

PubMed Best Match Codeathon Team 3

Vasileios Alevizos, Helen-Ann Epstein, Hacer
Karamese, Kate Majewski, Sarah Nabulsi,
Brandon Patterson, Susan Schmidt, Aidy Weeks

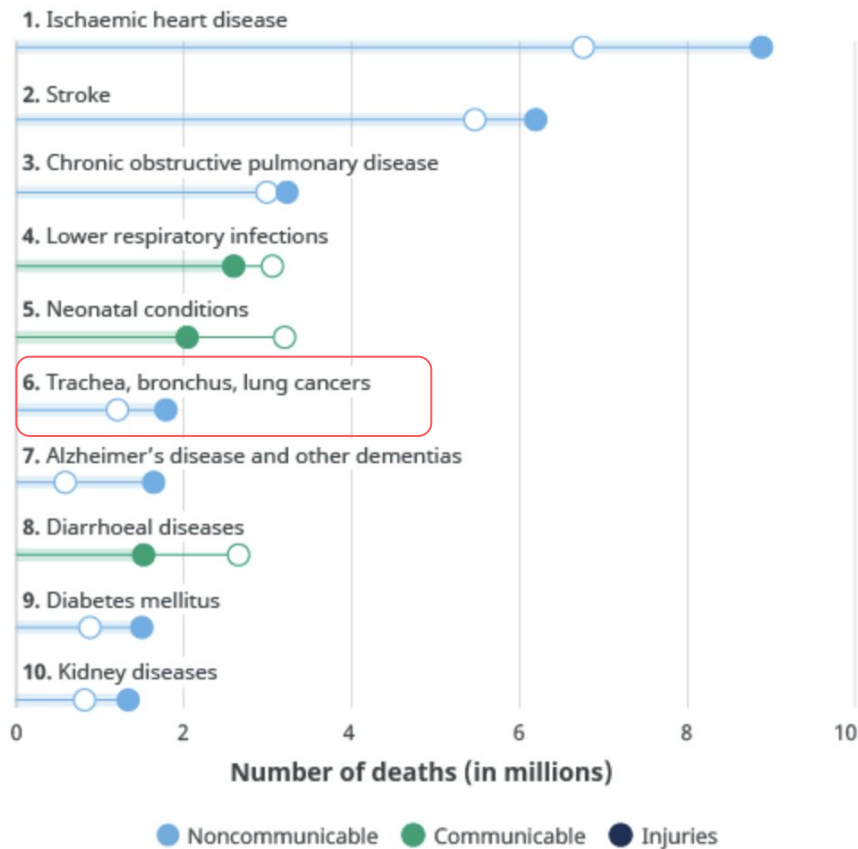
Hypothesis

Best Match increases the prominence of familiar journals by increasing the frequency of some journals over others in users' first page of results compared to using a date sort order.

Live Query

Leading causes of death globally

○ 2000 ● 2019



Source: WHO Global Health Estimates.



