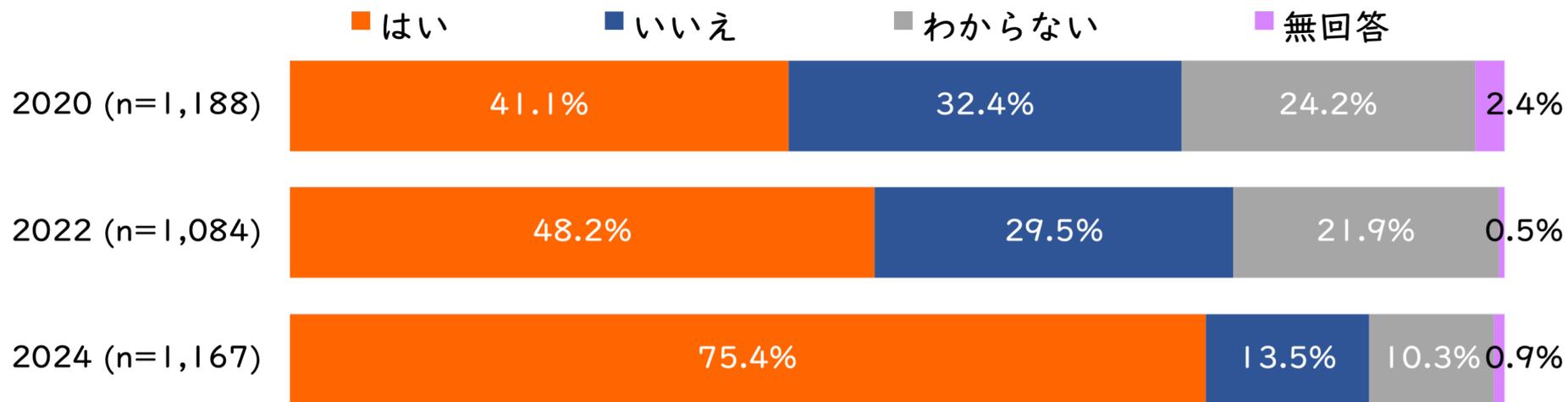
- 評価するべきである
- ある程度評価するべきである
- あまり評価するべきではない
- まったく評価するべきではない
- わからない
- その他



- 評価していると思う
- やや評価していると思う
- あまり評価していないと思う
- 全く評価していないと思う
- わからない
- その他

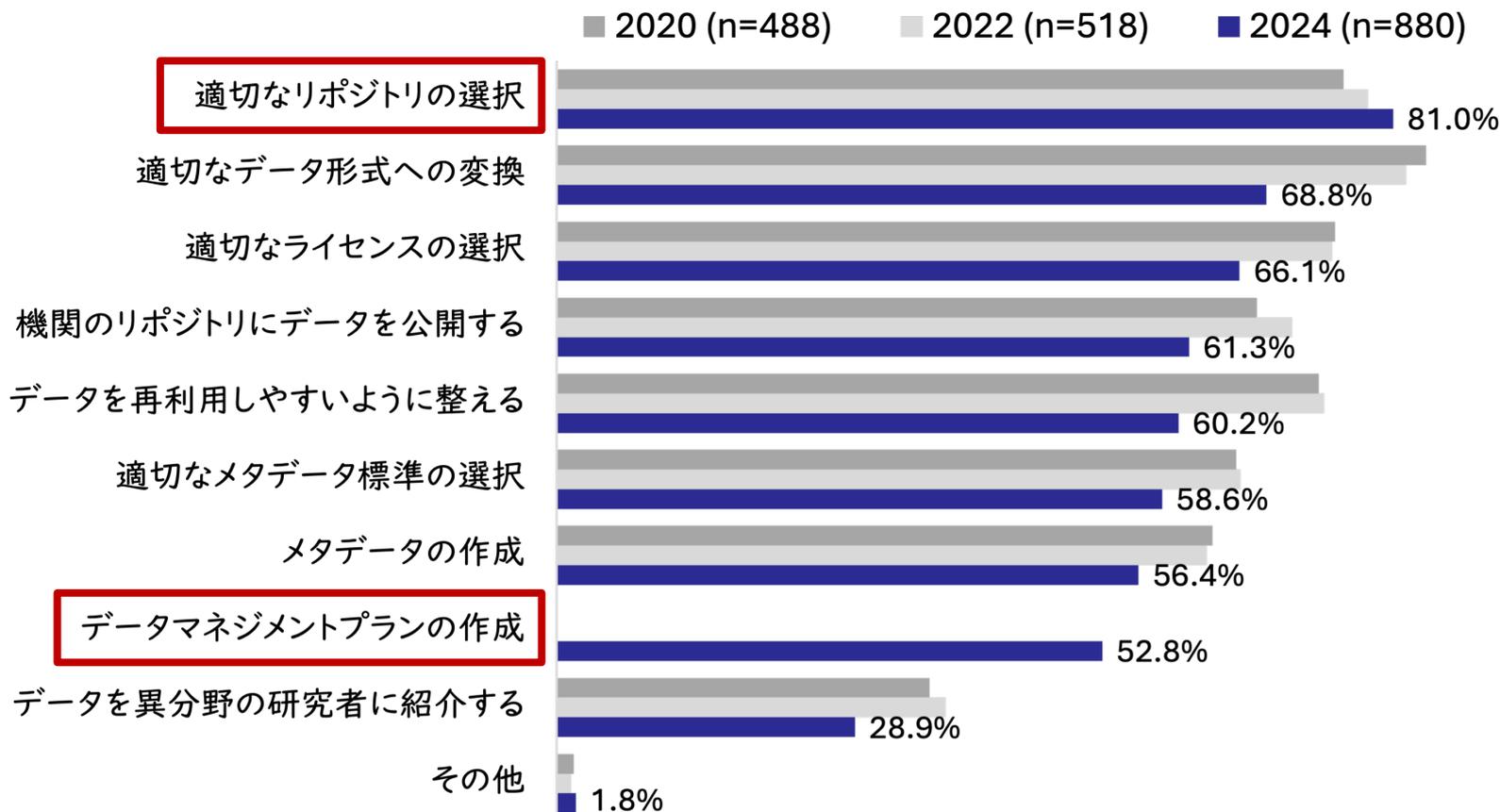


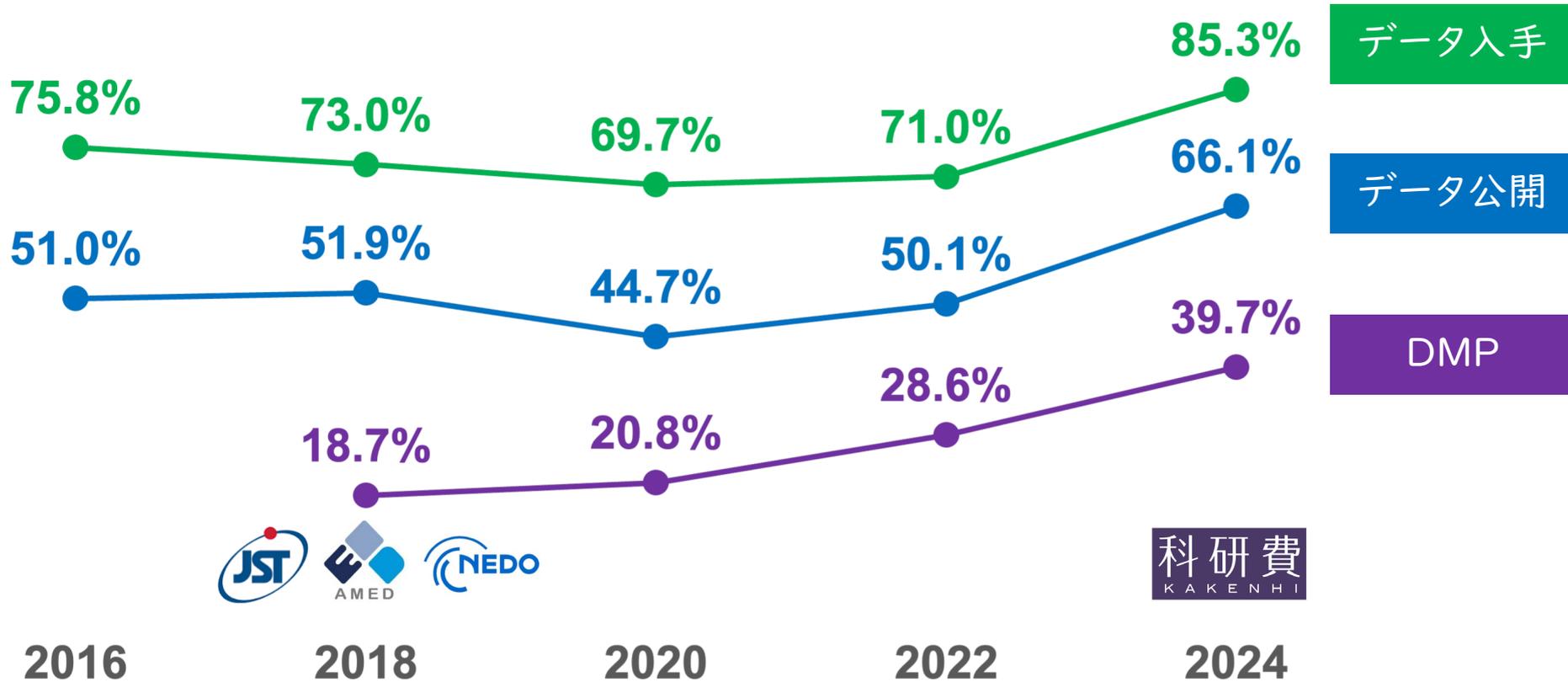
データ公開を業績として評価するべきであると考える研究者は72.5%
所属機関が業績として評価していると認識している研究者は16.4%

