Appendix

Appendix A: methodology and quantity of diaspora & returnee research

Method of distinguishing the diaspora papers

Appendix Figure A1 shows how we estimate the number of Chinese diaspora papers in 2018. We used the Scopus query string to download files that contain bibliographic data of papers from the Scopus online website[1]. The files include the paper's unique *EID*[2]; author information: last names and first name initials of authors, authors' addresses, *Author ID*[3]; and the number of citations received to the date of downloading. We determined the Chinese ethnicity of authors by whether their last names are common Chinese last names as reported in the Chinese Ministry of Public Security's list of Chinese last names[4].　By this count, 30.8% of papers with only US addresses had an ethnic Chinese author.

To differentiate mainland-born Chinese researchers from Chinese ethnic persons born outside the mainland on US-addressed papers, we divide authors with Chinese last names into those with Chinese first names indicative of being mainland born and those with non-Chinese first names. Because the online website files do not provide full *first names* of authors, we randomly sampled 2,000 papers of ethnic Chinese persons in each of our groups of US-addressed papers (USO, US-C, US-C-ROW, and US-ROW) and used the EIDs of sampled papers to retrieve author first names via the Scopus API portal. Appendix Table A1 gives the details of the sampled US addressed papers.

We use authors' first names to estimate the proportion of ethnic Chinese authors likely to have been born in mainland China based on whether their first names follow the grammar of the Hanyu Pinyin translation system used in mainland China[5]. Our scheme labels Jianguo Wang as mainland China born and John Wang as non-mainland China born. Because the structure of pinyin syllables used in mainland China differs from that in other Chinese language speaking areas, our program also differentiates mainland names from other Chinese language area names[6]. We identify authors with typical Chinese last and first names at a US address as a diaspora author in the US.[7]

---

[1]　https://www.scopus.com

[2]　See https://dev.elsevier.com/guides/ScopusSearchViews.html

[3]　Author ID is the unique identifier Scopus assigns to differentiate authors with similar names, see: https://service.elsevier.com/app/answers/detail/a_id/11212/c/10546/supporthub/scopus/kw/AU-ID/.

[4]The 2019 National Name Report: https://www.mps.gov.cn/n2254314/n6409334/c6874817/content.html lists the most common Chinese last names.　It covers the Chinese last names of 84.8% of the mainland population.

[5]　The program that distinguishes Chinese first names is available at GitHub: https://github.com/qingnanxie/Chinese-first-name.

[6]　For example, *Xie* is the mainland pinyin translation of 谢, which is translated as *Tes* in Hong Kong and *Hsieh* in Taiwan.

[7]　Our methodology misses the likely small number of authors who changed their names into non-Chinese names, Chinese ethnicity authors born outside of China but given a Chinese name, and authors with rare Chinese names.

Our identification includes US-addressed authors who may also have a China or other country address.

Multiplying those sample-based proportions of the four types papers (listed on the yellow arrows in Appendix Figure A1) to the number of each type of papers respectively, we estimated the number of USO papers with diaspora authors, the number of US-ROW papers with diaspora authors, the number of US-China-ROW papers with diaspora authors, and the number of US-China papers with diaspora authors.

Appendix Figure A1. Estimated number of diaspora papers in 2018: #US addressed papers with at least one Chinese diaspora author
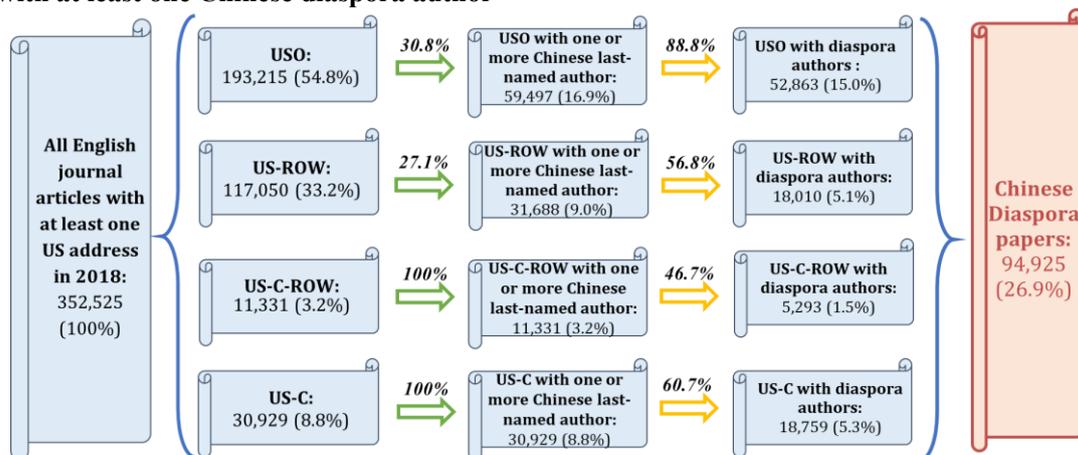
Note: Green arrows refer to calculations based on the query string results (population counts), yellow arrows refer to calculations based on sampling estimations.

**Appendix Table A1. Samples for US-addressed papers**

Data Sample	Purpose	Years Covered	Total number sampled
USO papers with Chinese last-named authors	・Estimate the number of USO papers with diaspora authors in 2015, 2016, 2017, 2018; ・Estimate the scientific impact of USO papers with diaspora authors (CiteScore & Citation for 2018 papers); ・Analyze the citing behavior of the 2016-2018 diaspora/returnee papers (3-year citation data of 2015 USO papers); ・Analyze diaspora and non-diaspora authors' co-authorship networks; ・Analyze the pre-2018 publications of diaspora and non-diaspora authors.	2015-2018	2,000 in each year for total of 8,000
US-ROW papers with Chinese last-named authors (No China address)	・Estimate the number of US-ROW papers with diaspora authors; ・Estimate the scientific impact of US-ROW papers with diaspora authors; ・Analyze diaspora and non-diaspora authors' co-authorship networks; ・Analyze the pre-2018 publications of diaspora and non-diaspora authors.	2018	2,000
US-C-ROW papers	・Estimate the quantity of US-C-ROW papers with diaspora authors; ・Estimate the quantity of US-C-ROW papers with returnee authors; ・Estimate the scientific impact of US-C-ROW papers with and without diaspora authors; ・Estimate the scientific impact of US-C-ROW papers with and without returnee authors; ・Analyze the role of diaspora authors and returnee-diaspora authors in the 2018 US-China collaboration; ・Analyze the co-authorship networks of diaspora, non-diaspora, returnee and non-returnee authors; ・Analyze the pre-2018 publications of diaspora, non-diaspora, returnee and non-returnee authors.	2018	2,000
US-C papers (No rest-of -world countries' address)	・Estimate the quantity of US-C papers with diaspora authors; ・Estimate the quantity of US-C papers with returnee authors; ・Estimate the scientific impact of US-C papers with and without diaspora authors; ・Estimate the scientific impact of US-C papers with and without returnee-diaspora authors; ・Analyze the role of diaspora and returnee-diaspora authors in the 2018 US-China collaboration; ・Analyze the co-authorship networks of diaspora, non-diaspora, returnee and non-returnee authors; ・Analyze the pre-2018 publications of diaspora, non-diaspora, returnee and non-returnee authors.	2018	2,000
USO papers	・Estimate the scientific impact of USO papers without diaspora authors	2018	2,000
US-ROW papers	・Estimate the scientific impact of US-ROW papers without diaspora authors	2018	2,000