csv-neoplasmOR-set

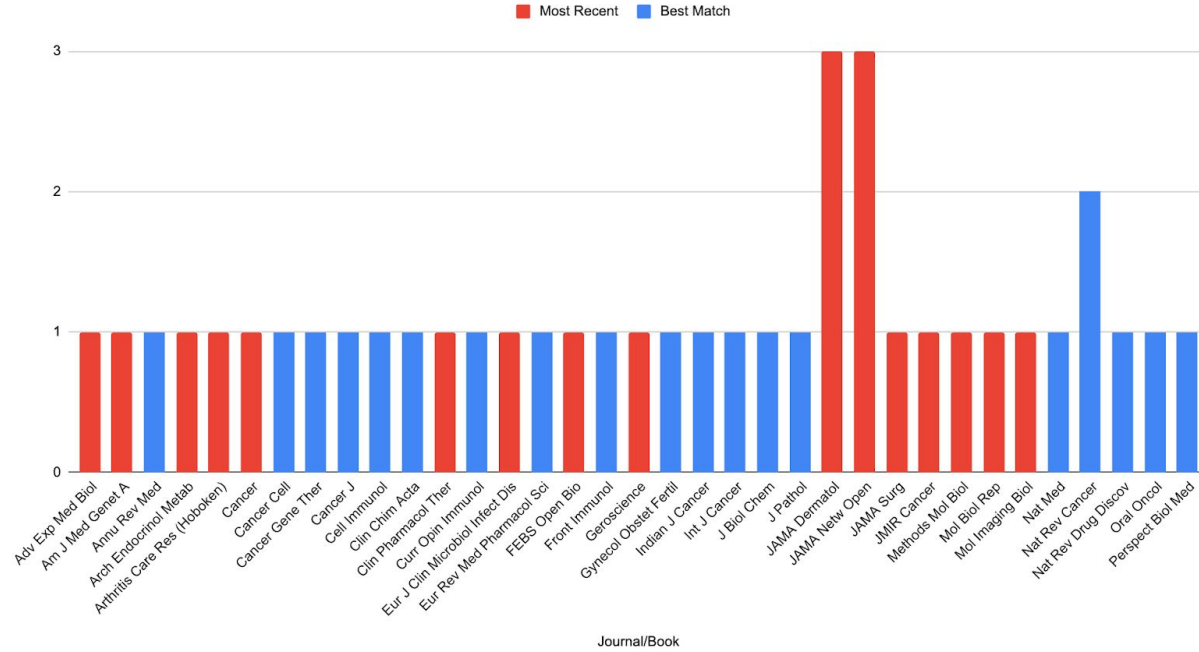


File Edit View Insert Format Data Tools Extensions Help Last edit was 5 hours ago

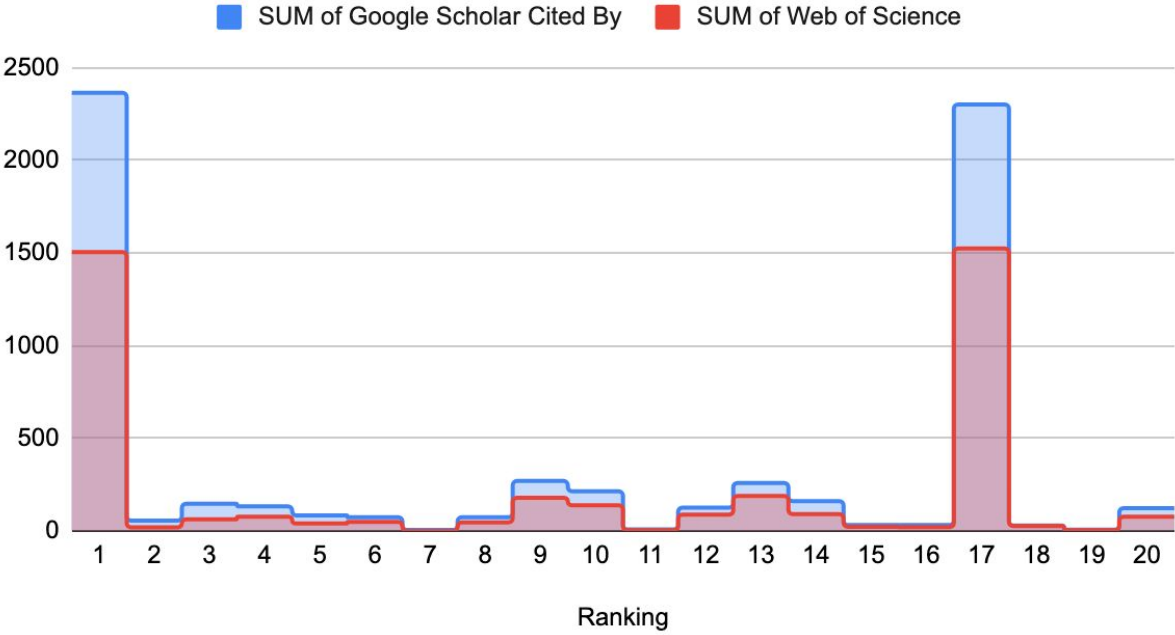
50% \$ % .0 .00 123 Default (Ari... 10 B I A