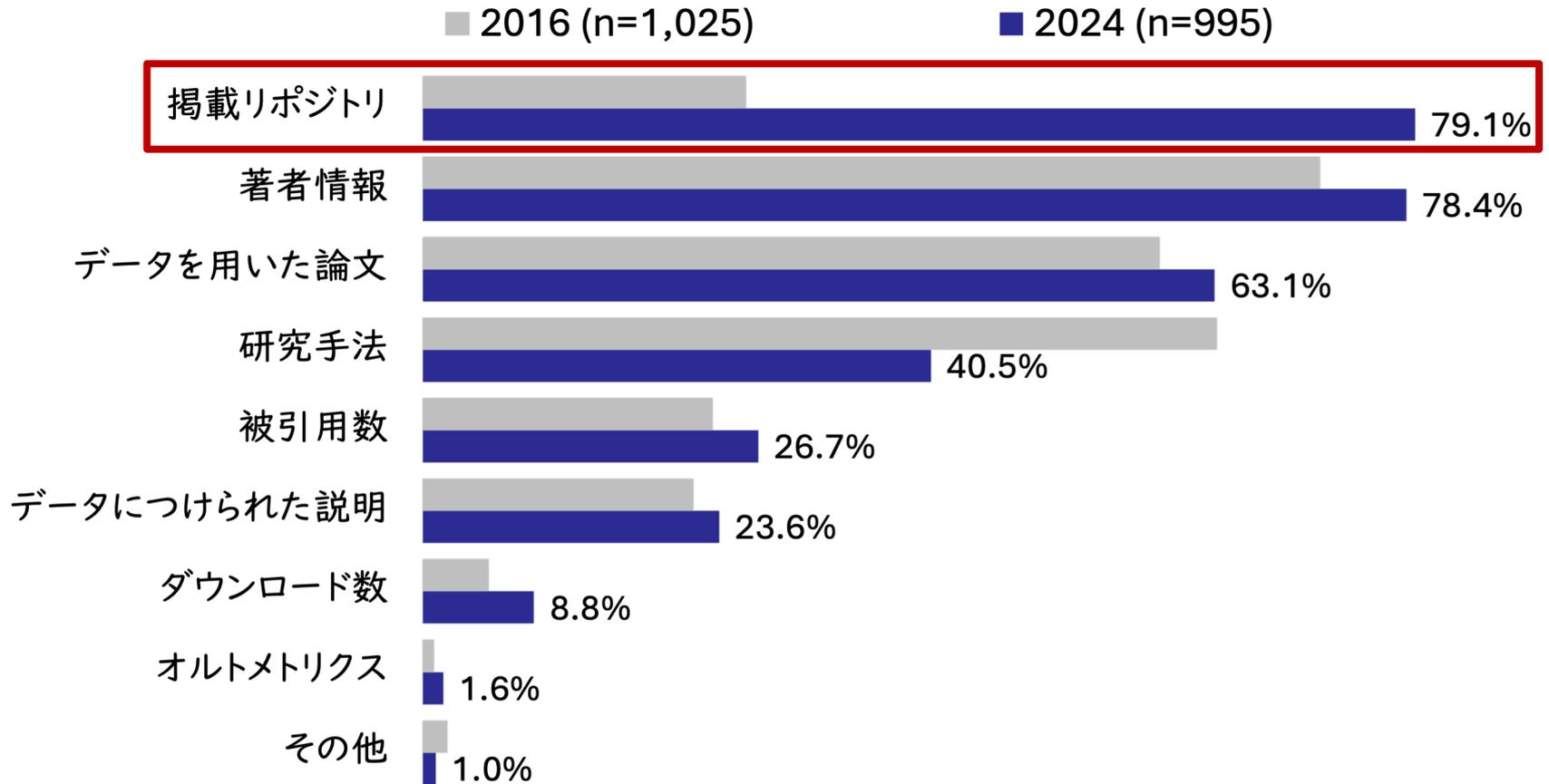


データ管理を第3者に任せたい研究者の増加（図書館やURAの支援の可能性）

データ管理を依頼したい内容（複数選択）







■ 研究データ公開の拡大

- ◆ ジャーナルのポリシーに応じた補足資料やリポジトリでの公開
- ◆ 依然としてデータ公開への懸念は強く、インセンティブは見えずらい

■ 研究データの流通におけるリポジトリ活用の深化

- ◆ データ公開・入手先として活用が進む
- ◆ データの信頼性の判断材料に
- ◆ 適切なリポジトリ選択に対する支援が求められる

■ 政策の効果と今後の課題

- ◆ データマネジメントプラン（DMP）の作成経験が増加
- ◆ 資源は依然として不足、第三者への依頼意思の向上
- ◆ 業績として評価することがデータ公開のインセンティブとなる可能性



撤回論文の動向調査:

OpenAlexとRetraction Watch Databaseを用いた分析

文部科学省 科学技術学術政策研究所 (NISTEP)

データ解析政策研究室

尾崎 翔美

■ 撤回論文のマクロな動態の調査を実施

- ◆ 研究成果DB OpenAlex と 撤回論文DB Retraction Watch Data とを掛け合わせて集計
- ◆ 2025年5月にNISTEP Discussion Paper として公開

■ 国・地域別の撤回論文数とジニ係数を計算

- ◆ 日本では少数の研究者が大量の論文撤回を行っている可能性が高い
- ◆ 他方、中国等では大量の研究者が少数の撤回を行っている可能性が高い

■ 撤回論文に関係する問題の一部を簡単に紹介

- ◆ 分野ごとの撤回数
- ◆ 撤回理由
- ◆ 撤回に掛かる期間
- ◆ 撤回論文が他の研究に与える影響（引用）

■ 一度公開されたものの、なんらかの理由によって撤回された論文

- ◆ 意図的ではないミス (e.g. 計算ミス, 図の誤り)
- ◆ 意図的なミス (≡研究不正 e.g. 剽窃, データ捏造)

■ 体系的な日本語文献は少ない

◆ 参考：撤回論文に関する研究動向

- 撤回論文が引用され続けることの問題(Bar-Ilan & Halevi, 2017, Schneider et al., 2020)
- 撤回論文の内容はSNS等で一般に拡散されやすい(H. Peng, D.M. Romero,& E. Horvát 2022)
- 統一的な論文撤回のフォーマット作成の必要性の示唆(Bar-Ilan & Halevi, 2017)
- 論文を一度でも撤回した若手研究者のキャリアパス
- 生成AIを用いた不正論文の事前検出 等

■ Retraction Watch (RW)

- ◆ 学術論文の撤回や研究の公正性に関する問題を報告するサイト
- ◆ “Center for Scientific Integrity”の1部門（米国NPO, 2010年～）
- ◆ 2023年9月にCrossref に買収され、データ無料公開を開始