Method of distinguishing the returnee papers

We estimated the number of China-addressed returnee papers in two steps, as shown in Appendix Figure 2. First, we searched English language journal articles for those with at least one China address and then divided those articles into: CO papers (China addresses only); C-ROW papers (China and rest-of-world countries addresses); US-C-ROW papers (US, China, and rest-of-world country addresses); and US-C papers (US and China addresses only). Appendix Table A2 gives the sampling details.

We sampled 2,000 of the papers in each group, giving a total sample of 8,000 papers, and took the Scopus Author ID associated with the papers to retrieve data on **all** Scopus-indexed publication of that author[8]. Our returnee authors are China-addressed authors who had an article with their name at a US address prior to their 2018 China-addressed article. The estimated proportion of China-addressed papers with a returnee author varied from 51.8% for US-C papers to 34.7% of C-ROW papers.   Multiplying the number of 2018 papers of each group by the estimated proportions of papers with returnee authors and summing across the 4 groups gives our bottom-line estimate that 38.3% of 2018 China-addressed papers had at least one author with prior US research experience.

Appendix Figure A2. Estimated number of returnee papers in 2018: #China addressed papers with at least one returnee author
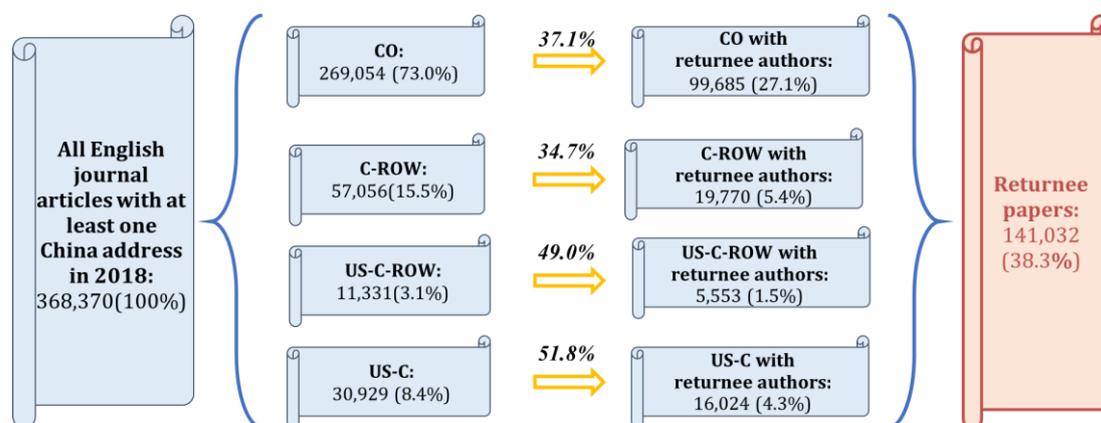
Note: Yellow arrows refer to calculations based on sampling estimations.

---

[8] We use all publications indexed in Scopus instead of just English journal S&E articles to capture as much as possible returnee authors.

Appendix Table A2. Samples for Additional China-addressed papers and ROW papers

Data Sample	Purpose	Years Covered	Total number sampled
CO papers	・Estimate the quantity of CO papers with and without returnee authors in 2015-2018; ・Estimate the scientific impact of CO papers with returnee authors (CiteScore & Citation for 2018 papers); ・Analyze the citing behavior of the 2016-2018 diaspora/returnee papers (3-year citation data of 2015 CO papers); ・Analyze the co-authorship networks of returnee and non-returnee authors; ・Analyze the pre-2018 publications of returnee and non-returnee authors.	2015-2018	2,000 in each year for total of 8,000
C-ROW papers	・Estimate the quantity of C-ROW papers with returnee authors; ・Estimate the scientific impact of C-ROW papers with and without returnee authors (CiteScore & Citation for 2018 papers); ・Analyze the co-authorship networks of returnee and non-returnee authors; ・Analyze the pre-2018 publications of returnee and non-returnee authors.	2018	2,000
ROW papers (The rest of the world papers)	・As the control group in our citation analysis to indicate the citation behaviors of papers without US and China address; ・The numbers of the rest of the world papers in 2016-2018 are counted based on the query string results (population counts).	2015	2,000

Appendix Table A3. US and China addressed papers with diaspora or returnee author, 2018

		#Papers	Proportion of all China-addressed papers	Proportion of all US-addressed papers
USO with Diaspora author		52,863	-	15.0%
US-ROW with Diaspora author		18,010	-	5.1%
USO without Diaspora author		140,352	-	39.8%
US-ROW without Diaspora author		99,040	-	28.1%
CO with Returnee author		99,685	27.1%	-
C-ROW with Returnee author		19,770	5.4%	-
CO without Returnee author		169,369	46.0%	-
C-ROW without Returnee author		37,286	10.1%	-
US-C-ROW with D	With Returnee authors	2,658	0.7%	0.8%
US-C with D		9,783	2.7%	2.8%
US-C-ROW without D		2,895	0.8%	0.8%
US-C without D		6,240	1.7%	1.8%
US-C-ROW with D	Without Returnee authors	2,635	0.7%	0.7%
US-C with D		8,976	2.4%	2.5%
US-C-ROW without D		3,143	0.9%	0.9%
US-C without D		5,930	1.6%	1.7%

Appendix A4. Estimating the Population of Diaspora Authors in the US, 2018

	Population based	Sample based		Estimated total #AU-IDs = #papers × avg# AU-ID per paper	
	#papers	Avg #US-ND AU-ID per paper	Avg #USD AU-ID per paper	#US-ND AU-ID	#USD AU-ID
USO with D	52,863	3.1	2.1	162,855	112,059
US-C with D	18,759	1.3	1.8	25,151	34,199
US-C-O with D	5,293	3.2	2.6	17,063	13,626
US-other with D	18,010	4.0	1.8	72,293	32,123
USO without D	140,352	5.2	0	729,901	0
US-C without D	12,170	2.2	0	26,636	0
US-C-O without D	6,038	3.8	0	22,733	0
US-other without D	99,040	3.3	0	326,981	0
Total	**352,525**	**3.9**	**0.5**	**1,383,613**	**192,007**

The total #AU-ID exceeds the #authors because one AU-ID could appear more than one time (this is also a measure of the productivity). To estimate the actual number of authors, we need to estimate the #times an AU-ID appears in 2018 papers. Dividing the total #AU-IDs by the average #times an AU-ID appears gives us the estimated author population.

