Supplementary Material for "Are disruptive papers more likely to impact technology and society?"

Alex J. Yang, Xiaohui Yan, Haotian Hu, Hanlin Hu, Jia Kong, Sanhong Deng

School of Information Management, Nanjing University, Nanjing, 210023, China.

Correspondence

Alex J. Yang, School of Information Management, Nanjing University, Nanjing 210023, China.

Email: alexjieyang@outlook.com (A.J.Y.)

Table S1. Description of the variables

Type	Variables	Description/Definition	Sources/References
Dependent variable	**P (Patent)**	The probability of papers cited by the patented inventions	USTPO, Marx and Fuegi (2022)
	P (Clinical trials)	The probability of papers integrated into clinical trials	ClinicalTrials.gov, SciSciNet (Lin et al., 2023)
	P (Newsfeed)	The probability of papers mentioned by news sources	Crossref (Hendricks et al., 2020)
	P (Tweet)	The probability of papers mentioned by Tweets	Crossref (Hendricks et al., 2020)
Independent variable	**Disruption index**	The proportion of disruptive citations in the total citations received by the paper.	(Funk & Owen-Smith, 2017; Park et al., 2023; Wu et al., 2019)
	Disruptive citation	The number of citations that reference the focal paper without citing its original sources.	(Lin et al., 2022; Yang et al., 2023)
Team level controls	**Team size**	The number of authors in a paper	(Wuchty et al., 2007)
	International team	Indicates whether the paper's authorship includes individuals from different countries.	(Lee et al., 2019)
	Interdisciplinary team	Indicates whether the paper's authorship spans multiple fields of study.	(Liu et al., 2024)
Paper level controls	**Funding**	Indicates whether the research was funded by the National Institutes of Health (NIH) or the National Science Foundation (NSF).	(Yang, 2024; Zhai et al., 2024)
	#Refercence	The total number of references cited in the paper.	(Yang et al., 2024)
	Rao-Stirling	A metric measuring the level of interdisciplinarity in a paper, typically calculated using methods developed by Rao and Stirling.	(Gates et al., 2019; Stirling, 2007)

Table S2. Effect of the 10-year disruption index on the likelihood of papers to be linked to technology.

Models	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Dependent Variables	**P (Patents)**				**P (Clinical trials)**			
10-year Disruption (Pct.)	-0.2358***	0.0967***	0.0978***	0.3146***	-0.0450***	0.3089***	0.2943***	0.6075***
	(0.0020)	(0.0023)	(0.0023)	(0.0025)	(0.0073)	(0.0080)	(0.0080)	(0.0093)
ln (Team size)			0.2270***	0.2092***			0.5115***	0.4254***
			(0.0013)	(0.0013)			(0.0033)	(0.0034)
International team			-0.0459***	-0.0691***			0.0955***	0.0493***
			(0.0019)	(0.0020)			(0.0052)	(0.0054)
Interdisciplinary team			0.1164***	0.1118***			-0.1097***	-0.1581***
			(0.0015)	(0.0015)			(0.0047)	(0.0047)
Funding				0.2766***				0.3980***
				(0.0021)				(0.0053)
ln (#Reference)				0.4396***				0.6142***
				(0.0012)				(0.0033)
Rao-Stirling				-1.802***				-1.065***
				(0.0104)				(0.0355)
Year FE	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Field FE	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**

Journal FE	No	**Yes**	**Yes**	**Yes**	No	**Yes**	**Yes**	**Yes**
Obs.	35,729,603	35,014,987	35,014,987	31,600,461	35,729,603	18,695,874	18,695,874	17,236,594
***Pseudo* R^2**	0.13961	0.23929	0.24173	0.24271	0.16735	0.16204	0.1715	0.18506

Note: robust standard errors are reported in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The absence of observations in certain models is attributed to instances where only one value is available for a particular year, field, and journal combination, often indicating a lack of associated papers within that technological/societal domain. Furthermore, missing values in the 'Rao-Stirling' variable contribute to the observed absence of data points.

Table S3. Effect of the 10-year disruption index on the likelihood of papers to be linked to society.

