A Online Appendix

A.1 Data

Our data are extracted from the OpenAlex database, as of June 2022, which includes approximately 238 million documents. To focus on AI, we narrowed our scope to publications tagged with the OpenAlex Concepts "Artificial Intelligence" and "Machine Learning", including their hierarchical ancestors. This selection results in a corpus of 16,493,355 articles. Following Arts et al. (2023), we removed articles with duplicate titles, abstracts, or DOIs, and limited the selection to journal articles, published conference proceedings, and preprints. This procedure reduced our dataset to 8,736,902 articles. We further refined the dataset to include only those articles for which at least one author's geolocation was identifiable, resulting in 5,759,007 articles. We confined our analysis to the period between 2000 and 2021, leaving us with 4,744,748 articles. We finally retain only articles with at least one author affiliated with a Census Metropolitan Area, reducing our dataset to 4,496,336 articles. Of this sample, 188,316 articles were authored by a Canadian-affiliated author. Table A1 presents the breakdown of the publications into the three organizational categories.

We then aggregate our analysis at the census metropolitan area (CMA). For Canada we rely on Statcan 2021 data to assign to each affiliation on an article a census metropolitan area based on the ROR geolocation[2], for other CMAs we make use of GHSL-OECD Functional Urban Areas from 2015.[3] Our final sample is composed of 2,910 metropolitan areas participating in AI worldwide, with 38 of them in Canada.[4]

Table A1: Number of articles in each classification

Category	**Total**	**Only PR**	**Comp.**	**Hops.**
Worldwide	4,496,336	3,771,645	223,439	187,527
Canada	188,316	156,764	3,518	11,509

A.2 Indicators used

Our variable of interest is the number of publications in AI involving a hospital from a given CMA after 2017. We calculate total research output per CMA before 2017, splitting between Only public research articles and articles involving at least one company from the given CMA. We also compute weighted measures that consider the number of citations received by a given article, the output measurement at the CMA level is then the sum of all average number of citations per year received by articles published in the given CMA. Second, we calculate the Revealed Symmetric Comparative Advantage (RSCA) Index derived from the revealed comparative advantage index [Balassa, 1965], to account for the specialization in knowledge production in a given CMA. For each CMA we derive three different metrics of specialization in AI, a specialization index for public research, one for companies, and one for hospitals.[5] Here also we compute weighted RSCA metrics by using the average number of citations per year received by articles. Finally, we constructed a collaboration network among researchers who publish AI. Our approach involves constructing a network at the

[2]Data can be found here https://www12.statcan.gc.ca/census-recensement/2021/geo/sip-pis/boundary-limites/index2021-eng.cfm?year=21

[3]Data can be found here https://jeodpp.jrc.ec.europa.eu/ftp/jrc-opendata/GHSL/GHS_FUA_UCDB2015_GLOBE_R2019A/V1-0/

[4]There are 41 CMA in Canada but Brantford, Chilliwack and Drummondville are not included in our sample as they are not active in AI. Some CMA includes cities from two different countries, in the case of US and Canada, we kept them separated

[5]Note that to compute RSCA we used the whole OpenAlex database to account for other publications not related to AI

article level and then aggregating it at the CMA level. Specifically, the strength of a link between two CMAs is determined by the number of articles involving researchers from both areas. We then calculate the nodes eigenvector centrality to assign scores to each CMA based on their centrality and influence, considering both the quantity and quality of connections. Thus, utilizing eigenvector centrality allows us to effectively assess how CMAs are interconnected with other central urban areas within the network. We explore two distinct networks: one incorporating public research and companies, and another focusing solely on hospitals. Additionally, in citation-weighted networks, links are weighted based on the average number of citations received per year, providing a more accurate reflection of the relevance of these collaborations.

Output: To measure a metropolitan area's scientific output, we count the number of scientific articles published by authors affiliated with institutions in that urban area.

Revealed Symmetric Comparative Advantage: The Revealed Comparative Advantage (RCA) proposed by Balassa [1965] is a quantitative measure used to identify the extent to which a region has a specialized advantage in a particular sector or activity, relative to a larger comparative framework. In this article the RCA index is adapted to quantify the CMA's specialization in specific fields of research.

$$RCA_{i,AI} = \frac{\left(\frac{P_{i,AI}}{P_{i\cdot}}\right)}{\left(\frac{P_{\cdot,AI}}{P_{\cdot\cdot}}\right)}$$

where $P_{i,AI}$ is the number of publications from city i in AI. $P_{i\cdot}$ is the total number of publications from city i in all fields. $P_{\cdot,AI}$ is the total number of global publications in AI. $P_{\cdot\cdot}$ is the total number of global publications in all fields.