#times an AU-ID appears in 2018 papers (#JA per AU-ID)		
	Sample based	
	US-ND AU-ID	USD AU-ID
USO with D	4.8	3.8
US-C with D	5.9	6.3
US-C-O with D	8.9	6.3
US-other with D	6.1	4.0
USO without D	4.9	-
US-C without D	6.1	-
US-C-O without D	9.6	-
US-other without D	5.6	-
Weighted #JA per AU-ID in 2018	**5.3**	**4.4**

Estimated Author Population in 2018 = *Total AU-IDs / Weighted #JA per AU-ID in 2018*

	US-ND author	USD author
Estimated Author Population	*1,383,613/5.3= 261,894 (85.8%)*	*192,007/4.4= 43,184 (14.2%)*

Note: Based on 8,000 sampled US-addressed papers per Appendix Table A1.

Appendix A5. Estimating the Population of Returnee Authors in China, 2018

	Population based	Sample based		Estimated total #AU-IDs = #papers $\times$ avg# AU-ID per paper	
	#papers	Avg #China-NR AU-ID per paper	Avg #Returnee AU-ID per paper	#China-NR AU-ID	#Returnee AU-ID
CO	269,054	6.2	0.5	1,659,794	137,621
China-other	57,056	3.9	0.5	223,089	27,273
US-China-other	11,331	3.9	0.8	44,308	9,594
US-China	30,929	4.9	0.8	151,739	25,116
Total	**368,370**	**5.6**	**0.5**	**2,078,930**	**199,604**

The total #AU-ID exceeds the #authors because one AU-ID could appear more than one time (this is also a measure of the productivity). To estimate the actual number of authors, we need to estimate the #times an AU-ID appears in 2018 papers. Dividing the total #AU-IDs by the average #times an AU-ID appears gives us the estimated author population.

#times an AU-ID appears in 2018 papers (#JA per AU-ID)		
	Sample based	
	China-NR AU-ID	Returnee AU-ID
CO	5.8	13.0
China-other	7.4	15.2
US-China-other	8.0	12.0
US-China	5.4	10.8
Weighted #JA per AU-ID in 2018	6.0	13.0

Estimated Author Population in 2018 = *Total AU-IDs / Weighted #JA per AU-ID in 2018*		
	China-NR author	Returnee author
Estimated Author Population	*2,078,930/6.0= 346,750 (95.8%)*	*199,604/13.0= 15,348 (4.2%)*

Note: Based on 8,000 sampled China-addressed papers per Appendix Table A1 & A2.

Appendix B: Scientific quality/impact of diaspora & returnee research in data set

This appendix documents the results of our analysis of CiteScores and citations, including various "robustness" on the functional form of the relations.

Appendix Figure B1. Power law distributions of CiteScores and 3-year citations of 2018 papers
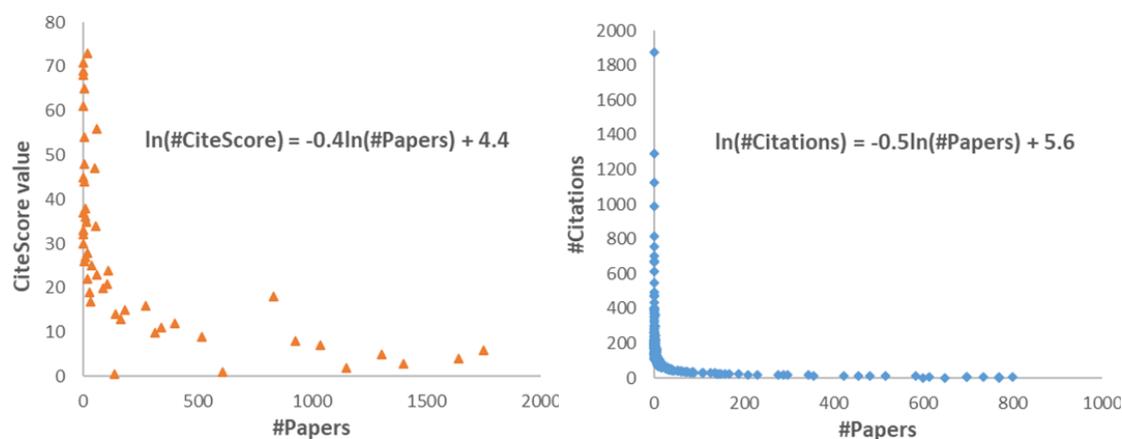
Note: For simplicity, we use Ln-Ln regression to estimate to powers of each distribution. Power law distributions of CiteScore are plotted based on 13,928 papers with valid author information, address information, and CiteScore values in our dataset. Power law distributions of 3-year citations are plotted based on 14,441 papers with valid author and address information in our dataset. The number of observations is smaller for CiteScores than for citations comes from papers that were published in newly established journals for which Scopus did not yet a CiteScore value.

Appendix Table B1. Average CiteScores and 3-year citations of 2018 US-addressed papers

	CiteScores	3-year citations
US-addressed papers	9.0	16.9
Non-diaspora papers	**8.4**	**14.9**
USO without D	8.3	13.9
US-ROW without D	8.7	15.7
US-C-ROW without D	9.4	24.4
US-C without D	7.1	14.1
Diaspora papers	**10.5**	**22.3**
USO with D	10.5	20.8
US-ROW with D	11.7	25.6
US-C-ROW with D	12.2	36.9
US-C with D	8.7	19.1

Appendix Table B2. Average CiteScores and 3-year citations of 2018 China-addressed papers

	CiteScores		3-year citations	
	Without returnee author	With returnee authors	Without returnee author	With returnee authors
CO	3.4	5.0	11.8	17.1
C-ROW	4.2	5.4	17.4	21.6
US-C-ROW with D	11.9	12.4	33.8	39.8
US-C with D	7.7	9.6	14.8	22.9
US-C-ROW without D	8.2	10.6	21.7	27.4
US-C without D	6.2	7.9	11.5	16.5
China-addressed papers	**3.9**	**5.8**	**13.2**	**18.8**
(Average for all)	(4.6)		(15.3)	

Appendix Table B3. Regression Estimates of Effect of Diaspora and Returnee Papers, Research Field, and Number of Authors of a Paper on CiteScore and Citations of 2018 US/China addressed Papers