■ OpenAlex

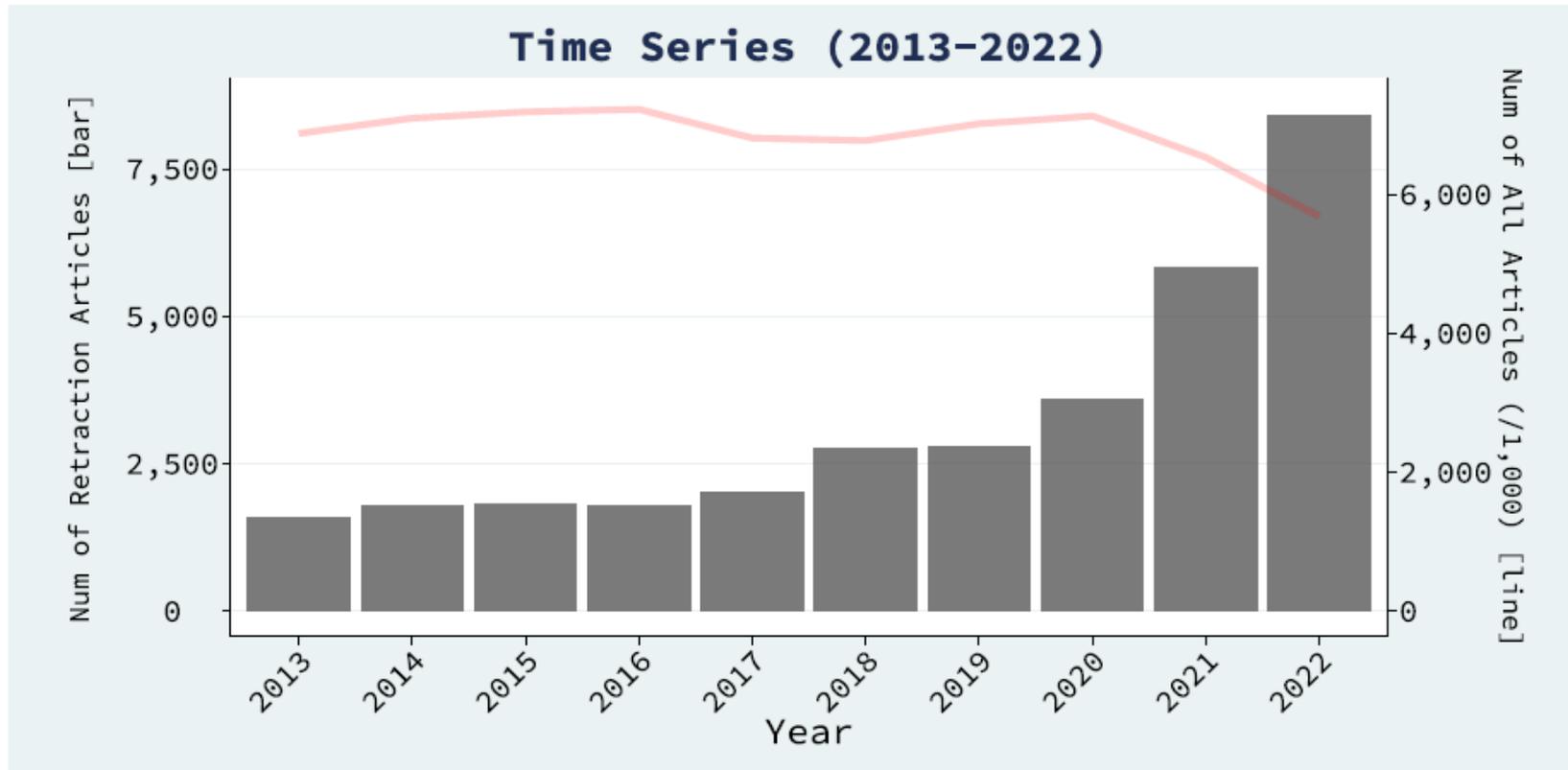
- ◆ 非営利団体Our Researchが学術研究の透明性とアクセシビリティの向上のため2022年に公開した無料の書誌情報公開データベース
- ◆ 学術論文のみならず書籍、プロシーディング、プレプリントの情報も含む



■ 2つのDBから相補的に情報を取得し、 約32,000件の独自のDBを構築

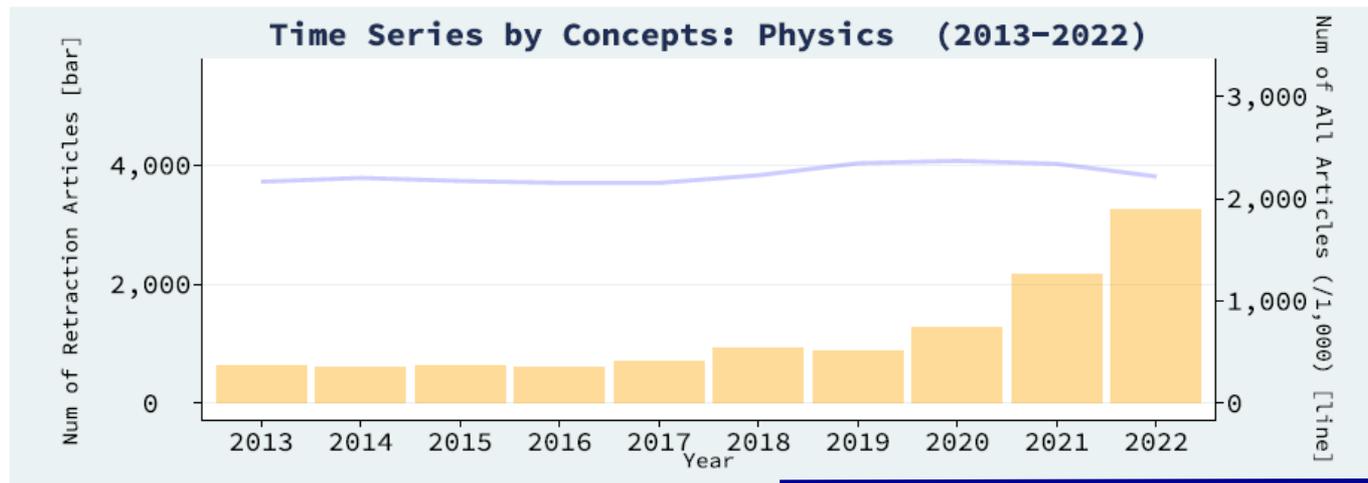
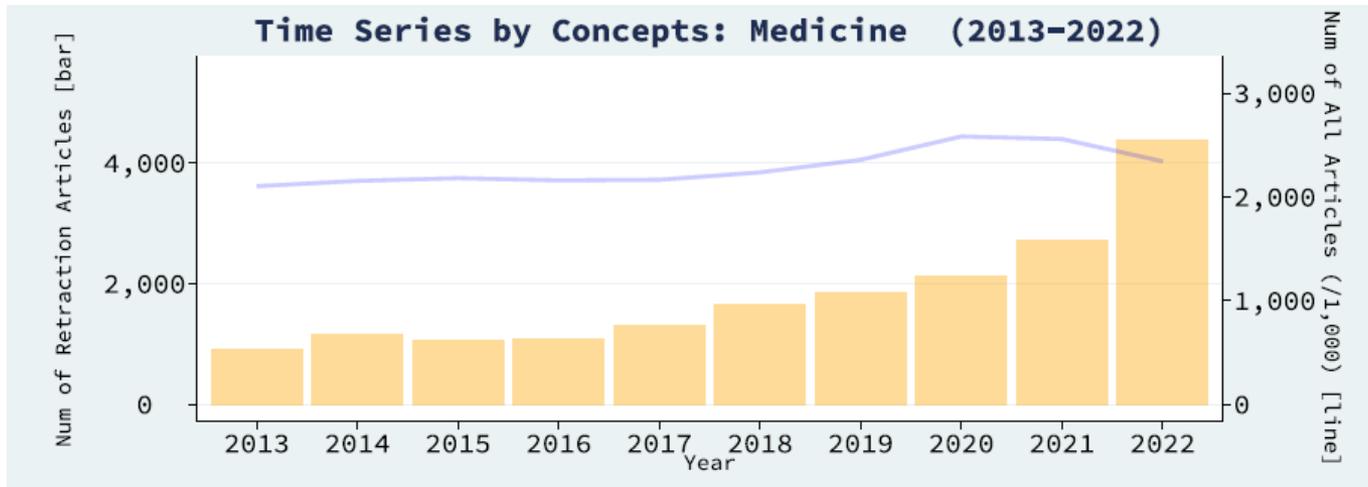
OpenAlexは RW の情報を取り込んでいるが、撤回理由などを取り込んでいなかったり、一部情報の不整合があることなどが知られている。

- 撤回論文数は年毎に増加傾向に対し、OpenAlex論文総数は微減傾向
 - ◆ 但し、WoSなどでは論文総数は右肩上がりに増加している点に注意
 - ◆ Nature誌によると、「2023年には撤回総数が1万件を超えた」との報告も