We used a symmetric version of the RCA as proposed by Laursen [2015], the Revealed Symmetric Comparative Advantage (RSCA):

$$RSCA_{i,AI} = \frac{RCA_{i,AI} - 1}{RCA_{i,AI} + 1}$$

Eigenvector centrality: Eigenvector centrality is a network analysis measure used to determine the relative importance of nodes within a network. Unlike simpler centrality measures that focus on immediate connections (such as degree centrality), eigenvector centrality considers both the quantity and quality of connections. The core idea is that a node is considered more central if it is connected to other nodes that are themselves central.

More formally the eigenvector centrality of a node x_i is:

$$x_i = \frac{1}{\lambda} \sum_{j=1}^{n} a_{ij} x_j$$

where x_i is the eigenvector centrality score of node i. λ is the largest eigenvalue of the adjacency matrix A. a_{ij} is an element of the adjacency matrix A of the network. a_{ij} is the number of coauthored articles from node i to node j if there is a connection, and 0 otherwise. Since we are using an undirected network, A is symmetric, and $a_{ij} = a_{ji}$.

Weighted metrics: For each article, we calculate the average number of citations it receives per year since its publication and use it as a weighting factor in our analysis. The cumulative metric for a given urban area is then derived by summing these weighted averages across all articles attributed to that location. Similarly, when assessing specialization indexes, each article's contribution is weighted by its annual citation rate. Finally, in the network analysis, the weight assigned to the link between two urban areas in a specific article is determined by the article's average yearly citation count.

A.3 Descriptives Statistics

Figure A1: Number and Share of AI publication over time

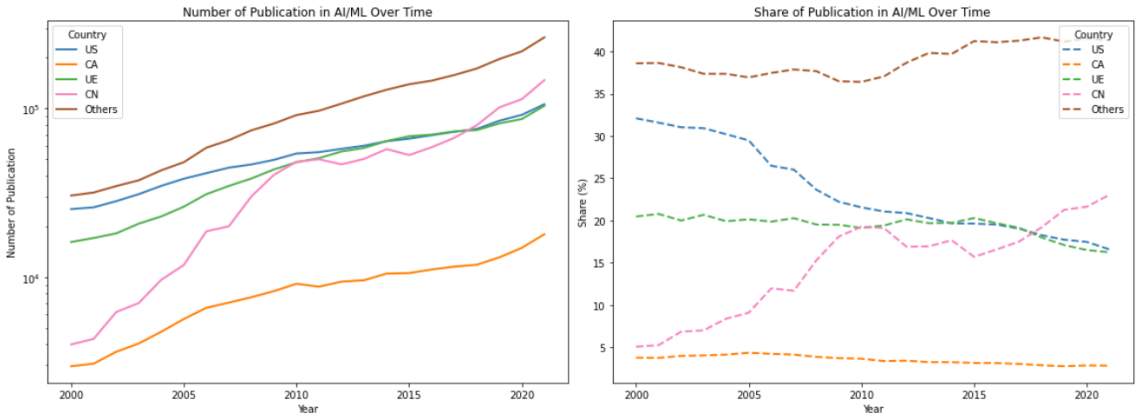


Table A2 and Table A3 show the descriptive statistics at census metropolitan area level for our sample for Canada and for the entire database respectively. In both cases, we kept in our analysis cities with more than 100 publications from public research institutions only between 2000 and 2016 included, along with hospitals having participated at least once in AI articles after 2017.