	US addressed papers					
	CiteScore	*CiteScore*	*CiteScore*	*Citations*	*Citations*	*Citations*
Diaspora paper dummy	2.10*** (0.171)	1.26*** (0.168)	1.94*** (0.055)	7.44*** (0.854)	5.92*** (0.872)	6.66*** (0.848)
#Authors	-		0.05*** (0.004)	-	-	0.28*** (0.019)
Field dummy (21)	NO	Yes	NO	NO	Yes	NO
#Obs	10,127	10,127	10,127	10,441	10,441	10,441
Adjusted R²	0.0146	0.0819	0.0345	0.0071	0.0177	0.0263

	China addressed papers					
	CiteScore	*CiteScore*	*CiteScore*	*Citations*	*Citations*	*Citations*
Returnee paper dummy	2.03*** (0.155)	1.33*** (0.145)	1.64*** (0.149)	6.21*** (0.867)	5.51*** (0.866)	5.20*** (0.860)
#Authors	-	-	0.09*** (0.005)	-	-	0.41*** (0.031)
CO	-	--	-	-	--	-
C-ROW	-	0.68*** (0.201)	-	-	5.21*** (1.192)	-
US-C	-	3.85*** (0.201)	-	-	2.54** (1.201)	-
US-C-ROW	-	6.52*** (0/202)	-	-	15.82*** (1.209)	-
Field dummy (21)	NO	NO	Yes	NO	NO	Yes
#Obs	7,583	7,583	7,583	7,908	7,908	7,908
Adjusted R²	0.0218	0.1665	0.1185	0.0063	0.0302	0.0379

Note: 95% confidence interval, * $p<0.1$; ** $p<0.05$; *** $p<0.01$. Standard errors are in brackets. Observations without valid address information, author information, or CiteScore value are dropped.

"--" is the benchmark in the regression and "-" means that variable is not added into the regression.

Appendix Table B4. Regression Estimates of Effect of Number of Diaspora and Returnee Authors on CiteScore and Citations of 2018 US and China addressed Papers

	USO&US-ROW papers		CO&C-ROW papers		USC&USC-ROW papers	
	CiteScore	*Citations*	*CiteScore*	*Citations*	*CiteScore*	*Citations*
#Diaspora author	0.41*** (0.067)	2.18*** (0.315)	-	-	0.79*** (0.086)	2.75*** (0.506)
#Non-Diaspora US addressed authors	0.07*** (0.011)	0.37*** (0.051)	-	-	-0.004 (0.0233)	-0.03 (0.139)
#Returnee authors	-	-	0.8*** (0.06)	2.56*** (0.492)	0.49*** (0.112)	1.92*** (0.666)
#Non-retunree China addressed authors	-	-	0.14*** (0.018)	0.99*** (0.146)	-0.03 (0.04)	-0.21 (0.234)
#Rest-of world country authors	0.03*** (0.005)	0.15*** (0.025)	0.17*** (0.014)	1.15*** (0.121)	0.04*** (0.012)	0.37*** (0.071)
USO	--	--	-	-	-	-
US-ROW	0.33 (0.222)	3.83*** (1.052)	-	-	-	-
US-C-ROW	-	-	-	-	2.36*** (0.271)	10.83*** (1.599)
US-C	-	-	-	-	--	--
C-ROW	-	-	0.52*** (0.109)	3.3*** (0.905)	-	-
CO	-	-	--	--	-	-
Field dummy (21)	Yes	Yes	Yes	Yes	Yes	Yes
#Obs	6,345	6,533	3,801	4,000	3,782	3,908
Adjusted R2	0.108	0.0446	0.1973	0.0689	0.1226	0.0519

Note: 95% confidence interval, * $p<0.1$; ** $p<0.05$; *** $p<0.01$. Standard errors are in brackets. Observations without valid address information, author information, or CiteScore value are dropped.

"--" is the benchmark in the regression and "-" means that variable have no meaning in the regression.

Appendix Table B5. Regression Estimates of Effect of Diaspora and Returnee Papers on Ln(CiteScore) and Ln(Citations) of 2018 US/China addressed Papers

	USO&US-ROW papers		CO&C-ROW papers		USC&USC-ROW papers	
	LN(CiteScore)	*LN(Citations)*	*LN(CiteScore)*	*LN(Citations)*	*LN(CiteScore)*	*LN(Citations)*
Diaspora paper dummy	0.08*** (0.02)	0.16*** (0.029)	-	-	0.1*** (0.024)	0.1*** (0.035)
Returnee paper dummy	-	-	0.23*** (0.026)	0.16*** (0.035)	0.12*** (0.024)	0.16*** (0.035)
Diaspora and Returnee dummy					-0.07 (0.047)	0.12* (0.068)
LN(#Authors)	0.18*** (0.013)	0.41*** (0.02)	0.39*** (0.027)	0.42*** (0.036)	0.26*** (0.02)	0.43*** (0.029)
USO	--	--	-	-	-	-
US-ROW	-0.05*** (0.02)	-0.01 (0.03)	-	-	-	-
US-C-ROW	-	-	-	-	0.18*** (0.025)	0.31*** (0.037)
US-C	-	-	-	-	--	--
C-ROW	-	-	0.18*** (0.025)	0.19*** (0.034)	-	-
CO	-	-	--	--	-	-
Field dummy (21)	Yes	Yes	Yes	Yes	Yes	Yes
#Obs	6,345	6,251	3,801	3,796	3,782	3,794
Adjusted R2	0.2191	0.1593	0.2502	0.0996	0.2036	0.1472

Note: 95% confidence interval, * p<0.1; ** p<0.05; *** p<0.01. Standard errors are in brackets. Observations without valid address information, author information, or CiteScore value are dropped.

"--" is the benchmark in the regression and "-" means that variable have no meaning in the regression.

Appendix Table B6. Regression Estimates of Effect of Number of Diaspora and Returnee Authors on Ln(CiteScore) and Ln(Citations) of 2018 US and China addressed Papers

	USO&US-ROW papers		CO&C-ROW papers		USC&USC-ROW papers	
	LN(CiteScore)	*LN(Citations)*	*LN(CiteScore)*	*LN(Citations)*	*LN(CiteScore)*	*LN(Citations)*
#Diaspora author	0.047*** (0.0056)	0.107*** (0.0083)	-	-	0.073*** (0.0079)	0.092*** (0.0116)
#Non-Diaspora US addressed authors	0.004*** (0.0009)	0.008*** (0.0013)	-	-	-0.002 (0.0022)	-0.003 (0.0032)
#Returnee authors	-	-	0.19*** (0.016)	0.14*** (0.021)	0.056*** (0.0104)	0.057*** (0.0153)
#Non-retunree China addressed authors	-	-	0.04*** (0.005)	0.04*** (0.006)	-0.003 (0.0037)	0.003 (0.0054)
#Rest-of world country authors	0.002*** (0.0004)	0.004*** (0.0007)	0.03*** (0.004)	0.05*** (0.005)	0.004*** (0.0011)	0.007*** (0.0016)
USO	--	--	-	-	-	-
US-ROW	0.042* (0.0186)	0.217*** (0.0281)	-	-	-	-
US-C-ROW	-	-	-	-	0.249*** (0.0251)	0.442*** (0.037)
US-C	-	-	-	-	--	--
C-ROW	-	-	0.21*** (0.029)	0.18*** (0.039)	-	-
CO	-	-	--	--	-	-
Field dummy (21)	Yes	Yes	Yes	Yes	Yes	Yes
#Obs	6,345	6,251	3,801	3,796	3,782	3,794
Adjusted R2	0.2078	0.1268	0.2365	0.0938	0.1822	0.108

Note: 95% confidence interval, * p<0.1; ** p<0.05; *** p<0.01. Standard errors are in brackets. Observations without valid address information, author information, or CiteScore value are dropped.