Models	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Dependent Variables	**P (News)**				**P (Tweets)**			
10-year Disruption (Pct.)	-0.1696*** (0.0087)	0.4434*** (0.0098)	0.4588*** (0.0098)	0.7656*** (0.0117)	-0.5336*** (0.0030)	0.0565*** (0.0036)	0.0649*** (0.0036)	0.3542*** (0.0044)
ln (Team size)			0.1499*** (0.0040)	0.0797*** (0.0043)			0.0205*** (0.0016)	-0.0209*** (0.0017)
International team			0.0880*** (0.0054)	0.0715*** (0.0057)			0.1628*** (0.0020)	0.1490*** (0.0022)
Interdisciplinary team			0.1230*** (0.0053)	0.1132*** (0.0055)			0.0901*** (0.0019)	0.0843*** (0.0021)
Funding				0.2126*** (0.0068)				0.1204*** (0.0030)
ln (#Reference)				0.3995*** (0.0045)				0.3576*** (0.0017)
Rao-Stirling				0.1660** (0.0566)				-0.0352 (0.0238)
Year FE	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Field FE	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Journal FE	No	**Yes**	**Yes**	**Yes**	No	**Yes**	**Yes**	**Yes**
Obs.	35,729,603	31,178,015	31,178,015	28,059,313	35,729,603	34,405,397	34,405,397	31,048,614
***Pseudo* R^2**	0.09508	0.22304	0.2247	0.23366	0.28393	0.41814	0.4188	0.43041

Note: robust standard errors are reported in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The absence of observations in certain models is attributed to instances where only one value is available for a particular year, field, and journal combination, often indicating a lack of associated papers within that technological/societal domain. Furthermore, missing values in the 'Rao-Stirling' variable contribute to the observed absence of data points.

Table S4. Effect of the 10-year disruptive citation on the likelihood of papers to be linked to technology.

Models	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Dependent Variables	**P (Patents)**				**P (Clinical trials)**			
ln (10-year Disruptive citation)	0.7271*** (0.0006)	0.6300*** (0.0007)	0.6222*** (0.0007)	0.6107*** (0.0008)	1.052*** (0.0018)	0.9742*** (0.0022)	0.9556*** (0.0023)	0.9235*** (0.0024)
ln (Team size)			0.1258*** (0.0012)	0.1190*** (0.0013)			0.2585*** (0.0034)	0.2485*** (0.0035)
International team			-0.0832*** (0.0019)	-0.0932*** (0.0020)			-0.0083 (0.0054)	-0.0148** (0.0056)
Interdisciplinary team			0.1132*** (0.0015)	0.1146*** (0.0015)			-0.0957*** (0.0047)	-0.1176*** (0.0048)
Funding				0.2315*** (0.0021)				0.3113*** (0.0054)
ln (#Reference)				0.1887*** (0.0012)				0.1502*** (0.0034)
Rao-Stirling				-1.071*** (0.0102)				0.0064 (0.0354)
Year FE	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Field FE	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Journal FE	No	**Yes**	**Yes**	**Yes**	No	**Yes**	**Yes**	**Yes**
Obs.	35,729,603	35,014,987	35,014,987	31,600,461	35,729,603	18,695,874	18,695,874	17,236,594
***Pseudo* R^2**	0.19941	0.27268	0.27366	0.27146	0.27627	0.24159	0.24355	0.24345

Note: robust standard errors are reported in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The absence of observations in certain models is attributed to instances where only one value is available for a particular year, field, and journal combination, often indicating a lack of associated papers within that technological/societal domain. Furthermore, missing values in the 'Rao-Stirling' variable contribute to the observed absence of data points.

Table S5. Effect of the 10-year disruptive citation on the likelihood of papers to be linked to society.