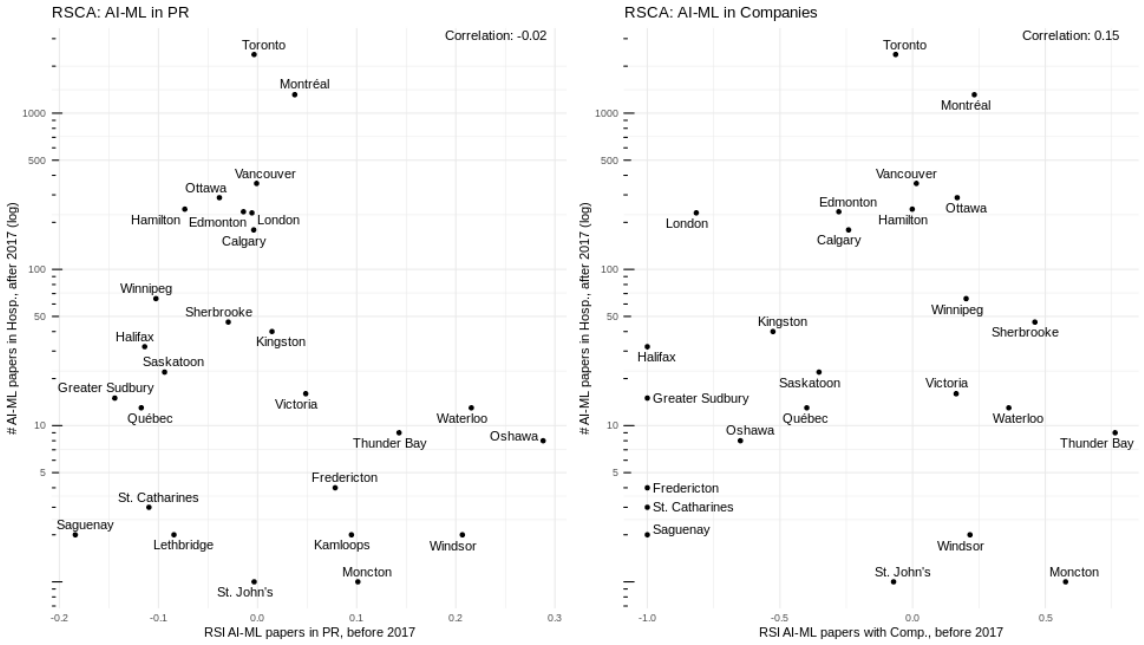
Table A2: Descriptive Statistics (Canada)

	Min	25%	Med	Mean	75%	Max
# Articles (Hosp. - before 2017)	1	3.5	16	204.2	204.5	2375
# Articles (PR - before 2017 - log)	118	778.5	2448	4020.4	4947	16940
# Articles (PR - before 2017 - log - Weighted)	153	1746	4784	9319	10801	40233
# Articles (Comp. - before 2017)	0	0.5	5	92.26	49.5	808
# Articles (Comp. - before 2017 - Weighted)	0	0	2.868	120.25	102.085	1146.753
Eigenvector cent. (PR/Comp. - before 2017)	0	0.003	0.008	0.015	0.017	0.071
Eigenvector cent. (PR/Comp. - before 2017 - Weighted)	0.001	0.002	0.006	0.014	0.013	0.076
Eigenvector cent. (Hosp. - before 2017)	0	0.001	0.003	0.015	0.015	0.153
Eigenvector cent. (Hosp. - before 2017 - Weighted)	0	0.001	0.004	0.019	0.015	0.214
Relative Spe. Index (PR - before 2017)	-0.184	-0.089	-0.004	0.004	0.063	0.288
Relative Spe. Index (PR - before 2017 - Weighted)	-0.357	-0.124	-0.011	-0.01	0.114	0.284
Relative Spe. Index (Comp. - before 2017)	-1	-0.649	-0.072	-0.21	0.202	0.763
Relative Spe. Index (Comp. - before 2017 - Weighted)	-1	-0.874	-0.226	-0.276	0.183	0.667
Relative Spe. Index (Hosp. - before 2017)	-1	-0.125	-0.059	-0.154	0.015	0.287
Relative Spe. Index (Hosp. - before 2017 - Weighted)	-1	-0.193	-0.035	-0.138	0.086	0.727

Table A3: Descriptive Statistics (World wide)