"--" is the benchmark in the regression and "-" means that variable have no meaning in the regression.

Appendix Table B7. Regression Estimates of Effect of Diaspora and Returnee Papers on CiteScore and Citations of all US addressed papers and all China addressed Papers, 2018

	US addressed papers				China addressed papers			
	CiteScore	*Citations*	*CiteScore*	*Citations*	*CiteScore*	*Citations*	*CiteScore*	*Citations*
Diaspora paper dummy	1.32*** (0.171)	5.95*** (0.884)	-	-	1.59*** (0.198)	6.65*** (1.219)	-	-
Returnee paper dummy	1.36*** (0.267)	5.63*** (1.376)	-	-	1.04*** (0.14)	4.71*** (0.861)	-	-
#Authors	0.04*** (0.004)	0.23*** (0.019)	-	-	0.06*** (0.005)	0.33*** (0.031)	-	-
#Diaspora author	-	-	0.53*** (0.052)	2.27*** (0.269)	-	-	0.83*** (0.064)	2.89*** (0.398)
#Non-Diaspora US addressed authors	-	-	0.06*** (0.009)	0.32*** (0.049)	-	-	-0.02 (0.017)	-0.12 (0.108)
#Returnee authors	-	-	0.47*** (0.116)	1.71*** (0.599)	-	-	0.52*** (0.068)	1.73*** (0.421)
#Non-retunree China addressed authors	-	-	-0.05 (0.041)	-0.28 (0.212)	-	-	0.01 (0.023)	0.17 (0.14)
#Rest-of world country authors	-	-	0.03*** (0.005)	0.17*** (0.024)	-	-	0.05*** (0.009)	0.42*** (0.054)
USO	--	--	--	--	-	-	-	-
US-ROW	0.27 (0.212)	3.00*** (1.095)	0.46** (0.213)	3.78*** (1.102)	-	-	-	-
US-C-ROW	0.26 (0.279)	9.27*** (1.438)	1.09*** (0.27)	13.34*** (1.393)	5.37*** (0.226)	10.08*** (1.389)	5.56*** (0.226)	11.49*** (1.395)
US-C	-2.05*** (0.278)	-2.77* (1.43)	-1.04*** (0.278)	2.19 (1.434)	2.89*** (0.23)	-1.91 (1.418)	3.09*** (0.212)	0.18 (1.31)
C-ROW	-	-	-	-	0.86*** (0.196)	5.11*** (1.196)	0.76*** (0.202)	4.38*** (1.241)
CO	-	-	-	-	--	--	--	--
Field dummy (21)	yes	yes	yes	yes	yes	yes	yes	yes
#Obs	10,127	10,441	10,127	10,441	7,583	7,908	7,583	7,908
Adjusted R²	0.1061	0.0452	0.1099	0.0475	0.2333	0.0558	0.2416	0.0551

Note: 95% confidence interval, * p<0.1; ** p<0.05; *** p<0.01. Standard errors are in brackets. Observations without valid address information, author information, or CiteScore value are dropped. "--" is the benchmark in the regression and "-" means that variable have no meaning in the regression. With one exception, the estimated coefficients on our categorization of papers by addresses show that US-China internationally collaborative papers receive higher CiteScores and more citations than other US-addressed papers (where USO is the base group) and receive higher CiteScores and more citations than other China-addressed papers (where CO is the base group). The exception are US-C collaborations, which obtain small negative estimated impacts on CiteScore and citations than USO that we trace mechanically to US-C collaborations having a smaller proportion of US-addressed authors than other US international collaborations. Calculations in Appendix Table B9 shows that conditional on the percentage of US researchers on collaborations and third country researchers, US-C papers obtain modestly higher CiteScore and many more citations than USO papers.

Appendix Table B8. Regression Estimates of Effect of Diaspora and Returnee Papers on Ln(CiteScore) and Ln(Citations) of all US addressed papers and all China addressed Papers, 2018

	US addressed papers				China addressed papers			
	Ln(CiteScore)	*Ln(Citations)*	*Ln(CiteScore)*	*Ln(Citations)*	*Ln(CiteScore)*	*Ln(Citations)*	*Ln(CiteScore)*	*Ln(Citations)*
Diaspora paper dummy	0.09*** (0.015)	0.14*** (0.022)	-	-	0.09*** (0.025)	0.12*** (0.034)	-	-
Returnee paper dummy	0.13*** (0.023)	0.16*** (0.034)	-	-	0.17*** (0.018)	0.15*** (0.024)	-	-
Ln(#Authors)	0.2*** (0.011)	0.42*** (0.016)	-	-	0.34*** (0.016)	0.44*** (0.022)	-	-
#Diaspora author	-	-	0.055*** (0.0046)	0.1*** (0.0067)	-	-	0.07*** (0.008)	0.1*** (0.011)
#Non-Diaspora US addressed authors	-	-	0.003*** (0.0008)	0.007*** (0.0012)	-	-	-0.01** (0.002)	-0.01*** (0.003)
#Returnee authors	-	-	0.051*** (0.0101)	0.045*** (0.0149)	-	-	0.08*** (0.009)	0.07*** (0.012)
#Non-retunree China addressed authors	-	-	-0.004 (0.0036)	0 (0.0053)	-	-	0.01*** (0.003)	0.01*** (0.004)
#Third country authors	-	-	0.002*** (0.0004)	0.004*** (0.0006)	-	-	0.01*** (0.001)	0.01*** (0.002)
USO	--	--	--	--	-	-	-	-
US-ROW	-0.06*** (0.019)	-0.01 (0.028)	0.05*** (0.019)	0.22*** (0.028)	-	-	-	-
US-C-ROW	-0.08*** (0.025)	0.16*** (0.036)	0.1*** (0.024)	0.5*** (0.035)	0.76*** (0.029)	0.31*** (0.04)	0.91*** (0.029)	0.49*** (0.04)
US-C	-0.27*** (0.024)	-0.14*** (0.036)	-0.13*** (0.024)	0.09** (0.036)	0.57*** (0.029)	-0.04 (0.04)	0.65*** (0.027)	0.05 (0.038)
C-ROW	-	-	-	-	0.2*** (0.024)	0.21*** (0.034)	0.24*** (0.026)	0.25*** (0.036)
CO	-	-	-	-	--	--	--	--
Field dummy (21)	yes	yes	yes	yes	yes	yes	yes	yes
#Obs	10127	10045	10127	10045	7583	7590	7583	7590
Adjusted R²	0.2068	0.158	0.1911	0.1208	0.3458	0.1328	0.3155	0.0992

