创业与管理学院

School of Entrepreneurship and Management

SHANGHAITECH SEM WORKING PAPER SERIES
No. 2022-002

Modernizing the Chinese Way: Effect of Big-Science Research Infrastructure on China's Indigenous Science Growth

Xiyi Yang

School of Entrepreneurship and Management, ShanghaiTech University

Xiaoyu Zhou

School of Entrepreneurship and Management, ShanghaiTech University

Cong Cao

Nottingham University Business School China, Univerisity of Nottingham Ningbo

December 2, 2021

https://ssrn.com/abstract=4057905

School of Entrepreneurship and Management

ShanghaiTech University

http://sem.shanghaitech.edu.cn

Modernizing the Chinese Way: Effect of Big-Science Research Infrastructure on China's Indigenous Science Growth

Xiyi Yang[1], Xiaoyu Zhou[1]*, Cong Cao[2]*

1 School of Entrepreneurship and Management, ShanghaiTech University; 393 Huaxia Road, Shanghai, 201210, China

2 Nottingham University Business School China, University of Nottingham Ningbo China; 199 East Taikang Road, Ningbo, 315100, China

*Corresponding authot

Abstract

Since the mid-twentieth century, big-science research infrastructure (RI) has become an indispensable component for scientists to explore fundamental scientific breakthroughs. However, due to its extremely expensive constructional and operational cost, whether and how RIs can effectively and efficiently boost the development of science and technology remains a controversial debate in emerging countries. To address this concern, we focus our investigation on the impact of a major RI, the Shanghai Synchrotron Radiation Facility (SSRF), on China's production of science. We assemble a unique dataset of China's scientific publications and apply a difference-in-difference design to assess the casual relationship between the SSRF and Chinese scientists' indigenous publications. We find that for affected scientific disciplines, the establishment of SSRF caused 43.5% and 18% increases in the number and percentage of high-quality indigenous publications, respectively, and a 21% increase in the overall impact factor of indigenous publications. Combined, our findings suggest that SSRF has led to a quantitative and qualitative growth of China's indigenous science. Such findings have important implications for public policy design and a better understanding of the current global technological competition. The study also addresses the concern that China lacks originality and indigenous innovation capability.

Keywords: big-science, research infrastructure, indigenous science, difference-in-difference, emerging markets.

Main Text
Since the mid-twentieth century, an increasing percentage of scientific breakthroughs are achieved utilizing big-science research infrastructures (RIs) (1). The broad experimental opportunities offered by synchrotron radiation, neutron spallation, free-electron laser, among others, have been essential for the development of disciplines such as nanotechnology, proteomics, and drug development (2). Accordingly, countries are investing huge amounts of resources in constructing RIs to enhance their global technological competitiveness. Initially, the United States, Japan, and European Union were the key players in hosting big-science RIs. Major facilities include Stanford Linear Accelerator Center (SLAC) in the US, CERN in Europe and KEK in Japan for research on particle physics, as well as major synchrotron light sources such as Advanced Photon Source (APS) in the US, Super Photon Ring-8 GeV (SPRING-8) in Japan, and European Synchrotron Radiation Facility (ESRF) in France.

Big-science RIs have drawn increasing scholarly attention. Studies find that RIs have contributed to the significant improvement of the quality of natural sciences research in countries hosting the RIs and their national competitiveness (3-5). Increasing international collaboration among top natural scientists from all over the world is also attributed to the rising number of big-science RIs (6). There also have been discussions on scientific and socio-economic merits of such facilities against the extremely huge costs incurred to finance, organize, and operate the RIs (7).

Yet, extant studies are primarily grounded in Western developed countries (5, 8) and focusing on productivity outcomes such as publication and citation counts as well as prestigious awards such as Nobel Prizes (2, 9). In general, establishing such a causal relationship is difficult due to various endogeneity issues. Hence, except for Helmers and Overman's work on the British Diamond Light Source, findings from the effect of big-science RIs are mostly draw from correlations (1) instead of causal inferences. Of particular interests are the following questions: Can RIs increase the productivity of their users or are more productive researchers have better access to the RIs? Do RIs promote regional innovation or are they more likely to be located in more prosperous regions? Finally, can RIs truly enhance nations' production of science or is it due to other confounding factors such as human capital accumulation? These questions are especially acute as there is also very little knowledge about whether investments in the RIs can improve emerging countries' indigenous innovative capability and accelerate the process of their catching-up with developed countries.