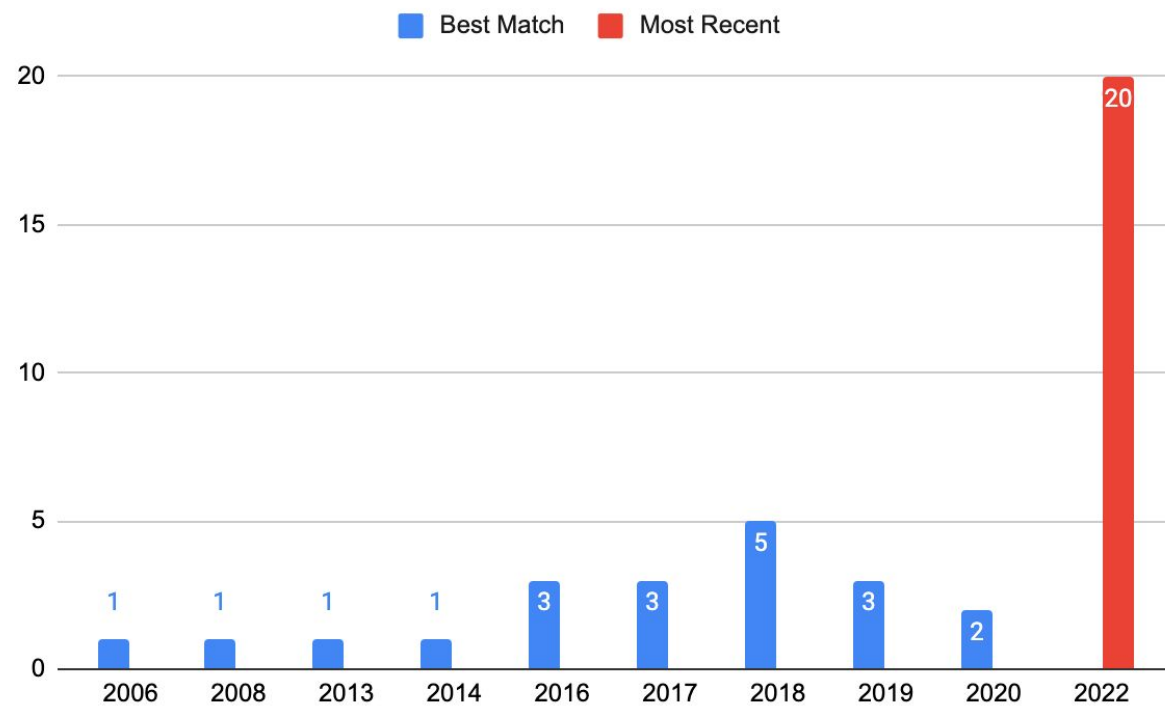
A1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	Ranking	BJR	PMID	Title	Authors	Citation	First Author	Journal/Book	Publication Year	Create Date	PMCID	NIHMS ID	DOI		
1	Best Match	18538731	Tumor cell metal	Kroemer G, Pou	Cancer Cell. 200	Kroemer G	Cancer Cell	2008	2008/06/10				10.1016/j.ccr.2008.05.005		
2	Best Match	31761807	What Is Cancer?	Hausman DM.	Perspect Biol Med	Hausman DM	Perspect Biol Med	2019	2019/11/26				10.1363/pbm.2019.0046		
3	Best Match	28244479	Cancer and cure	Roy PS, Salkia E	Indian J Cancer.	Roy PS	Indian J Cancer	2016	2017/03/01				10.4103/0019-509X.200658		
4	Best Match	29949179	Tumor microenv	Wang JJ, Lei KF	Eur Rev Med Ph	Wang JJ	Eur Rev Med Ph	2018	2018/06/28				10.26355/eurrev.201806_15270		
5	Best Match	27741350	Measuring canci	Graham TA, Sot J	Pathol. 2017 J	Graham TA	J Pathol	2017	2016/10/15				10.1002/path.4821		
6	Best Match	28574057	New methods in	Zaimy MA, Saffa	Cancer Gene Th	Zaimy MA	Cancer Gene Th	2017	2017/06/03				10.1038/ctgt.2017.16		
7	Best Match	27839715	[Lucy's cancer]	s Chene G, Lambi	Gynecol Obstet	Chene G	Gynecol Obstet	2016	2016/11/15				10.1016/j.gyobfe.2016.10.001		
8	Best Match	33024261	Cancer progress	Klein CA.	Nat Rev Cancer.	Klein CA	Nat Rev Cancer	2020	2020/10/07				10.1038/s41568-020-00300-6		
9	Best Match	24274180	Thyroid cancer	Carling T, Udelst	Annu Rev Med.	Carling T	Annu Rev Med	2014	2013/11/27				10.1146/annurev-med-061512-105739		
10	Best Match	29770138	NY-ESO-1 Base	Thomas R, Al-Ki	Front Immunol.	Thomas R	Front Immunol	2018	2018/05/18		PMC5941317		10.3389/fimmu.2018.00947		
11	Best Match	33173229	Addressing can	Mullard A.	Nat Rev Drug Di	Mullard A	Nat Rev Drug Di	2020	2020/11/11				10.1038/d41573-020-00202-0		
12	Best Match	26111503	L1CAM in huma	Altavogt P, Dobe	Int J Cancer. 20	Altavogt P	Int J Cancer	2016	2015/06/27				10.1002/ijc.29658		