	Min	25%	Med	Mean	75%	Max
\# Articles (Hosp. - before 2017)	1	5	18	124.7	76.5	7834
\# Articles (PR - before 2017 - log)	101	399	1172	3256	2938	117101
\# Articles (PR - before 2017 - log - Weighted)	26	510.6	1954.2	6580.6	5709.3	162916.1
\# Articles (Comp. - before 2017)	0	0	1	173.7	23.5	13391
\# Articles (Comp. - before 2017 - Weighted)	0	0	0	411.49	29.51	65065.49
Eigenvector cent. (PR/Comp. - before 2017)	0	0.001	0.004	0.014	0.011	0.363
Eigenvector cent. (PR/Comp. - before 2017 - Weighted)	0	0.001	0.004	0.012	0.01	0.377
Eigenvector cent. (Hosp. - before 2017)	0	0	0.001	0.012	0.007	0.458
Eigenvector cent. (Hosp. - before 2017 - Weighted)	0	0	0.001	0.011	0.006	0.447
Relative Spe. Index (PR - before 2017)	-1	-0.155	-0.022	-0.023	0.114	0.537
Relative Spe. Index (PR - before 2017 - Weighted)	-1	-0.162	-0.013	-0.021	0.115	0.68
Relative Spe. Index (Comp. - before 2017)	-1	-0.574	-0.184	-0.248	0.109	0.771
Relative Spe. Index (Comp. - before 2017 - Weighted)	-1	-0.796	-0.254	-0.287	0.124	0.805
Relative Spe. Index (Hosp. - before 2017)	-1	-0.357	-0.112	-0.199	0.041	0.772
Relative Spe. Index (Hosp. - before 2017 - Weighted)	-1	-0.532	-0.118	-0.229	0.099	0.916

Figure A2: Specialization metrics in public research, companies and hospitals

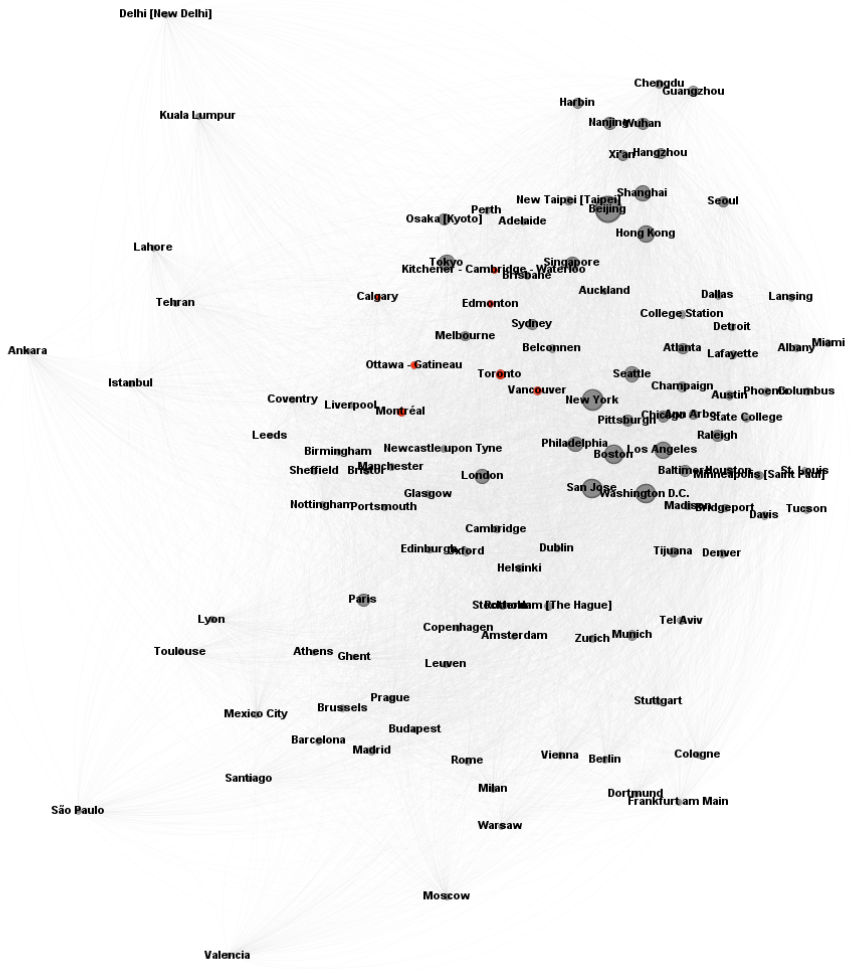


A.4 Network Analysis

Network graphs in this section show CMAs with the highest eigenvector centrality. Each node corresponds to an urban area (CMA). The size of each node reflects the urban area's eigenvector centrality. Node positions are determined using ForceAtlas2 and nodes positioned centrally hold greater influence. Edge sizes denote the number of articles involving researchers from both connected areas. We construct two distinct networks, the first is based on articles from public research and companies in a given CMA (PR + Comp.), while the second is restricted to collaborations between hospitals in different CMAs (Hosp.). In the international network (See Figure A3 for PR + Comp. and Figure 2 for Hosp. network), we represented the top 5% and top 10% most central CMAs respectively, to maintain readability. Nodes and edges are color-coded by geographical region, with red representing Canadian urban areas.