Note: 95% confidence interval, * $p<0.1$; ** $p<0.05$; *** $p<0.01$. Standard errors are in brackets. Observations without valid address information, author information, or CiteScore value are dropped. The lower numbers of observations for regressions using Ln of dependent variables than using absolute number is due to the observations with 0 citation or 0 CiteScore.   "--" is the benchmark in the regression and "-" means that variable have no meaning in the regression.

Appendix Table B9. Regression of CiteScores and Citations of US addressed papers on percent of US addressed authors and percent of third country authors

	CiteScore	3-year citations
Precent of US addressed authors	3.42***(0.532)	9.24***(2.741)
Precent of third country addressed authors	1.13*(0.61)	8.79***(3.143)
#Authors	0.05***(0.004)	0.24***(0.02)
US-ROW	1.21***(0.312)	1.9(1.613)
US-C-ROW	2.71***(0.354)	14.74***(1.828)
US-C	0.84**(0.408)	6.14***(2.101)
Field dummy (21)	Yes	Yes
#Obs	10,142	10,441
Adjusted R²	0.1036	0.0405

Note: 95% confidence interval, * p<0.1; ** p<0.05; *** p<0.01. Standard errors are in brackets.

Appendix Table B10. Diaspora and returnee papers published on Science and Nature, 2018

	Presence count						Fractional count			
	All articles	Articles with valid name and address info	US addressed articles	US/all	Diaspora articles	Diaspora/US	US addressed author	US/all	Diaspora author	diaspora/US
Science	967	859	581	67.6%	251	43.2%	444.4	51.7%	55.6	12.5%
Nature	1,230	987	658	66.7%	280	42.6%	476.2	48.2%	57.6	12.1%
S&N	2,197	1,846	1,239	67.1%	531	42.9%	920.6	49.9%	113.2	12.3%
	All articles	Articles with valid name and address info	China addressed articles	China/all	Returnee articles	Returnee/China	China addressed author	China/all	Returnee author	returnee/China
Science	967	859	115	13.4%	87	75.65%	55.4	6.4%	15.7	28.3%
Nature	1,230	987	136	13.8%	103	75.74%	51.9	5.3%	14.8	28.6%
S&N	2,197	1,846	251	13.6%	190	75.70%	107.3	5.8%	30.5	28.5%

Note: The difference between column 1 and 2 is number of Scopus journal article records without author name and address information. By checking those papers on Science and Nature websites, we found that those papers are non-research articles (news, comments, etc.) but mis-recorded as research articles in Scopus. We exclude those papers from our analysis

Appendix C: Diaspora and returnee authors in US-China collaboration

This appendix details how we used fractional counts to create the pooled samples for testing whether the observed distribution of diaspora and returnee authors on US-China collaborations could have come from random selection from pooled samples of possible authors.

Method of fractional count calculation

We created a pool of US-addressed authors and a pool of China-addressed authors from the 2018 papers based on a fractional count of Author IDs for Table 2 analysis. The fractional count adjusts the number of authors in the pool upwards by the number of papers they wrote in 2018 and downwards by the number of co-authors on papers. For example, if a paper has one diaspora author, one US addressed non-diaspora author and one UK author, we count 1/3 for diaspora author and 2/3 for all US addressed authors, then the diaspora share of fractionated US addressed author equals one-third divided by two-thirds = 1/2. As shown in Appendix Table C2 and C3, the diaspora share of fractionated US-addressed authors is 12%. The returnee share of fractionated China-addressed authors is 8.5% in 2018.

Appendix Table C1. Proportions of types of authors on US-addressed papers, 2018

	%US-addressed and non-Chinese named author	%Diaspora author	%China-addressed author	% ROW-addressed author
USO without D	100.0%	0.0%	0.0%	0.0%
US-ROW without D	36.5%	0.0%	0.0%	63.5%
US-C-ROW without D	25.6%	0.0%	35.0%	39.4%
US-C without D	29.8%	0.0%	70.2%	0.0%
USO with D	59.2%	*40.8%*	0.0%	0.0%
US-ROW with D	43.4%	19.3%	0.0%	37.3%
US-C-ROW with D	21.9%	17.5%	34.8%	25.9%
US-C with D	18.2%	24.8%	57.0%	0.0%

Appendix Table C2. Fractional count of US-addressed papers by types of authors, 2018

	#Papers	a. Fractional count by % US-addressed and non-Chinese named author	b. Fractional count by % Diaspora author	c. Fractional count by % China-addressed author	d. Fractional count by % ROW-addressed author	e. Fractional by % all US-addressed author (a+b)	b/e
US-addressed papers	352,525	225,424 (63.9%)	30,600 (8.7%)	23,186 (6.6%)	73,316 (20.8%)	**256,024**	**12.0%**
USO without D	140,352	140,352	0	0	0	**140,352**	**0.0%**
US-ROW without D	99,040	36,190	0	0	62,850	**36,190**	**0.0%**
US-C-ROW without D	6,038	1,544	0	2,115	2,379	**1,544**	**0.0%**
US-C without D	12,170	3,624	0	8,546	0	**3,624**	**0.0%**
USO with D	*52,863*	31,315	*21,548*	0	0	**52,863**	**40.8%**
US-ROW with D	18,010	7,818	3,474	0	6,718	**11,291**	**30.8%**
US-C-ROW with D	5,293	1,159	925	1,840	1,369	**2,084**	**44.4%**
US-C with D	18,759	3,422	4,653	10,684	0	**8,075**	**57.6%**

Note: The fractional count of papers is calculated by multiplying the #papers in the first column by percent of authors on the same types of papers shown in Appendix Table C1. For example, the fractional count of USO with D by diaspora authors = 52,863 * 40.8% from Appendix Table C1 = 21,548

Appendix Table C3. Fractional count of China-addressed papers by returnee-diaspora authors, 2018