13	Best Match	29217838	An evolutionary	Curtius K, Wrigh	Nat Rev Cancer.	Curtius K	Nat Rev Cancer	2018	2017/12/09				10.1038/nrc.2017.102		
14	Best Match	29397066	Cancer-associ	Liao Z, Tan ZW,	Cell Immunol. 2	Liao Z	Cell Immunol	2019	2018/02/05				10.1016/j.cellimm.2017.12.003		
15	Best Match	30480571	Adolescent and	Tricoli JV, Bleyer	Cancer J. 2018	Tricoli JV	Cancer J	2018	2018/11/28				10.1097/PPQ.0000000000000343		
16	Best Match	28351568	Is there an incre	Minghani H, Stur	Oral Oncol. 201	Minghani H	Oral Oncol	2017	2017/03/30				10.1016/j.oraloncology.2017.02.024		
17	Best Match	16892035	Tumor metastas	Steege PS.	Nat Med. 2006	Steege PS	Nat Med	2006	2006/08/08				10.1038/nm1469		
18	Best Match	30593508	Bladder cancer	Wu CH, Silvers	J Biol Chem. 20	Wu CH	J Biol Chem	2019	2018/12/30		PMC6398136		10.1074/jbc.RA118.006682		
19	Best Match	29032295	Innate sensing	Seeliger R, Searl	Curr Opin Immu	Seeliger R	Curr Opin Immu	2018	2017/10/17		PMC5857403	NIHMS912844	10.1016/j.coli.2017.09.005		
20	Best Match	23266771	ACTB in cancer	Guo C, Liu S, W	Clin Chim Acta.	Guo C	Clin Chim Acta	2013	2012/12/26				10.1016/j.cca.2012.12.012		
21	Most Recent	35612878	Identifying Ther	Kessler SR, Henr	JMIR Cancer. 2	Kessler SR	JMIR Cancer	2022	2022/05/25				10.2196/34828		
22	Most Recent	35612872	Risk of Prostate	Wheeler AM, Ro	Arthritis Care	Wheeler AM	Arthritis Care	2022	2022/05/25				10.1002/acr.24890		
23	Most Recent	35612856	Development an	Deng H, Ellekha	JAMA Netw Ope	Deng H	JAMA Netw Ope	2022	2022/05/25				10.1001/jamanetworkopen.2022.12930		
24	Most Recent	35612854	Test Performanc	Teoh D, Nam G,	JAMA Netw Ope	Teoh D	JAMA Netw Ope	2022	2022/05/25				10.1001/jamanetworkopen.2022.14020		
25	Most Recent	35612853	Association Bet	Popat S, Liu SV,	JAMA Netw Ope	Popat S	JAMA Netw Ope	2022	2022/05/25				10.1001/jamanetworkopen.2022.14046		
26	Most Recent	35612849	Non-melanoma	S Maisel-Campbe	JAMA Dermatol.	Maisel-Campbe	JAMA Dermatol	2022	2022/05/25				10.1001/jamadermatol.2022.1733		