Models	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variables	P (News)				P (Tweets)			
ln (10-year Disruptive citation)	0.9422*** (0.0021)	0.7226*** (0.0026)	0.7233*** (0.0026)	0.7273*** (0.0029)	0.4816*** (0.0008)	0.4100*** (0.0010)	0.4089*** (0.0010)	0.4084*** (0.0011)
ln (Team size)			-0.0533*** (0.0041)	-0.0650*** (0.0043)			-0.0564*** (0.0016)	-0.0815*** (0.0017)
International team			0.0479*** (0.0055)	0.0470*** (0.0057)			0.1438*** (0.0020)	0.1356*** (0.0022)
Interdisciplinary team			0.1225*** (0.0053)	0.1194*** (0.0056)			0.0842*** (0.0019)	0.0836*** (0.0021)
Funding				0.1502*** (0.0069)				0.0857*** (0.0030)
ln (#Reference)				0.0691*** (0.0043)				0.1935*** (0.0017)
Rao-Stirling				0.9516*** (0.0577)				0.3210*** (0.0237)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Journal FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Obs.	35,729,603	31,178,015	31,178,015	28,059,313	35,729,603	34,405,397	34,405,397	31,048,614
Pseudo R^2	0.17883	0.26331	0.26355	0.26872	0.30942	0.43343	0.4338	0.44379

Note: robust standard errors are reported in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The absence of observations in certain models is attributed to instances where only one value is available for a particular year, field, and journal combination, often indicating a lack of associated papers within that technological/societal domain. Furthermore, missing values in the 'Rao-Stirling' variable contribute to the observed absence of data points.

Table S6. Effect of the 5-year disruption index on the likelihood of papers to be linked to technology.

Models	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variables	P (Patents)				P (Clinical trials)			
5-year Disruption (Pct.)	-0.3561*** (0.0019)	-0.0170*** (0.0022)	0.0009 (0.0022)	0.2069*** (0.0025)	-0.6990*** (0.0074)	-0.2954*** (0.0078)	-0.1989*** (0.0079)	0.0662*** (0.0090)
ln (Team size)			0.2334*** (0.0012)	0.2161*** (0.0013)			0.5199*** (0.0033)	0.4360*** (0.0034)
International team			-0.0451*** (0.0019)	-0.0684*** (0.0019)			0.0877*** (0.0052)	0.0422*** (0.0054)
Interdisciplinary team			0.1154*** (0.0014)	0.1103*** (0.0015)			-0.1075*** (0.0047)	-0.1583*** (0.0047)
Funding				0.2792*** (0.0020)				0.3983*** (0.0053)
ln (#Reference)				0.4366*** (0.0012)				0.5912*** (0.0032)
Rao-Stirling				-1.729*** (0.0101)				-0.6644*** (0.0347)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Journal FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Obs.	38,104,430	37,356,439	37,356,439	33,335,435	38,104,430	19,533,842	19,533,842	17,899,089
Pseudo R^2	0.1358	0.23739	0.23991	0.241	0.16924	0.1617	0.17139	0.184

Note: robust standard errors are reported in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The absence of observations in certain models is attributed to instances where only one value is available for a particular year, field, and journal combination, often indicating a lack of associated papers within that technological/societal domain. Furthermore, missing values in the 'Rao-Stirling' variable contribute to the observed absence of data points.

Table S7. Effect of the 5-year disruption index on the likelihood of papers to be linked to society.

Models	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variables	P (News)				P (Tweets)			
5-year Disruption (Pct.)	-0.4443*** (0.0098)	0.1947*** (0.0108)	0.2433*** (0.0109)	0.5299*** (0.0130)	-0.5143*** (0.0033)	0.0666*** (0.0040)	0.0837*** (0.0040)	0.3272*** (0.0048)
ln (Team size)			0.1540*** (0.0040)	0.0870*** (0.0043)			0.0237*** (0.0016)	-0.0176*** (0.0017)

International team			0.0870*** (0.0054)	0.0710*** (0.0057)			0.1635*** (0.0020)	0.1491*** (0.0022)
Interdisciplinary team			0.1255*** (0.0053)	0.1156*** (0.0055)			0.0898*** (0.0019)	0.0843*** (0.0021)
Funding				0.2148*** (0.0068)				0.1220*** (0.0030)
ln (#Reference)				0.3846*** (0.0044)				0.3538*** (0.0017)
Rao-Stirling				0.2903*** (0.0552)				-0.0204 (0.0232)
Year FE	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Field FE	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Journal FE	No	**Yes**	**Yes**	**Yes**	No	**Yes**	**Yes**	**Yes**
Obs.	38,104,430	32,986,537	32,986,537	29,378,326	38,104,430	36,575,555	36,575,555	32,637,778
***Pseudo* R^2**	0.09789	0.22395	0.22567	0.23424	0.28691	0.42075	0.42141	0.43286