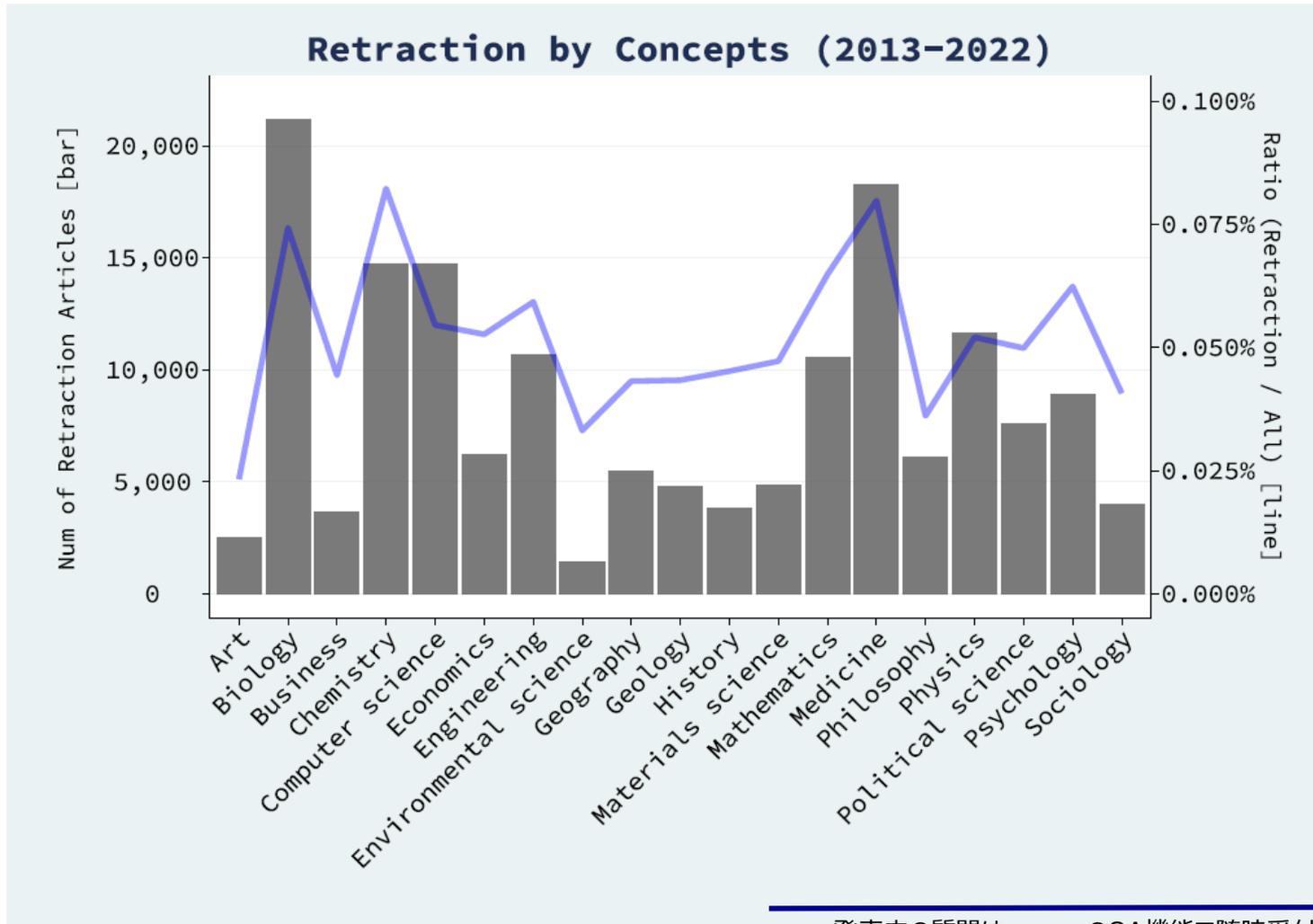
論文出版日ベース: 2022年発行の論文のうち、7,500件超が撤回… を意味する。

- 例えば、医学系と物理系は論文数は近いが、撤回数には多少の差
 - ◆ 臨床系と理論系など元々の学問の性質などに起因する可能性がある



■ 前掲の通り分野差は顕著に存在

◆ 撤回 = 不正 ではなく、かつ、分野特性に起因する可能性には留意が必要



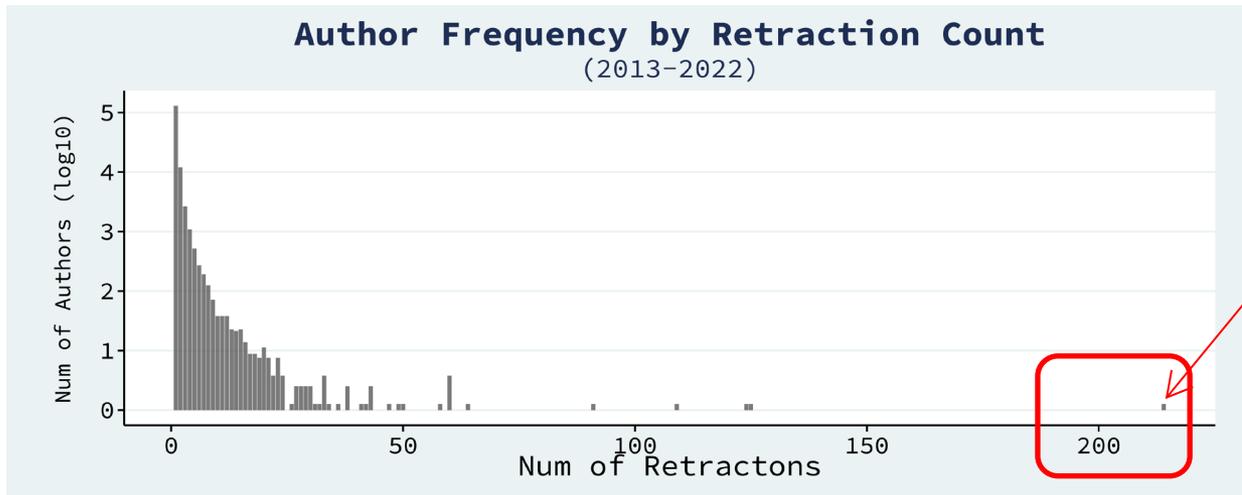
- 全著者（左表），筆頭著者（右表）ごとの撤回論文数では中国が圧倒的。日本はそれぞれ9位，6位

All Authors				
#	Region	Over All	Retract	Ratio
1	CN (中国)	7,483,331	17,885	0.24%
2	US (米国)	7,831,430	2,779	0.04%
3	IN (インド)	2,164,691	2,593	0.12%
4	SA (サウジアラビア)	214,645	1,159	0.54%
5	IR (イラン)	882,849	952	0.11%
6	PK (パキスタン)	263,132	749	0.28%
7	GB (英国)	1,819,769	728	0.04%
8	KR (韓国)	1,034,972	691	0.07%
9	JP (日本)	1,653,995	606	0.04%
10	FR (フランス)	1,239,900	543	0.04%
11	EG (エジプト)	332,469	538	0.16%
12	DE (ドイツ)	1,672,600	501	0.03%
13	IT (イタリア)	1,092,458	465	0.04%
14	MY (マレーシア)	373,023	434	0.12%
15	AU (オーストラリア)	873,062	412	0.05%
16	CA (カナダ)	1,026,400	366	0.04%
17	RU (ロシア)	1,281,893	336	0.03%
18	ET (エチオピア)	63,859	335	0.52%
19	TR (トルコ)	859,395	314	0.04%
20	ES (スペイン)	1,112,671	300	0.03%

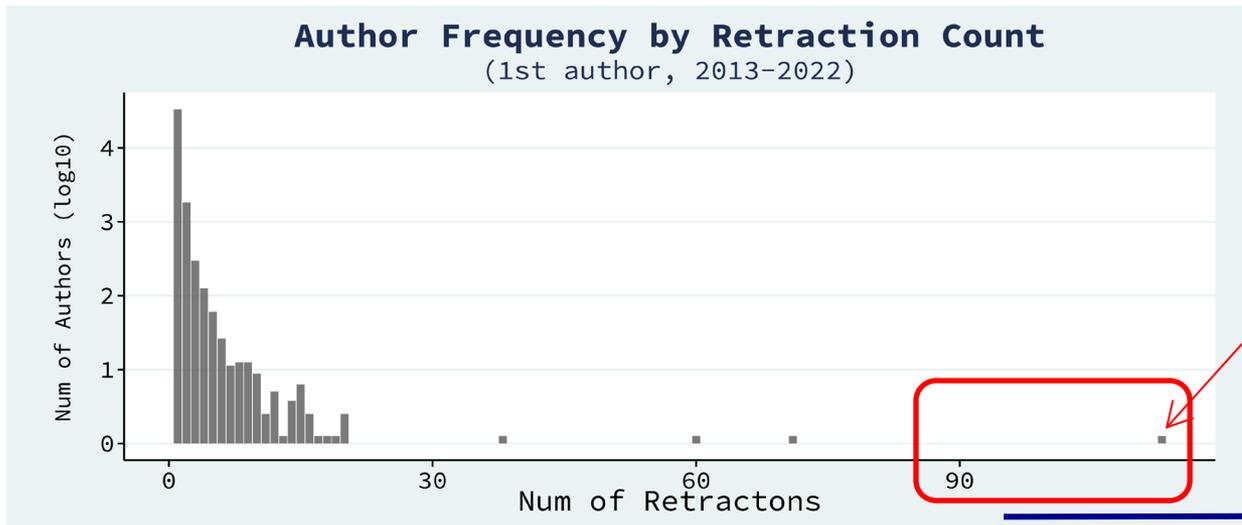
1st Author Only				
#	Region	Over All	Retract	Ratio
1	CN (中国)	5,777,015	17,019	0.29%
2	IN (インド)	1,700,378	1,973	0.12%
3	US (米国)	6,056,018	1,795	0.03%
4	IR (イラン)	688,791	777	0.11%
5	PK (パキスタン)	204,486	499	0.24%
6	JP (日本)	1,316,693	492	0.04%
7	KR (韓国)	800,371	480	0.06%
8	SA (サウジアラビア)	164,185	471	0.29%
9	FR (フランス)	972,401	395	0.04%
10	GB (英国)	1,400,644	378	0.03%
11	EG (エジプト)	262,605	349	0.13%
12	IT (イタリア)	839,281	287	0.03%
13	MY (マレーシア)	294,733	260	0.09%
14	AU (オーストラリア)	665,989	217	0.03%
15	DE (ドイツ)	1,289,552	214	0.02%
16	RU (ロシア)	1,052,802	211	0.02%
17	TR (トルコ)	695,912	199	0.03%
18	CA (カナダ)	793,267	174	0.02%
19	ES (スペイン)	874,745	172	0.02%
20	TW (台湾)	326,184	165	0.05%