International Network

Figure A3: AI/ML collaboration network among public research institutes and companies across urban areas. Top 5%



Canadian Network

Figure A4: AI/ML collaboration network among public research institutes and companies across urban areas in Canada

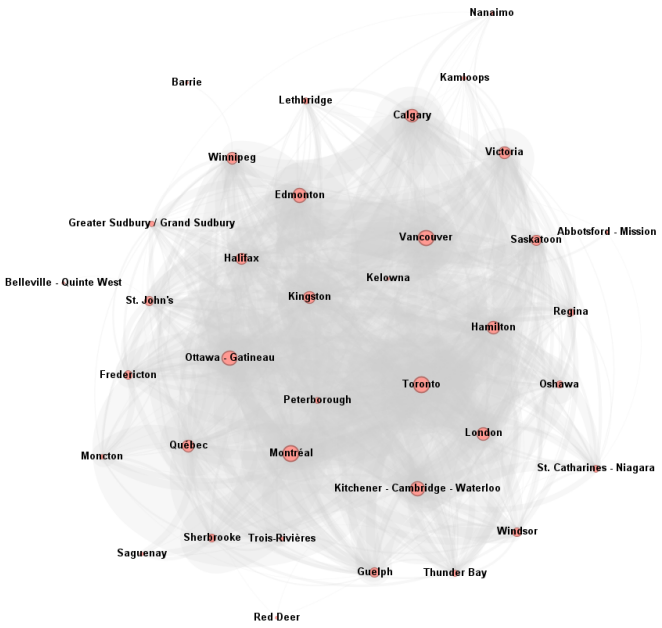
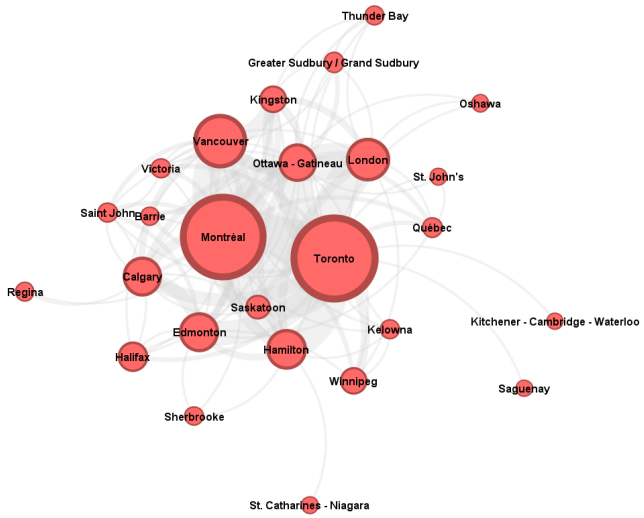


Figure A5: AI/ML collaboration network among hospitals across urban areas in Canada



A.5 Regression Analysis

In this section, we present the tables of the regressions (OLS), the dependent variable is the number of AI articles written in a given CMA with at least one hospital. We also did the same analysis using the number of AI articles weighted by the mean annual number of citations. The analysis was conducted at the Canadian level, and similar regressions were run at the global level to provide a broader perspective.