Note: robust standard errors are reported in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The absence of observations in certain models is attributed to instances where only one value is available for a particular year, field, and journal combination, often indicating a lack of associated papers within that technological/societal domain. Furthermore, missing values in the 'Rao-Stirling' variable contribute to the observed absence of data points.

Table S8. Effect of the 5-year disruptive citation on the likelihood of papers to be linked to technology.

Models	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Dependent Variables	**P (Patents)**				**P (Clinical trials)**			
ln (5-year Disruptive citation)	0.6762*** (0.0006)	0.6148*** (0.0007)	0.6083*** (0.0007)	0.6016*** (0.0007)	1.009*** (0.0016)	0.9443*** (0.0020)	0.9295*** (0.0021)	0.8988*** (0.0022)
ln (Team size)			0.1417*** (0.0012)	0.1337*** (0.0013)			0.2970*** (0.0034)	0.2846*** (0.0035)
International team			-0.0782*** (0.0019)	-0.0920*** (0.0020)			-0.0019 (0.0054)	-0.0161** (0.0056)
Interdisciplinary team			0.1096*** (0.0015)	0.1108*** (0.0015)			-0.0975*** (0.0047)	-0.1212*** (0.0048)
Funding				0.2385*** (0.0021)				0.3050*** (0.0055)
ln (#Reference)				0.2312*** (0.0012)				0.2053*** (0.0033)
Rao-Stirling				-1.405*** (0.0101)				-0.7387*** (0.0351)
Year FE	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Field FE	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Journal FE	No	**Yes**	**Yes**	**Yes**	No	**Yes**	**Yes**	**Yes**
Obs.	38,104,430	37,356,439	37,356,439	33,335,435	38,104,430	19,533,842	19,533,842	17,899,089
***Pseudo* R^2**	0.19891	0.27707	0.27817	0.277	0.2784	0.24735	0.24999	0.25113

Note: robust standard errors are reported in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The absence of observations in certain models is attributed to instances where only one value is available for a particular year, field, and journal combination, often indicating a lack of associated papers within that technological/societal domain. Furthermore, missing values in the 'Rao-Stirling' variable contribute to the observed absence of data points.

Table S9. Effect of the 5-year disruptive citation on the likelihood of papers to be linked to society.

Models	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Dependent Variables	**P (News)**				**P (Tweets)**			
ln (5-year Disruptive citation)	0.9262*** (0.0020)	0.7388*** (0.0024)	0.7372*** (0.0024)	0.7392*** (0.0026)	0.4744*** (0.0009)	0.4376*** (0.0010)	0.4357*** (0.0010)	0.4424*** (0.0011)
ln (Team size)			-0.0302*** (0.0041)	-0.0480*** (0.0043)			-0.0438*** (0.0016)	-0.0736*** (0.0017)
International team			0.0580*** (0.0055)	0.0520*** (0.0058)			0.1489*** (0.0020)	0.1383*** (0.0022)
Interdisciplinary team			0.1154*** (0.0053)	0.1126*** (0.0056)			0.0804*** (0.0019)	0.0798*** (0.0021)
Funding				0.1553*** (0.0069)				0.0878*** (0.0030)
ln (#Reference)				0.1188*** (0.0042)				0.2165*** (0.0017)

Rao-Stirling				0.4785*** (0.0573)				0.0544* (0.0237)
Year FE	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Field FE	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Journal FE	No	**Yes**	**Yes**	**Yes**	No	**Yes**	**Yes**	**Yes**
Obs.	38,104,430	32,986,537	32,986,537	29,378,326	38,104,430	36,575,555	36,575,555	32,637,778
***Pseudo* R^2**	0.18094	0.26829	0.26851	0.27433	0.30857	0.43518	0.43557	0.44616

Note: robust standard errors are reported in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The absence of observations in certain models is attributed to instances where only one value is available for a particular year, field, and journal combination, often indicating a lack of associated papers within that technological/societal domain. Furthermore, missing values in the 'Rao-Stirling' variable contribute to the observed absence of data points.