■ 例えば、日本の撤回論文数が10件として…

- ◆ 10件とも同一の著者なのか、10件とも別の著者なのか…では意味が異なる



全著者だと200件以上の撤回に名前が挙がる著者が存在



第1著者のみでも100件以上の撤回に名前が挙がる著者が存在

■ 筆頭著者ごとの撤回論文数についてジニ係数を計算

◆ 直観的には、異常値 = 一部の大量撤回者が存在するほど大きくなる係数

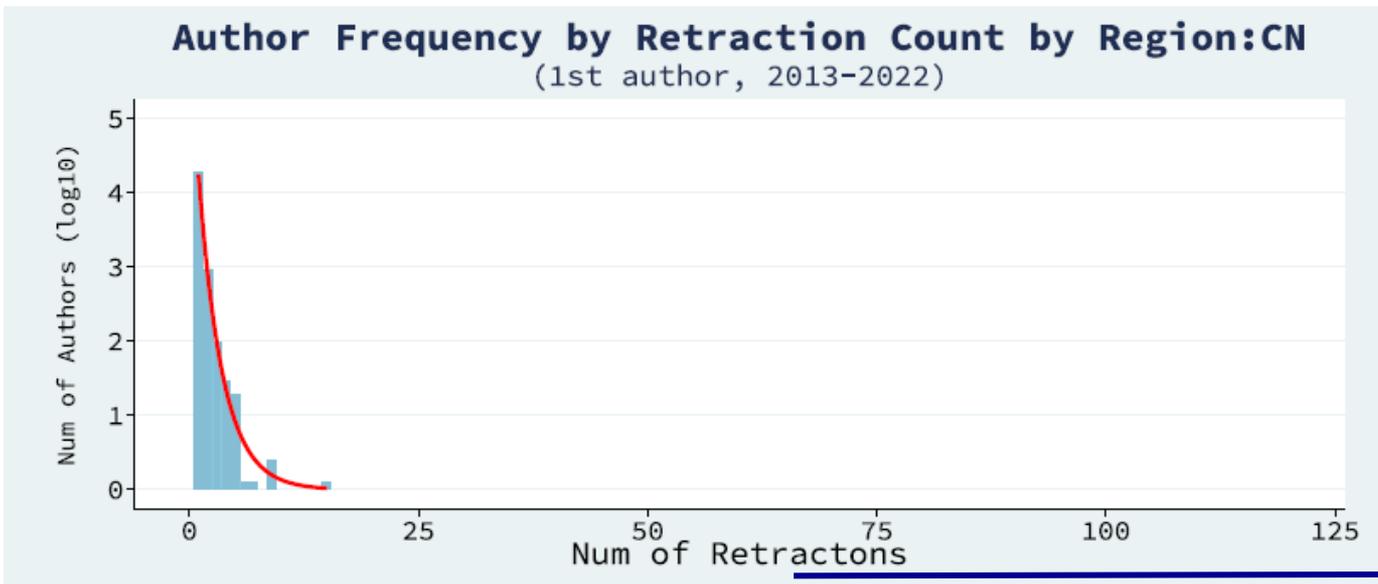
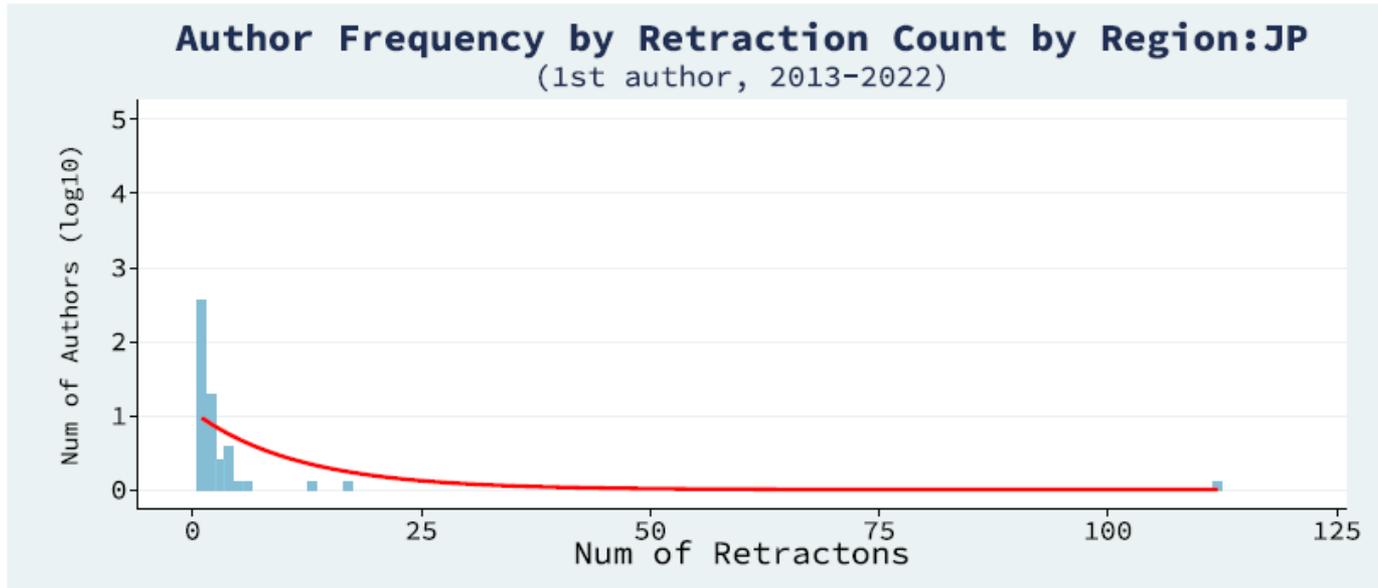