27	Most Recent	35612841	Non-alpine primi	Bala NM, Simde	Arch Endocrinol	Bala NM	Arch Endocrinol	2022	2022/05/25				10.20945/2359-3997000000460		
28	Most Recent	35612839	Risk of Hematol	Deng J, Parthas	JAMA Dermatol.	Deng J	JAMA Dermatol	2022	2022/05/25				10.1001/jamadermatol.2022.1562		
29	Most Recent	35612832	Association of M	Shoucair S, Che	JAMA Surg. 202	Shoucair S	JAMA Surg	2022	2022/05/25				10.1001/jamasurg.2022.1362		
30	Most Recent	35612829	Association of B	Asdourian MS, S	JAMA Dermatol.	Asdourian MS	JAMA Dermatol	2022	2022/05/25				10.1001/jamadermatol.2022.1624		
31	Most Recent	35612824	Tuberous sclero	Evans LM, Geer	Am J Med Gene	Evans LM	Am J Med Gene	2022	2022/05/25				10.1002/ajmg.a.62850		
32	Most Recent	35612815	Additional cytogi	Loke J, Labopin	Cancer. 2022	Loke J	Cancer	2022	2022/05/25				10.1002/oncr.34268		
33	Most Recent	35612791	Microbiota in Pe	Fong SB, Boyer	Adv Exp Med Bi	Fong SB	Adv Exp Med Bi	2022	2022/05/25				10.1007/978-3-030-96881-6_2		
34	Most Recent	35612789	G-CSF/GM-CSF	He K, Liu X, Hof	FEBES Open Bi	He K	FEBES Open Bi	2022	2022/05/25				10.1002/2211-5463.13445		
35	Most Recent	35612777	pIRNAs and PIV	AmelMojarad M	Mol Biol Rep. 20	AmelMojarad M	Mol Biol Rep	2022	2022/05/25				10.1007/s11033-022-07506-x		
36	Most Recent	35612774	Association of	Bagi Z, Kroenke	Geroscience. 20	Bagi Z	Geroscience	2022	2022/05/25				10.1007/s11357-022-00585-5		
37	Most Recent	35612771	Early Diagnosis	Zhu T, Jiang J, X	Mol Imaging Biol	Zhu T	Mol Imaging Biol	2022	2022/05/25				10.1007/s11037-022-01743-z		
38	Most Recent	35612766	Use of DosR ani	Mao L, Xu L, Wz	Eur J Clin Micro	Mao L	Eur J Clin Micro	2022	2022/05/25				10.1007/s10096-022-04459-8		
39	Most Recent	35612761	Clinical relevanc	Taskar KS, Yang	Clin Pharmacol	Taskar KS	Clin Pharmacol	2022	2022/05/25				10.1002/cpt.2670		
40	Most Recent	35612745	Comprehensive	Agbo L, Blanche	Methods Mol Biol	Agbo L	Methods Mol Biol	2022	2022/05/25				10.1007/978-1-0716-2124-0_15		