A.5.1 Canada level

Table A4: Number of AI articles written by hospitals in Canadian CMAs after 2017 (OLS)

	Dependent variable:							
	# Articles (Hosp. - after 2017 - log)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Pop. 2016 (log)	0.981***	1.110***	0.810**	0.745*	0.766**	0.545	0.014	0.057
	(0.326)	(0.351)	(0.385)	(0.394)	(0.318)	(0.459)	(0.412)	(0.406)
# Articles (PR - before 2017 - log)	0.558**		0.472			0.291	0.486*	0.142
	(0.265)		(0.286)			(0.406)	(0.248)	(0.364)
# Articles (Comp. - before 2017)		0.265	0.156			0.105	0.147	0.114
		(0.178)	(0.185)			(0.192)	(0.157)	(0.160)
Eigenvector cent. (PR/Comp. - before 2017)				107.260**		55.916		108.495
				(47.728)		(75.122)		(82.146)
Eigenvector cent.^{2} (PR/Comp. - before 2017)				−909.931		−338.585		−2,002.118
				(589.253)		(884.390)		(1,232.890)
Eigenvector cent. (Hosp. - before 2017)					98.253***		93.570***	107.214***
					(33.392)		(30.369)	(33.383)
Eigenvector cent.^{2} (Hosp. - before 2017)					−526.839**		−479.293**	−401.985*
					(196.981)		(179.929)	(192.977)
Constant	−13.619***	−11.712**	−11.126**	−7.436	−7.286*	−6.844	−1.546	−0.197
	(2.807)	(4.208)	(4.079)	(4.756)	(3.951)	(5.381)	(4.554)	(4.818)
Observations	27	27	27	27	27	27	27	27
R^2	0.762	0.742	0.769	0.776	0.798	0.785	0.848	0.869
Adjusted R^2	0.742	0.720	0.739	0.746	0.771	0.734	0.812	0.820
Residual Std. Error	1.041	1.084	1.047	1.032	0.980	1.058	0.890	0.870
F Statistic	38.416***	34.474***	25.548***	26.510***	30.243***	15.324***	23.403***	17.934***

Notes: This table reports coefficients of the effect of Canadian metropolitan area characteristics on AI publication in local hospitals. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively. Effects are estimated using OLS.

Table A5: Number of AI articles written by hospitals in Canadian CMAs after 2017 (OLS – Citation Weighted)

	Dependent variable:							
	# Articles (Hosp. - after 2017 - log)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Pop. 2016 (log)	0.947***	1.278***	0.848**	0.953**	1.150***	0.628	0.317	0.255
	(0.307)	(0.336)	(0.370)	(0.369)	(0.305)	(0.441)	(0.435)	(0.431)
# Articles (PR - before 2017 - Weighted)	0.545**		0.515**			0.482	0.527**	0.376
	(0.227)		(0.237)			(0.313)	(0.229)	(0.290)
# Articles (Comp. - before 2017 - Weighted)		0.153	0.074			0.062	0.109	0.117
		(0.154)	(0.148)			(0.157)	(0.142)	(0.146)
Eigenvector cent. (PR/Comp. - before 2017 - Weighted)				100.942*		20.153		52.753
				(55.014)		(74.560)		(71.558)
Eigenvector cent.^{2} (PR/Comp. - before 2017 - Weighted)				−919.767		−31.860		−1,148.932
				(612.233)		(824.406)		(916.602)
Eigenvector cent. (Hosp. - before 2017 - Weighted)					45.657		47.781*	98.572**
					(27.624)		(25.198)	(40.664)
Eigenvector cent.^{2} (Hosp. - before 2017 - Weighted)					−175.893		−174.991	−332.788**
					(117.991)		(107.775)	(147.478)
Constant	−13.452***	−13.610***	−12.104***	−9.856**	−11.967***	−9.215*	−5.806	−4.294
	(2.700)	(4.076)	(3.856)	(4.471)	(3.804)	(4.973)	(4.743)	(4.964)
Observations	27	27	27	27	27	27	27	27
R^2	0.773	0.729	0.775	0.759	0.750	0.785	0.814	0.837
Adjusted R^2	0.754	0.707	0.746	0.727	0.718	0.734	0.770	0.777
Residual Std. Error	1.017	1.111	1.034	1.070	1.090	1.058	0.983	0.967
F Statistic	40.804***	32.296***	26.432***	24.135***	23.014***	15.340***	18.399***	13.974***

Notes: This table reports coefficients of the effect of Canadian metropolitan area characteristics on AI publication in local hospitals. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively. Effects are estimated using OLS.