- 日本、フランス等は「少数が多数の論文撤回を行っている」可能性が高い
- 中国、英国等は「多数が少数の論文撤回を行っている」可能性が高い

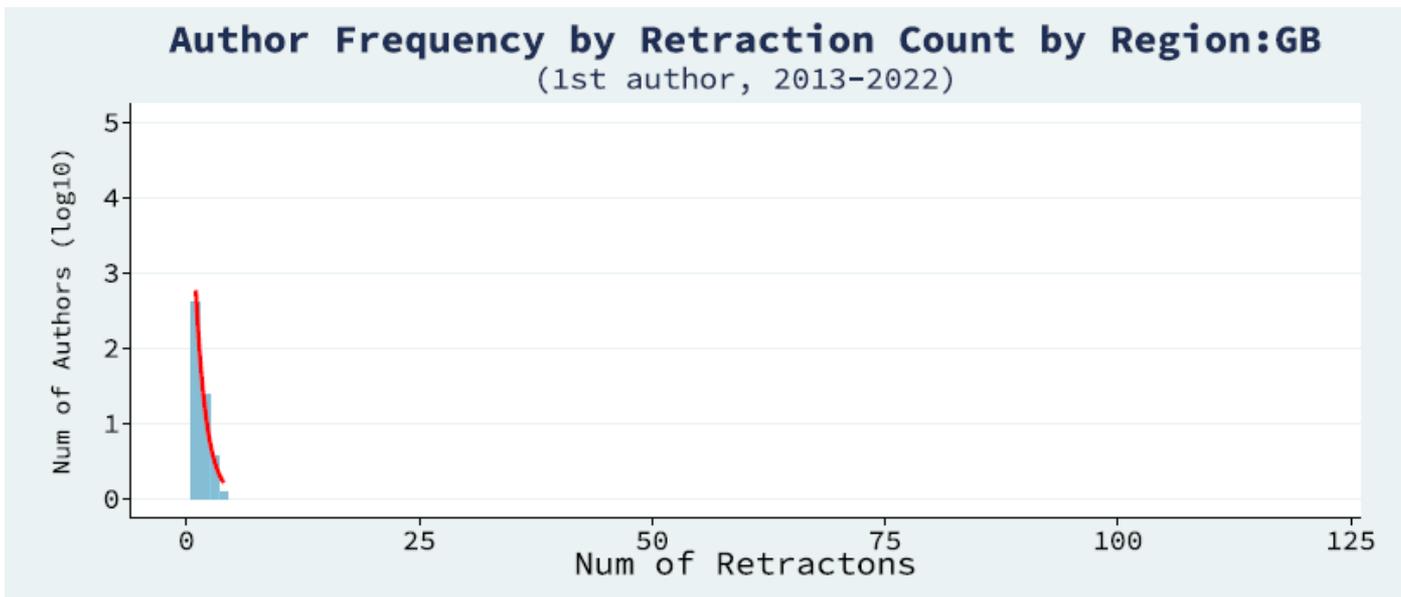
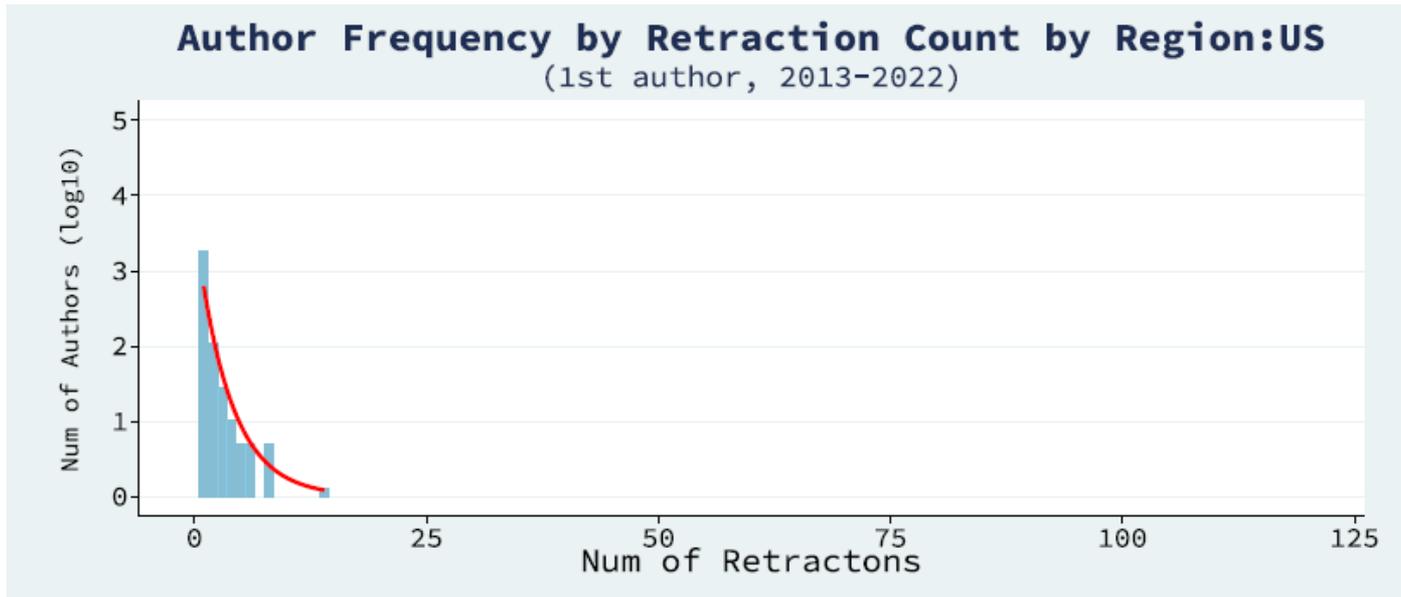
#	Region	Gini
1	CN (中国)	0.0586
2	IN (インド)	0.1097
3	US (米国)	0.1228
4	IR (イラン)	0.1881
5	PK (パキスタン)	0.1163
6	JP (日本)	0.3513
7	KR (韓国)	0.0947
8	SA (サウジアラビア)	0.1689
9	FR (フランス)	0.3337
10	GB (英国)	0.0699

#	Region	Gini
11	EG (エジプト)	0.2128
12	IT (イタリア)	0.2343
13	MY (マレーシア)	0.1106
14	AU (オーストラリア)	0.1367
15	DE (ドイツ)	0.1095
16	RU (ロシア)	0.0775
17	TR (トルコ)	0.0663
18	CA (カナダ)	0.0836
19	ES (スペイン)	0.1725
20	TW (台湾)	0.1117

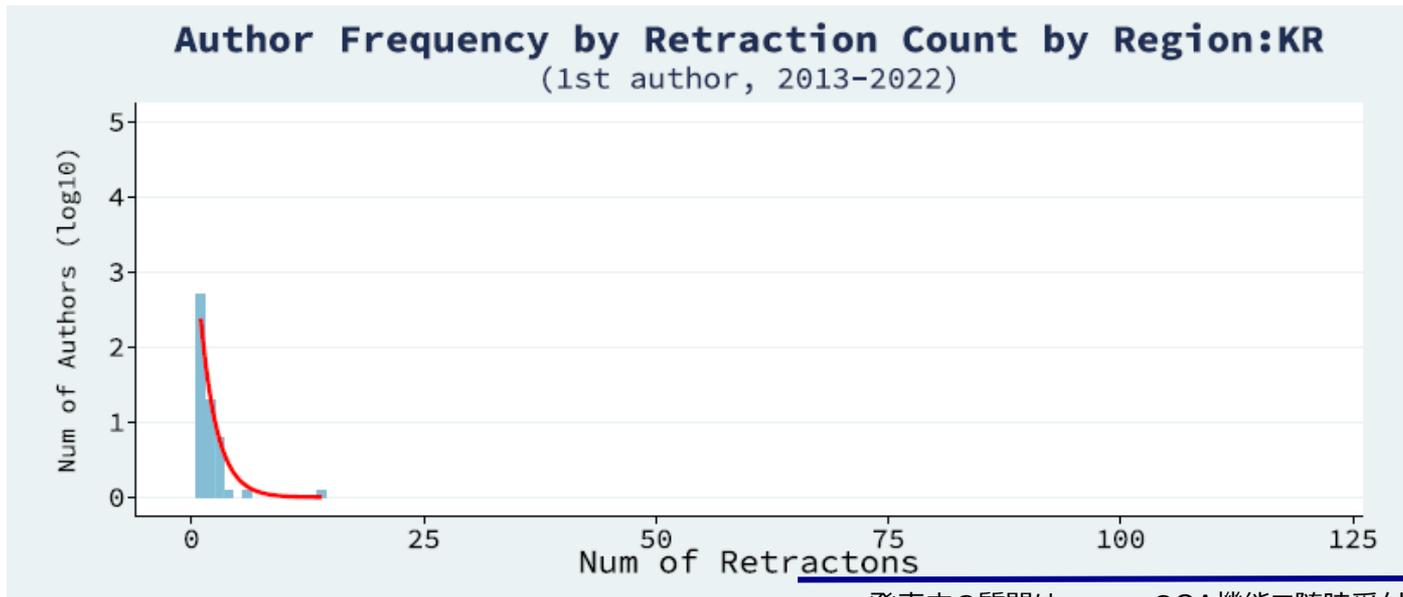
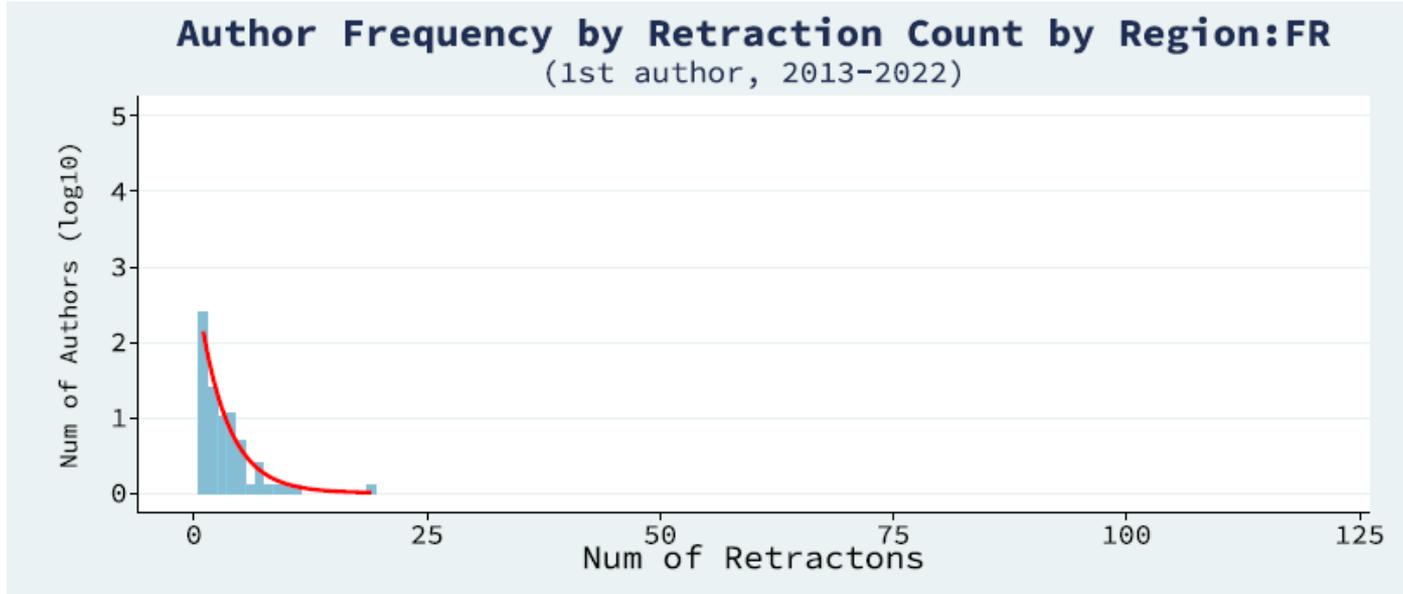
分析結果：国・地域別撤回論文数の分布（抜粋）