Best Match and Most Recent



Best Match Ranking: GS Cited By vs. WoS Citations



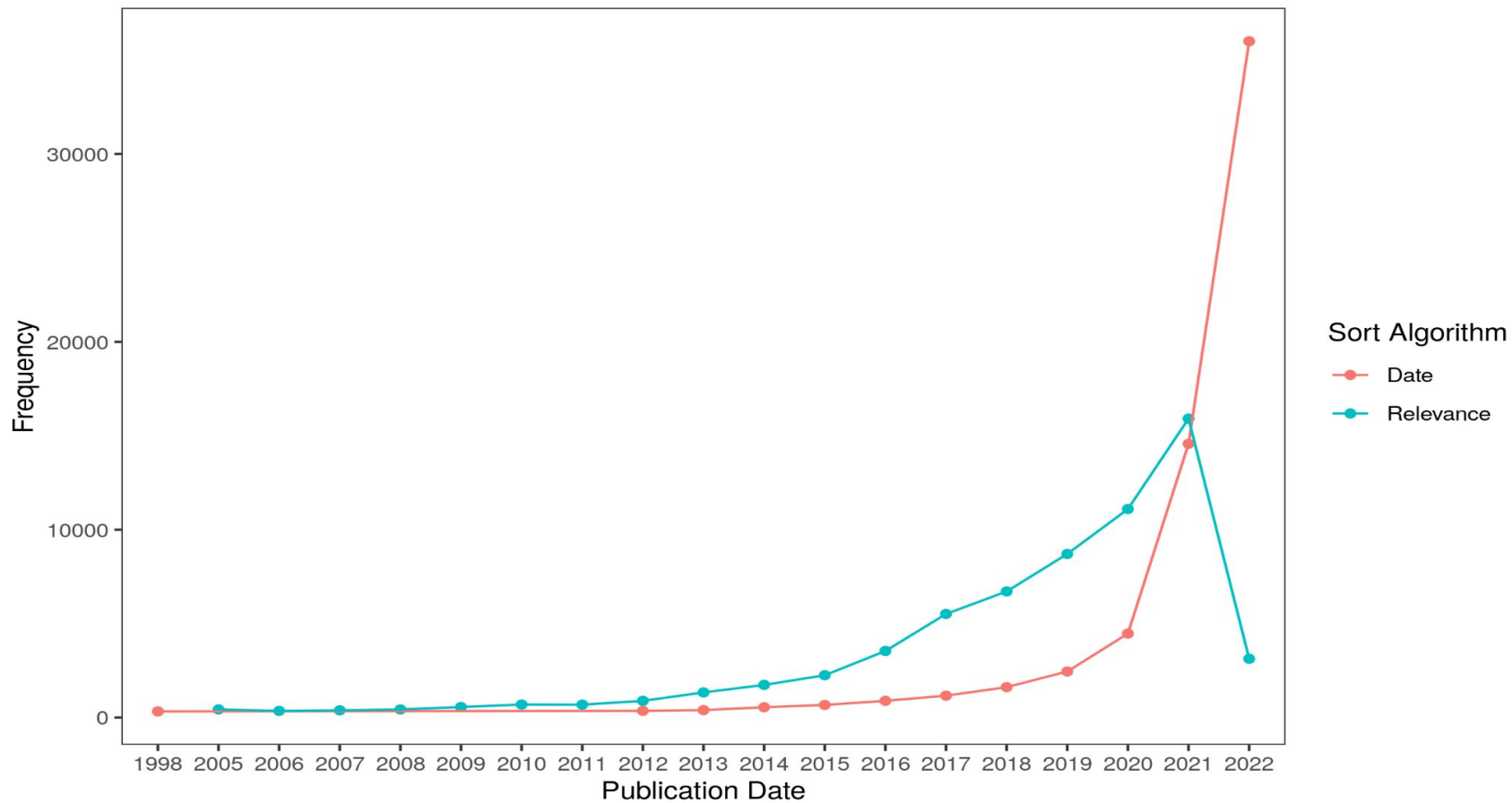


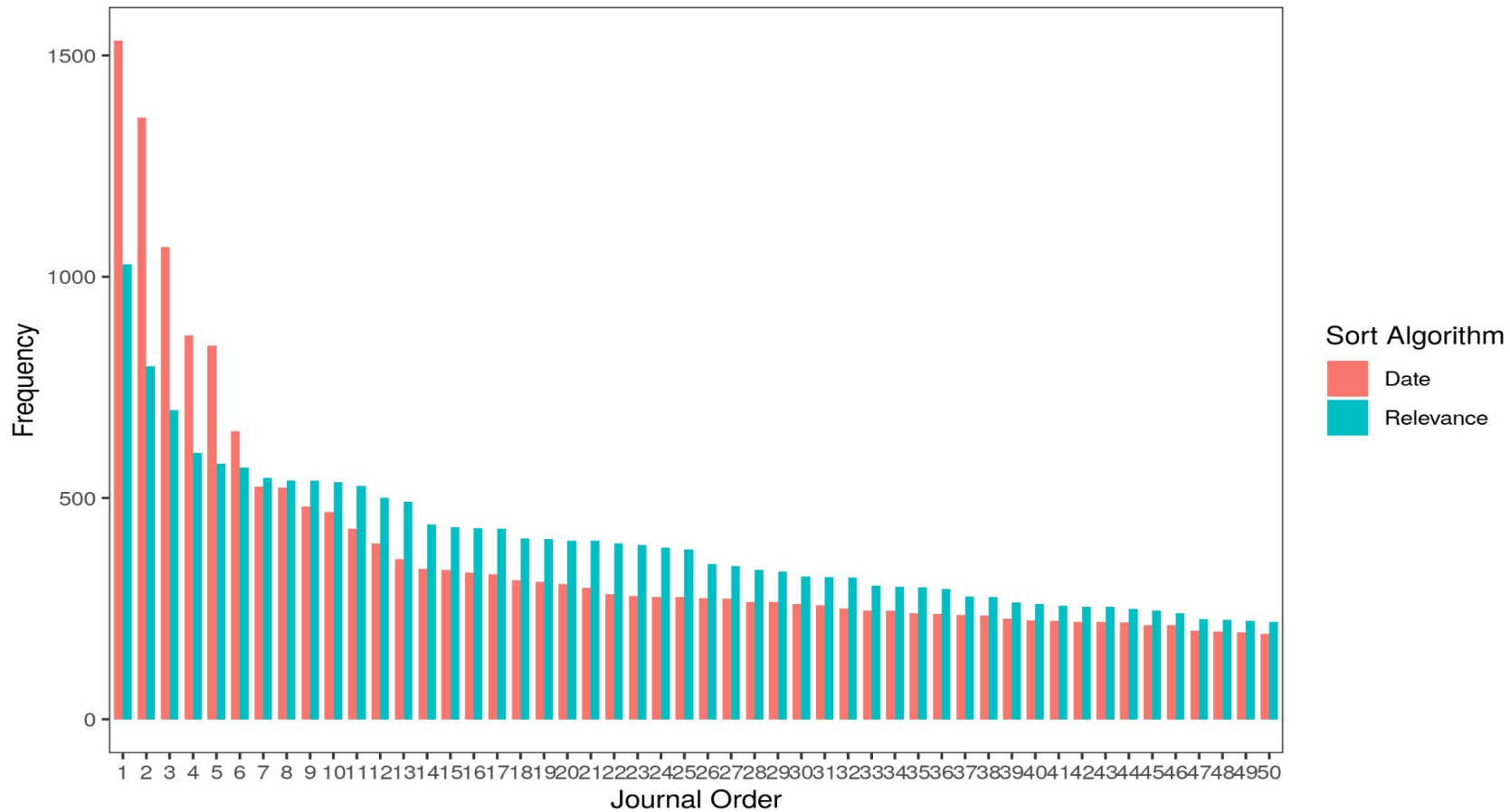
Another consideration

In order to reduce barriers for other researchers and scholars who are not able to access the training dataset and would still like to do testing in the live PubMed environment, we recommend having additional fields exportable when downloading .csv files:

- Author affiliation
- Language
- Publication Type
- MeSH Terms

Applied to Training Dataset





Outcomes

- Significant different in frequency distribution between Best Match and Date sort order
- Best Match seems to favor a more distributed set of journals in the first page of results
- Best Match has a larger distribution of dates with older articles occasionally coming up our broader search topic
- We can't confirm our hypothesis which was that the distributions should be similar for Best Match and Date sort for journal title.

Future Work

Determine if the skewed distribution represents:

- A trend in publishing, where fewer journals are published for this content area currently than in the past.
- An artifact of issue frequency: A subset of titles issuing more citations than others in recent years.
- Click through rate satisfaction by combining good reputable publishers and traffic engagement as percentage of impression.

What else?

- Encourage more participatory/broader research
 - Consider funding/providing more opportunities for more research on the behavioral side of information seeking, to help inform system development
- Improve the algorithm
 - Concise Title, broad MeSH terms, few MeSH terms, lack of abstract may over-rank Old MEDLINE records
 - Do special characters in author names or non-english articles have an effect?
 - Crowdsourcing (catching things like older articles in best match)
 - Create a mechanism to report when older non-relevant records show up in the first 10 results for Best Match