A.5.2 World level

Table A6: Number of AI articles written by hospitals in CMAs after 2017 (OLS)

	Dependent variable:							
	# Articles (Hosp. - after 2017 - log)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Pop. 2016 (log)	0.877***	0.966***	0.799***	0.941***	0.919***	0.794***	0.661***	0.658***
	(0.052)	(0.053)	(0.058)	(0.048)	(0.041)	(0.058)	(0.055)	(0.055)
# Articles (PR - before 2017 - log)	0.311***		0.271***			0.203***	0.258***	0.295***
	(0.041)		(0.043)			(0.052)	(0.040)	(0.049)
# Articles (Comp. - before 2017)		0.134***	0.082***			0.061**	0.034	0.043
		(0.027)	(0.027)			(0.028)	(0.026)	(0.026)
Eigenvector cent. (PR/Comp. - before 2017)				19.468***		7.941*		−4.544
				(3.336)		(4.119)		(4.011)
Eigenvector cent.^{2} (PR/Comp. - before 2017)				−44.318***		−14.448		9.479
				(12.749)		(14.156)		(13.582)
Eigenvector cent. (Hosp. - before 2017)					30.187***		28.477***	29.561***
					(3.045)		(2.975)	(3.101)
Eigenvector cent.^{2} (Hosp. - before 2017)					−55.112***		−53.190***	−54.204***
					(8.274)		(8.021)	(8.283)
Constant	−12.052***	−11.528***	−10.773***	−11.195***	−10.902***	−10.331***	−8.849***	−9.008***
	(1.190)	(1.285)	(1.257)	(1.234)	(1.142)	(1.268)	(1.179)	(1.187)
Observations	807	807	807	807	807	807	807	807
R^2	0.684	0.671	0.688	0.682	0.713	0.691	0.733	0.734
Adjusted R^2	0.637	0.622	0.641	0.634	0.670	0.644	0.692	0.692
Residual Std. Error	1.045	1.067	1.039	1.050	0.997	1.035	0.963	0.963
F Statistic	14.620***	13.739***	14.735***	14.295***	16.570***	14.629***	17.938***	17.621***

Notes: This table reports coefficients of the effect of Canadian metropolitan area characteristics on AI publication in local hospitals. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively. Effects are estimated using OLS.

Table A7: Number of AI articles written by hospitals in CMAs after 2017 (OLS – Citation Weighted)

	Dependent variable:							
	# Articles (Hosp. - after 2017 - log)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Pop. 2016 (log)	0.897***	0.963***	0.808***	0.993***	0.961***	0.812***	0.693***	0.685***
	(0.050)	(0.052)	(0.056)	(0.045)	(0.041)	(0.056)	(0.054)	(0.054)
# Articles (PR - before 2017 - Weighted)	0.272***		0.237***			0.196***	0.228***	0.251***
	(0.036)		(0.037)			(0.041)	(0.035)	(0.039)
# Articles (Comp. - before 2017 - Weighted)		0.128***	0.083***			0.062**	0.049**	0.058**
		(0.024)	(0.025)			(0.026)	(0.023)	(0.024)
Eigenvector cent. (PR/Comp. - before 2017 - Weighted)				17.864***		6.405*		−4.174
				(3.367)		(3.886)		(3.986)
Eigenvector cent.^{2} (PR/Comp. - before 2017 - Weighted)				−36.500***		−8.991		7.246
				(11.383)		(12.189)		(13.160)
Eigenvector cent. (Hosp. - before 2017 - Weighted)					26.300***		24.749***	26.026***
					(3.038)		(2.943)	(3.180)
Eigenvector cent.^{2} (Hosp. - before 2017 - Weighted)					−46.243***		−45.131***	−46.309***
					(8.379)		(8.071)	(9.168)
Constant	−12.049***	−11.488***	−10.658***	−11.898***	−11.461***	−10.492***	−9.047***	−9.061***
	(1.188)	(1.277)	(1.248)	(1.215)	(1.155)	(1.248)	(1.187)	(1.188)
Observations	807	807	807	807	807	807	807	807
R^2	0.685	0.672	0.690	0.679	0.703	0.693	0.726	0.727
Adjusted R^2	0.638	0.623	0.644	0.631	0.659	0.646	0.685	0.685
Residual Std. Error	1.043	1.065	1.036	1.053	1.014	1.032	0.974	0.974
F Statistic	14.679***	13.815***	14.867***	14.149***	15.817***	14.757***	17.352***	17.052***

Notes: This table reports coefficients of the effect of Canadian metropolitan area characteristics on AI publication in local hospitals. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively. Effects are estimated using OLS.