分析結果：国・地域別撤回論文数の分布（抜粋）



分析結果：国・地域別撤回論文数の分布（抜粋）



■ 複数付与可能だが、第三者からの指摘によるものが目立つ

◆ ペーパーミル、偽のピアレビューなど、明らかに不適切なものも存在も

#	Reason	理由（機械翻訳）	Count	%
1	Investigation by Journal/Publisher	ジャーナル／出版社による調査	16,129	49.8%
2	Unreliable Results	信頼性の低い結果	12,670	39.1%
3	Investigation by Third Party	第三者による調査	11,120	34.4%
4	Concerns/Issues About Data	データに関する懸念／問題	10,641	32.9%
5	Concerns/Issues about Referencing/Attributions	参照・帰属に関する懸念／問題	8,797	27.2%
6	Paper Mill	ペーパーミル	8,021	24.8%
7	Concerns/Issues with Peer Review	ピアレビューに関する懸念／問題	7,984	24.7%
8	Concerns/Issues About Results	結果に関する懸念／問題	6,707	20.7%
9	Randomly Generated Content	ランダム生成されたコンテンツ	6,105	18.9%
10	Fake Peer Review	偽のピアレビュー	6,068	18.7%
11	Duplication of Image	画像の重複	3,015	9.3%
12	Date of Retraction/Other Unknown	撤回日／その他不明	2,025	6.3%
13	Lack of IRB/IACUC Approval	IRB／IACUCの承認不足	1,829	5.7%
14	Rogue Editor	不正な編集者	1,764	5.4%
15	Concerns/Issues About Image	画像に関する懸念／問題	1,751	5.4%
16	Notice - Limited or No Information	お知らせ — 情報が限定的または存在しない	1,678	5.2%
17	Author Unresponsive	著者が応答しない	1,553	4.8%
18	Removed	削除された	1,544	4.8%
19	Original Data not Provided	元データが提供されていない	1,492	4.6%
20	Duplication of Article	論文の重複	1,491	4.6%
21	Plagiarism of Article	論文の盗用	1,470	4.5%
22	Euphemisms for Plagiarism	盗用に対する婉曲表現	1,363	4.2%
23	Concerns/Issues About Authorship/Affiliation	著者名／所属に関する懸念／問題	1,347	4.2%
24	Informed/Patient Consent - None/Withdrawn	インフォームド／患者同意 — 無し／撤回済み	1,171	3.6%
25	Unreliable Data	信頼性の低いデータ	1,091	3.4%

■ 出版社の偏りも存在

◆ ペーパーミルなどに関連した出版社の偏りが存在

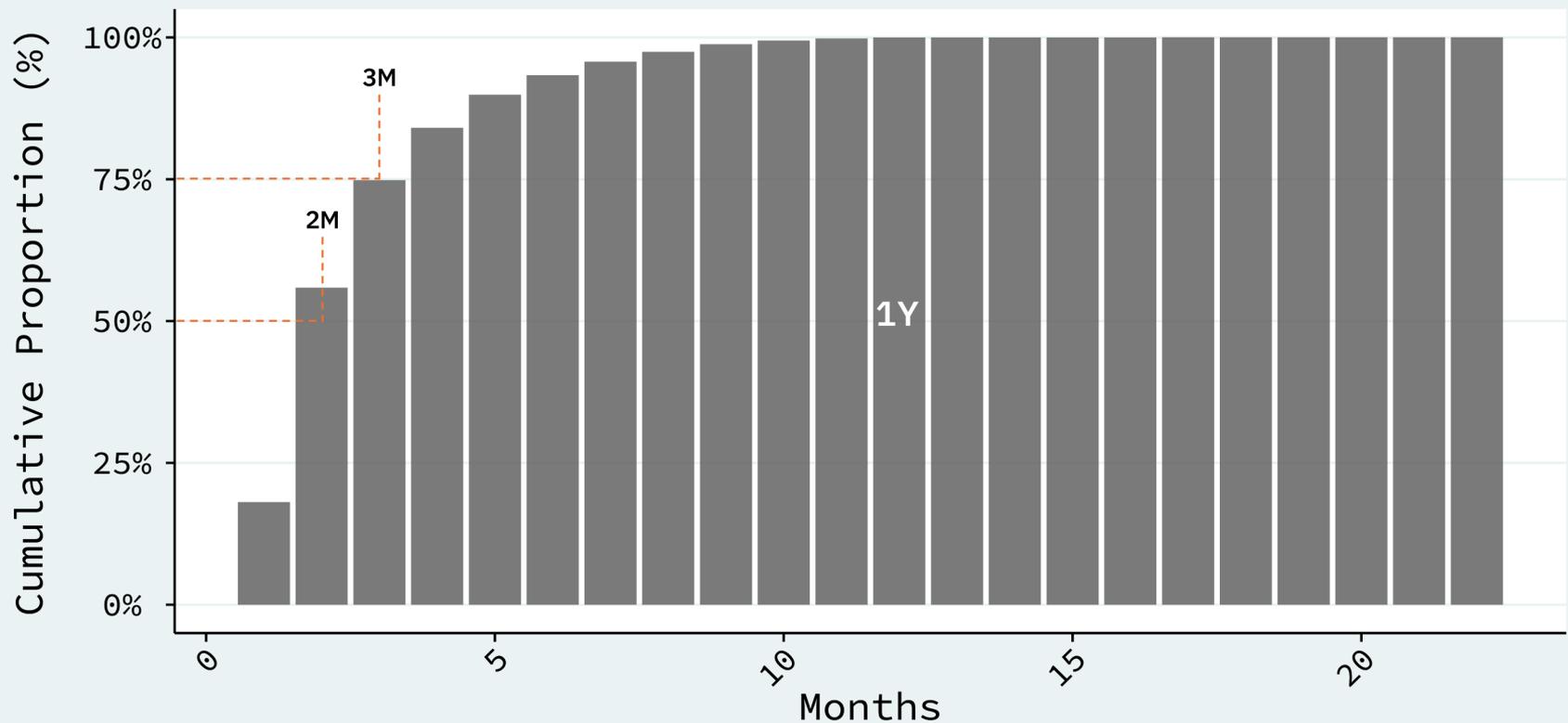
#	Publisher (Anonymization)	Over All	Retraction	% of All Retraction	Retraction /Over All
1	A	228,161	7,989	24.7%	3.5%
2	B	6,290,976	4,281	13.2%	0.1%
3	C	2,138,567	3,283	10.1%	0.2%
4	D	2,226,500	2,364	7.3%	0.1%
5	E	710,728	1,208	3.7%	0.2%
6	F	539,491	1,204	3.7%	0.2%
7	#N/A	25,814,012	1,125	3.5%	0.0%
8	G	1,217,201	1,050	3.2%	0.1%
9	H	236,308	815	2.5%	0.3%
10	I	334,353	623	1.9%	0.2%
11	J	395,351	543	1.7%	0.1%
12	K	56,101	466	1.4%	0.8%
13	L	58,611	434	1.3%	0.7%
14	M	44,168	422	1.3%	1.0%
15	N	744,352	378	1.2%	0.1%
16	O	367,303	357	1.1%	0.1%
17	P	316,247	355	1.1%	0.1%
18	Q	325,614	350	1.1%	0.1%
19	R	860,603	242	0.7%	0.0%
20	S	2,659	236	0.7%	8.9%

■ 同様にジャーナルの偏りも存在

#	Journal (Anonymization)	Over All Retraction	% of All Retraction	Retraction /Over All	
1	A	4,808	867	2.7%	18.0%
2	B	147,544	846	2.6%	0.6%
3	#N/A	13,090,461	784	2.4%	0.0%
4	C	3,486	768	2.4%	22.0%
5	D	4,193	766	2.4%	18.3%
6	E	209,906	759	2.3%	0.4%
7	F	2,538	689	2.1%	27.1%
8	G	5,599	680	2.1%	12.1%
9	H	9,154	670	2.1%	7.3%
10	I	18,941	549	1.7%	2.9%
11	J	745	442	1.4%	59.3%
12	K	4,206	431	1.3%	10.2%
13	L	10,128	416	1.3%	4.1%
14	M	3,000	371	1.1%	12.4%
15	N	1,158	367	1.1%	31.7%
16	O	18,470	339	1.0%	1.8%
17	P	3,367	295	0.9%	8.8%
18	Q	5,835	261	0.8%	4.5%
19	R	93,855	244	0.8%	0.3%
20	S	5,220	237	0.7%	4.5%