	#Papers	a. Fractional count of papers by China-addressed authors	b. Fractional count of papers by returnee authors	b/a
China-addressed papers	368,370	325,572	27,692	**8.5%**
CO	269,054	269,054	22,142	**8.2%**
C-ROW	57,056	33,332	4,022	**12.1%**
US-C-ROW with D	5,293	1,840	204	**11.1%**
US-C with D	18,759	10,684	670	**6.3%**
US-C-ROW without D	6,038	2,115	226	**10.7%**
US-C without D	12,170	8,546	428	**5.0%**

We randomly sampled 2,000 diaspora authors & 2,000 non-diaspora authors from the US addressed papers described in Appendix Table A1 and 2,000 returnee authors & 2,000 non-returnee authors from the China addressed papers described in Appendix Table A1 and A2. We retrieved English journal articles published in 2018 for 8,000 randomly chosen authors from the two (which goes beyond the papers in our initial sample) and calculated the total number of papers each wrote in 2018 and the number that were US-China collaborations. About the heterogeneity in authors' propensity to work on US-C collaborations. While some of the variation reflects the number of papers that authors wrote in 2018, as the proportion of collaborations for an author with one paper is necessarily 0% or 100%, and so on, there is greater variation in the US-China collaboration proportion of papers among authors with the same numbers of papers than would occur if each had the same propensity to work on a US-China collaboration.

Appendix Table C4. Sampled Diaspora, Non-Diaspora, Returnee, and Non-Returnee authors for Table 3 Analysis

Author type	#Sampled Author IDs	#Valid Author IDs at 2022	#English journal articles by authors in 2018
Diaspora	2,000	1,972	8,987
Non-Diaspora	2,000	1,966	9,889
Returnee	2,000	1,974	22,836
Non-Returnee	2,000	1,936	11,485

Note: 1.8% of Author IDs sampled in early 2020 became invalid in 2022 because Scopus updates Author ID over time.

Appendix D: Author career and citation network

Appendix Table D1. Estimated number of US-China dual addressed authors in 2018

	Fractional count of authors	Fractional share of US addressed author
Dual diaspora author	*7,264*	*0.461%*
Dual non-diaspora author	*1,232*	*0.078%*
Non-dual diaspora author	181,811	11.5%
Non-dual non-diaspora author	1,385,313	87.9%
	Fractional count of authors	**Fractional share of China addressed authors**
Dual returnee author	*4,273*	*0.188%*
Dual non-returnee author	*150*	*0.007%*
Non-dual returnee author	189,402	8.3%
Non-dual non-returnee author	2,084,709	91.5%
Total dual	***12,919***	**-**

Note: US-China dual addressed authors are count as part of US addressed author and part of China addressed authors based on their affiliation address, if one dual author listed one US affiliation, one China affiliation and one UK affiliation, this dual author will be count as 1/3 US addressed author and 1/3 of China addressed author. Estimated number of US-China dual addressed authors is based sampled 2,000 USC and 2,000 US-C-ROW papers per Appendix Table A1. Estimated number of non-dual US and China addressed authors is based on 12,000 sampled US and China addressed papers per Appendix Table A1 and A2.

In Appendix A4, we estimate the number of US addressed authors on 2018 US-China collaborative papers is 139,408 (authors will be count n times if their AU-IDs appeared on n papers), so dual authors' share of US addressed authors on collaborative papers = (7264+1232)/139408 = 6.1%. Similarly, the estimated number of China addressed authors on 2018 US-China collaborative papers is 230,757 (see Appendix A5) and the dual authors' share of China addressed authors on collaborative papers = (4273+150)/230757 = 1.9%.

Appendix Table D2. Pre-2018 publications of dual addressed authors

Dual authors with at least one pre-2018 publication where the author had	Pct of authors
1. A dual address	71%
2. A solo China address	89%
3. A solo US address	61%
Union of the 1-3 columns	100%
4. A solo US address and a China addressed coauthor	55%
5. A solo China address and a US addressed coauthor	63%
Union of the 4-5 columns	79%

Note: Estimated number of US-China dual addressed authors is based on the sampled 2,000 US-C and 2,000 US-C-ROW papers per Appendix Table A1.

Appendix Table D3. Percent of Pre-2018 Publications by 2018 US-Addressed and China-Addressed Authors, by Diaspora and Returnee Status

Diaspora and returnee status	% of pre-2018 publications by 2018 US-addressed authors where the author had				
	Solo-China address and no US addressed coauthors	Solo-China address and US addressed coauthors	US-China dual address	Solo-US address and no China addressed coauthors	Solo-US address and China addressed coauthors
Diaspora	8.6%	1.4%	67.4%	17.3%	2.4%
Non-D	0.1%	0.1%	86.6%	3.7%	0.1%
Returnee	68.6%	10.5%	6.7%	1.9%	2.9%
Non-R	77.6%	4.9%	0.0%	0.0%	0.0%

Note: Based on all pre-2018 publications by randomly sampled 2000 diaspora and 2000 non-diaspora authors from the 8000 sampled US-addressed papers described in Appendix Table A1, and 2000 returnee authors and 2000 non-returnee authors from the 8000 sampled China-addressed papers per Appendix Table A1 and A2. In these calculations, to focus on US-China connections, we treat papers where author had China and Rest of World addresses and no US address as "solo China" and papers where author had US and Rest-of-World address and no China address as "solo US address" and treat a paper where author had US, China, and Rest-of-World addresses as US-China dual address.

Appendix Table D4. Citation preference measures between US and China

2015 papers	Citations to 2015 papers from 2016-18 papers		
	from USO D	from USO ND	*DD Ratio*
CO	26,546	27,467	-
ROW	145,823	263,932	-
from D/from ND	*0.182*	*0.104*	*1.75*
	from CO	from ROW	*DD Ratio*
USO D	99,660	185,690	-
USO ND	136,023	516,658	-
from CO/from ROW	*0.733*	*0.359*	*2.04*
	from CO R	from CO NR	*DD Ratio*
USO	107,587	128,096	-
ROW	263,961	530,562	-
from R/from NR	*0.408*	*0.241*	*1.69*
	from USO	from ROW	*DD Ratio*
CO R	28,790	123,472	-
CO NR	25,223	217,526	-
from USO/from ROW	*1.141*	*0.568*	*2.01*

Note: To estimate the diaspora preference to CO vs ROW compared to non-diaspora papers, we randomly sampled 2,000 CO and 2,000 ROW papers published in 2015, then we retrieved 27,443 papers in 2016-2018 that cite our sampled 2015 CO papers and ROW papers. We select the 2016-18 USO papers from those 27,443 papers, and use the same name-based method to calculate the average citations per paper from 2016-18 USO diaspora/non-diaspora papers for our sampled 2015 CO and ROW papers. Multiplying number of CO papers and ROW papers in 2015 (CO: 203,602; ROW: 894,849), we estimated the number of citations from all 2016-2018 diaspora and non-diaspora USO papers to all 2015 CO and ROW papers, respectively.