References

Funk, R. J., & Owen-Smith, J. (2017). A dynamic network measure of technological change. *Management Science*, *63*(3), 791-817. https://doi.org/10.1287/mnsc.2015.2366

Gates, A. J., Ke, Q., Varol, O., & Barabási, A. L. (2019). Nature's reach: narrow work has broad impact. *Nature*, *575*(7781), 32-34. https://doi.org/10.1038/d41586-019-03308-7

Hendricks, G., Tkaczyk, D., Lin, J., & Feeney, P. (2020). Crossref: The sustainable source of community-owned scholarly metadata. *Quantitative Science Studies*, *1*(1), 414-427. https://doi.org/10.1162/qss_a_00022

Lee, C., Kogler, D. F., & Lee, D. (2019). Capturing information on technology convergence, international collaboration, and knowledge flow from patent documents: A case of information and communication technology. *Information Processing & Management*, *56*(4), 1576-1591. https://doi.org/10.1016/j.ipm.2018.09.007

Lin, Y., Evans, J. A., & Wu, L. (2022). New directions in science emerge from disconnection and discord. *Journal of Informetrics*, *16*(1), 101234. https://doi.org/10.1016/j.joi.2021.101234

Lin, Z., Yin, Y., Liu, L., & Wang, D. (2023). SciSciNet: A large-scale open data lake for the science of science research. *Scientific Data*, *10*(1), 315. https://doi.org/10.1038/s41597-023-02198-9

Liu, X., Bu, Y., Li, M., & Li, J. (2024). Monodisciplinary collaboration disrupts science more than multidisciplinary collaboration. *Journal of the Association for Information Science and Technology*, *75*(1), 59-78. https://doi.org/https://doi.org/10.1002/asi.24840

Marx, M., & Fuegi, A. (2022). Reliance on science by inventors: Hybrid extraction of in-text patent-to-article citations. *Journal of Economics & Management Strategy*, *31*(2), 369-392. https://doi.org/10.1111/jems.12455

Park, M., Leahey, E., & Funk, R. J. (2023). Papers and patents are becoming less disruptive over time. *Nature*, *613*(7942), 138-144. https://doi.org/10.1038/s41586-022-05543-x

Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of The Royal Society Interface*, *4*(15), 707-719. https://doi.org/10.1098/rsif.2007.0213

Wu, L. F., Wang, D. S., & Evans, J. A. (2019). Large teams develop and small teams disrupt science and technology. *Nature*, *566*(7744), 378-+. https://doi.org/10.1038/s41586-019-0941-9

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The Increasing Dominance of Teams in Production of Knowledge. *Science*, *316*(5827), 1036-1039. https://doi.org/10.1126/science.1136099

Yang, A. J. (2024). Unveiling the impact and dual innovation of funded research. *Journal of Informetrics*, *18*(1), 101480. https://doi.org/10.1016/j.joi.2023.101480

Yang, A. J., Gong, H., Wang, Y., Zhang, C., & Deng, S. (2024). Rescaling the disruption index reveals the universality of disruption distributions in science. *Scientometrics*, *129*(1), 561-580. https://doi.org/10.1007/s11192-023-04889-x

Yang, A. J., Hu, H., Zhao, Y., Wang, H., & Deng, S. (2023). From consolidation to disruption: A novel way to measure the impact of scientists and identify laureates. *Information Processing & Management*, *60*(5), 103420. https://doi.org/10.1016/j.ipm.2023.103420

Zhai, S., Xia, L., Chi, M., & Li, X. (2024). Configurational determinants of time-to-win in NSFC youth program funding: Insights from Chinese library and information science. *Information Processing & Management*, *61*(3), 103649. https://doi.org/10.1016/j.ipm.2024.103649