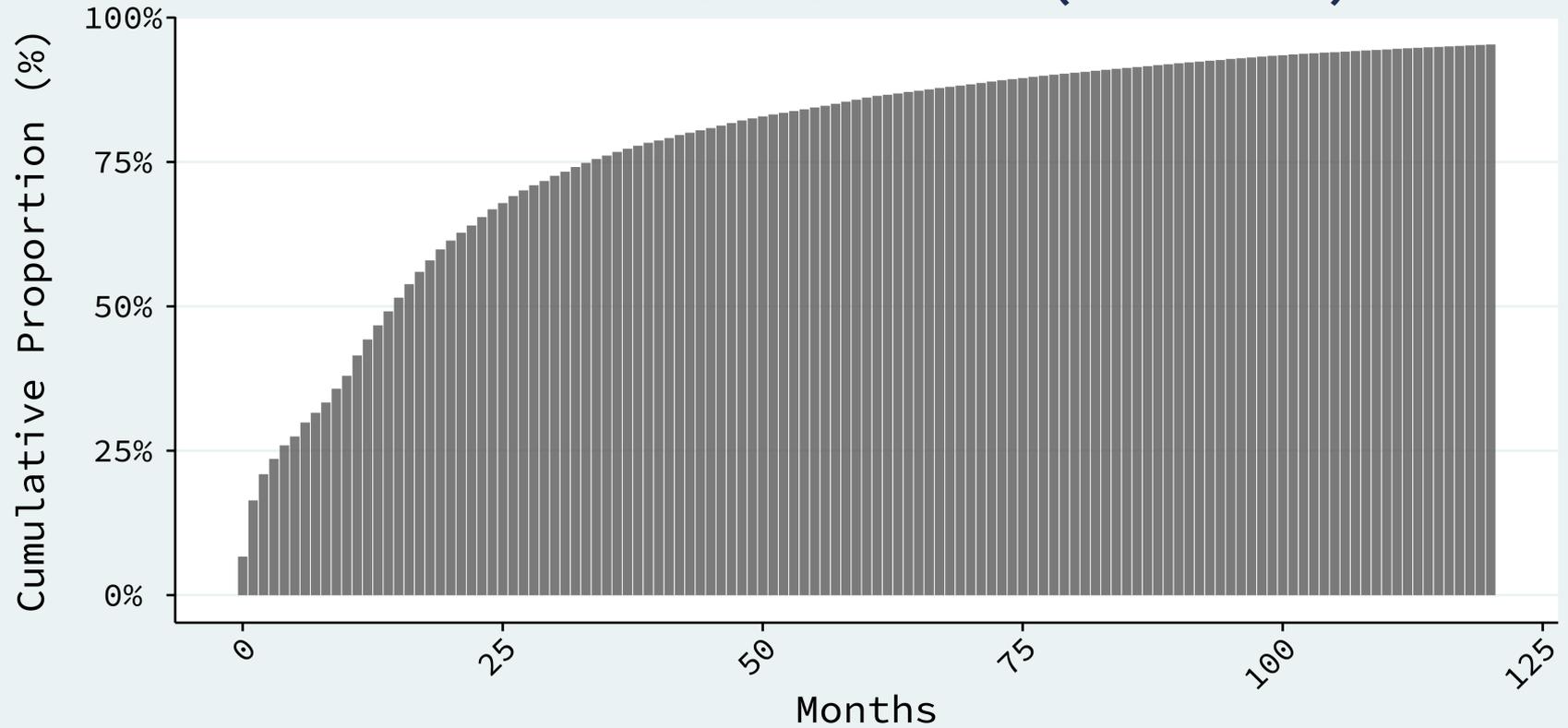
- およそ75%の論文が3か月以内に撤回されている
 - ◆ 撤回されるべき論文の大半は、比較的早くに退場させられてはいる…？

Months Until Retraction (2013-2022)



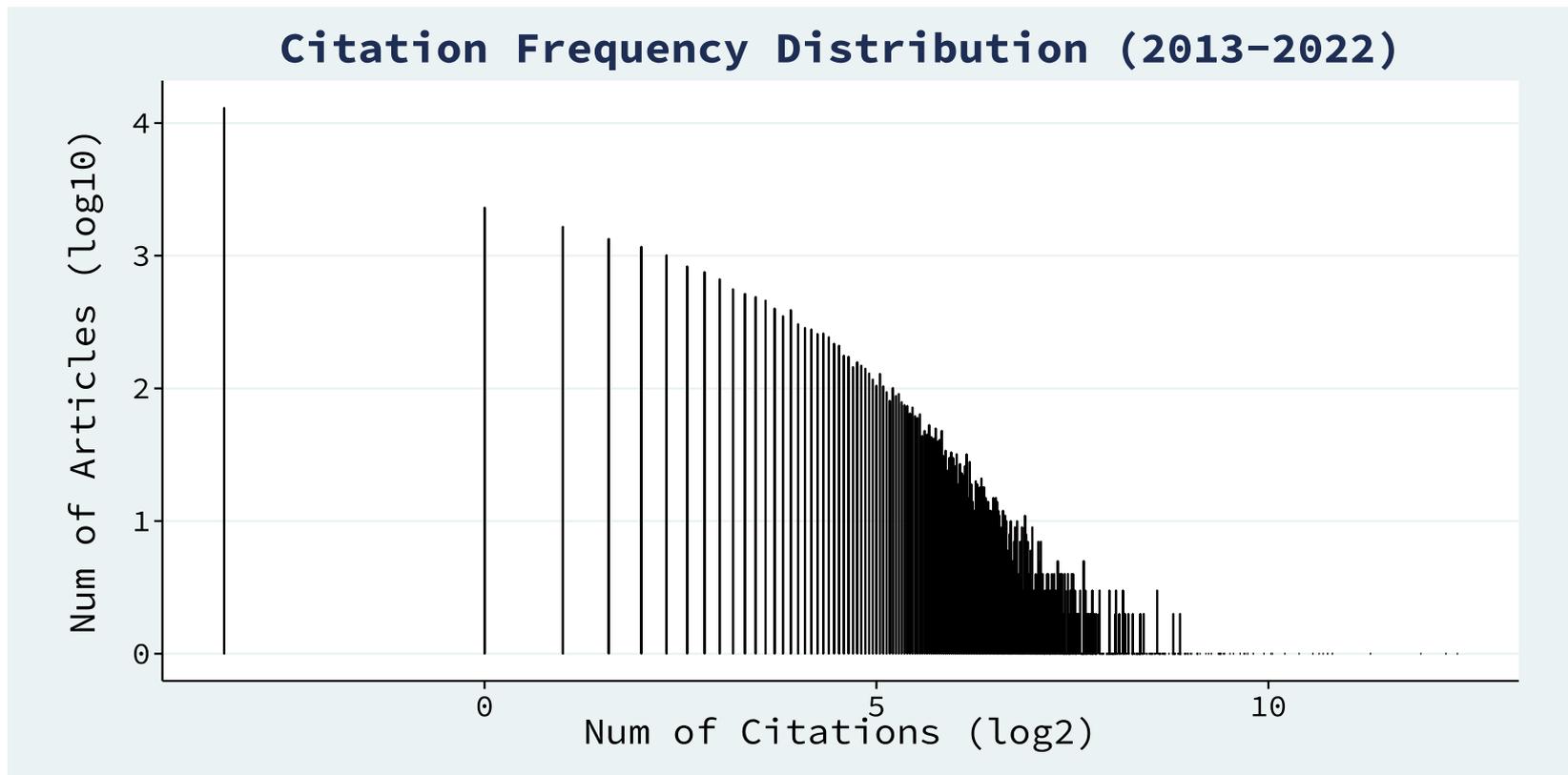
- 他方で，長期間にわたって生存するものも一定数は存在

Months Until Retraction (1940-2024)



■ 撤回されるべき論文を引用してしまう＝汚染の拡散

- ◆ 多くの論文は0回だが，中には5千件超の引用を稼いでいるもの等も存在
- ◆ 撤回されて数年たつのに，引用されているような論文のケースも存在



対数に0回は使用できないため，便宜上，0回を0.001として図示

■ 撤回論文のマクロな動態の調査を実施

- ◆ 研究成果DB OpenAlex と 撤回論文DB Retraction Watch Data とを掛け合わせて集計

■ 国・地域別の撤回論文数とジニ係数を計算

- ◆ 日本では少数の研究者が大量の論文撤回を行っている可能性が高い
- ◆ 他方, 中国等では大量の研究者が少数の撤回を行っている可能性が高い

■ 撤回論文に関係する問題の一部を簡単に紹介

- ◆ 分野ごとの撤回数
- ◆ 撤回理由
- ◆ 撤回に掛かる期間
- ◆ 撤回論文が他の研究に与える影響 (引用)

- オープンサイエンス等の新たな研究の潮流をとらえた調査研究
 - ◆ 研究論文の公開とプレプリントに関する実態調査2024（仮）
（3月末公開予定）
- データサイエンス、AI関連技術等を用いた、新たな調査研究・データ解析手法の開発
 - ◆ プレプリントによる研究トピック早期検知検証（取りまとめ中）
- オープンサイエンスが指向する科学、社会、科学と社会の変容を促す対話と共創の場の構築
 - ◆ Japan Open Science Summit（JOSS）の企画セッション等のイベント開催