To estimate the CO preference to diaspora vs non-diaspora compared to ROW, we randomly sampled 2,000 USO papers published in 2015 and use our name-based method to divide the USO papers into 2015 diaspora USO papers and non-diaspora USO papers. Then we retrieved 35,108 papers in 2016-18 that cite our sampled 2015 USO papers. We select the 2016-18 CO and ROW papers from those 35,108 papers and calculate the average citations per paper from 2016-18 CO/ROW papers for our sampled 2015 diaspora and non-diaspora USO papers. Multiplying

number of diaspora and non-diaspora USO papers in 2015 (diaspora USO: 48,629; non-diaspora USO: 147,661), we estimated the number of citations from all 2016-2018 CO and ROW papers to all 2015 diaspora and non-diaspora USO papers, respectively.

To estimate the returnee preference to USO vs ROW compared to non-returnee papers, we use the randomly sampled 2,000 USO and 2,000 ROW papers in 2015, and retrieved 47,883 papers in 2016-2018 that cite our sampled 2015 USO papers and ROW papers. We select the 2016-18 CO papers from those 47,883 papers, and use the same publication-history-based method to calculate the average citations per paper from 2016-18 CO returnee/non-returnee papers for our sampled 2015 USO and ROW papers. Multiplying number of USO papers and ROW papers in 2015 (USO: 196,290; ROW: 894,849), we estimated the number of citations from all 2016-2018 returnee and non-returnee CO papers to all 2015 USO and ROW papers, respectively.

To estimate the USO preference to returnee vs non-returnee compared to ROW, we use the randomly sampled 2,000 CO papers in 2015 and apply our publication-history-based method to divide the CO papers into 2015 returnee CO papers and non-returnee CO papers. Then we retrieved 14,669 papers in 2016-18 that cite our sampled 2015 CO papers. We select the 2016-18 USO and ROW papers from those 14,669 papers and calculate the average citations per paper from 2016-18 USO/ROW papers for our sampled 2015 returnee and non-returnee CO papers. Multiplying number of returnee and non-returnee CO papers in 2015 (returnee CO: 56,935; non-returnee CO: 146,667), we estimated the number of citations from all 2016-2018 USO and ROW papers to all 2015 returnee and non-returnee CO papers, respectively.

Appendix E: Numbers of US F1 and J1 Visas Issued per month to Chinese citizens

Appendix Table E1 shows that the numbers of F1 visas issued to Chinese citizens between January to April 2022 exceed numbers in the same months of 2021, and the monthly F1 visa issuances started declining in May 2022 after the Shanghai lockdown and the US consulate in Shanghai was closing to public in April 2022.

Appendix Table E1. Number and percentage of F1 visas issued by the US to Chinese citizens

a. #F1 visa issued to Chinese citizens by the US						
	2018	**2019**	**2020**	**2021**	**2022**	**2023**
Jan	3,094	3,001	2,918	**480**	**936**	2,356
Feb	925	664	112	**231**	**554**	1,282
March	1,359	1,443	135	**304**	**585**	1,775
Apr	3,700	4,281	17	**321**	**1,550**	-
May	18,271	20,871	7	***23,066***	***8,570***	-
June	29,555	34,001	8	33,896	18,860	-
July	24,666	21,781	145	21,163	15,090	-
Aug	7,118	824	363	8,699	4,909	-
Sept	1,351	1,231	268	1,355	980	-
Oct	922	1,150	217	1,604	948	-
Nov	1,948	2,335	195	4,349	1,983	-
Dec	4,774	7,002	468	3,963	2,546	-
Total	97,683	98,584	4,853	99,431	57,511	5,413
b. %F1 visa issued to Chinese citizens by the US						
	2018	**2019**	**2020**	**2021**	**2022**	**2023**
Jan	17%	17%	19%	**5%**	**7%**	16%
Feb	13%	10%	2%	**6%**	**8%**	19%
March	16%	18%	4%	**6%**	**7%**	18%
Apr	27%	30%	17%	**4%**	**12%**	-
May	42%	45%	3%	***59%***	***23%***	-
June	38%	43%	1%	43%	20%	-
July	28%	25%	1%	22%	16%	-
Aug	17%	2%	2%	14%	9%	-
Sept	14%	15%	4%	11%	10%	-
Oct	13%	15%	3%	13%	11%	-
Nov	14%	17%	2%	16%	8%	-
Dec	18%	23%	2%	11%	6%	-
Total	27%	27%	5%	25%	14%	17%

Note: The numbers of F1 visa issuances to Chinese nationality are from the Monthly Nonimmigrant Visa Issuance Statistics by U.S. Bureau of Consular Affairs (https://travel.state.gov/content/travel/en/legal/visa-law0/visa-statistics/nonimmigrant-visa-statistics/monthly-nonimmigrant-visa-issuances.html).

Appendix Table E2. Number and percentage of J1 visas issued by the US to Chinese citizens

	2018	2019	2020	2021	2022	2023
a. #J1 visa issued to Chinese citizens by the US						
Jan	2,021	2,114	1,421	33	300	605
Feb	1,251	1,218	105	70	318	499
March	2,476	2,325	71	80	323	515
Apr	3,561	4,722	18	82	284	-
May	4,402	4,503	11	133	416	-
June	4,526	5,732	8	458	934	-
July	5,746	4,389	12	966	989	-
Aug	5,091	4,713	45	647	881	-
Sept	2,812	2,836	60	419	601	-
Oct	2,365	2,113	54	334	446	-
Nov	2,298	1,809	55	728	679	-
Dec	2,560	2,693	65	726	678	-
Total	39,109	39,167	1,925	4,676	6,849	1,619
b. %J1 visa issued to Chinese citizens by the US						
	2018	2019	2020	2021	2022	2023
Jan	14%	15%	10%	1%	3%	4%
Feb	10%	9%	1%	2%	2%	3%
March	8%	7%	1%	1%	1%	1%
Apr	7%	8%	16%	1%	1%	-
May	7%	7%	2%	1%	1%	-
June	11%	14%	1%	2%	3%	-
July	14%	11%	1%	4%	3%	-
Aug	21%	19%	1%	4%	4%	-
Sept	23%	22%	3%	5%	5%	-
Oct	15%	14%	3%	4%	3%	-
Nov	11%	9%	3%	4%	3%	-
Dec	13%	13%	1%	4%	4%	-
Total	11%	11%	3%	3%	2%	2%

Note: The numbers of J1 visa issuances to Chinese nationality are from the Monthly Nonimmigrant Visa Issuance Statistics by U.S. Bureau of Consular Affairs (https://travel.state.gov/content/travel/en/legal/visa-law0/visa-statistics/nonimmigrant-visa-statistics/monthly-nonimmigrant-visa-issuances.html).