We try to answer some of these questions by investigating the causal influence of Shanghai Synchrotron Radiation Facility (SSRF), one of the big-science RIs in China, on the country's indigenous science growth. In its early stage of development, China could not afford to build the RIs so that domestic researchers had to travel abroad or collaborate with foreign teams that had access to overseas RIs. With its rapid transformation from a peripheral player to one of the world's most active science systems, evident by its sharp and sustained increases in R&D expenditure (10, 11), the rising number of scientific publications (12, 13) and patents (14), China since the 1980s has initiated a series of megaprojects involving substantial government investments in constructing big-science RIs (15). Indeed, generous government budgets have now made China home to a few of the world's largest RIs, such as the Five-hundred-meter Aperture Spherical Telescope (FAST) and the X-Ray Free-Electron Laser Facility (appendix A).

Our study examines the effect of SSRF from the perspective of indigenous science, that is, scientific knowledge produced by domestic scholars, or those working in China-based institutions, without the aid of foreign collaborators. Amid the ongoing US-China trade war and intensified global technology competition, the Chinese government has increasingly recognized the importance of self-strengthening in science and indigenous innovation. Given that the big science RIs in China are almost fully funded by the government, it is important to examine whether these RIs can help fulfill the national agenda of promoting and strengthening indigenous science. SSRF is the first and only 3rd generation synchrotron facility in Mainland China. It is a circular particle accelerator that produces beams of X-rays, infrared light and ultraviolet light, so that it is vital to study small objects, such as molecules and atoms, whose visualization requires light with shorter wavelengths than what is available in microscopes (appendix B). At the time of construction, SSRF was China's costliest single research infrastructure (16). Since its operation in 2009, over 26,000 scientists have conducted experiments there and produced more than 6,000 academic publications, including 109 in *Nature*, *Science*, or *Cell*.

To quantify the causal influence of SSRF, we hinge our identification strategy on the fact that only certain scientific disciplines rely on synchrotron light for experimentation and hence are affected by the construction of SSRF. These affected disciplines include condensed matter physics, atomic and molecular physics, chemistry, materials science, life science, medicine, and so on and can be regarded as the 'treatment group'. On the other hand, the development of many other disciplines, such as math, computer science, clinical medicine, psychology, and social sciences, does not depend on the availability of synchrotron light and hence is not directly affected by the establishment of SSRF. These disciplines can be used as the 'control group'. We then treat the opening of the SSRF in 2009 as a quasi-natural experiment and use a difference-in-differences (DID) estimation model to quantify its impact on the production of science of affected disciplines.

Our outcome variables measure the production of high-quality indigenous science by various disciplines and include (i) the number of high-quality indigenous publications by China-based scholars without foreign coauthors (*Numb_highquality_indigenous*); (ii) the percentage of high-quality indigenous publications to total publications (*Percent_highquality_indigenous*); and (iii) the average impact factor (IF) of indigenous publications (*Mean_IF_indigenous*). Scientific publication is a particularly suitable measure of scientific knowledge production in the context of our analysis given that research conducted at SSRF can be regarded as 'cutting edge', which makes it likely to result in findings publishable in leading scientific journals. High-quality publications refer to those that appear in journals with an IF larger than 10. We acknowledge that journal IF is an imperfect measure of publication quality, though it highly correlates with other impact or quality measures such as relative citation rates and h-index. We focus on high-quality publications for the reason that, although China's total publications have been skyrocketing recently, there is concern that a large proportion of them lacks originality or value to the scientific community(17).

To control for other confounding factors, we have included several control variables in our estimations. *NSFC fund* is the logarithm of the total amount of funding each broad scientific discipline receives from the National Natural Science Foundation of China (NSFC) in each year. This controls for the effect of government funding on discipline's production of science. *National key discipline* is a dummy variable that equals 1 if the concerned discipline belongs to the National Basic Research Program (also known as the '973' Program) in any given year and hence receives more support from the government. Finally, we have controlled for *Overseas returnee*, the logarithm of the total number of returnees that received education in foreign countries in each year, to account for influence of returnees' human capital. Returnees include those studying and/or carrying research overseas for at least one year, most of which have formal higher education credentials at undergraduate, postgraduate, doctoral, to post-doctoral levels.

Our sample consists of 1,830,731 English journal publications with at least one author based in China (including Hong Kong, Macau and Taiwan) from the Web of Science in the period of 1998-2015. For each publication, we have information on its quality (such as IF and citation), whether it is produced solely by domestic scholars without foreign coauthors, and its corresponding scientific discipline. According to the Web of Science, our sample fall into 252 disciplines. With the help of the managers of the beamline stations at SSRF, we categorized the disciplines based on SSRF's relevance to them. 117 disciplines that have carried out experiments at SSRF are considered affected by the facility and hence constitute our 'treatment group'. The remaining 135 disciplines constitute the 'control group' (see Methods section for full details on data and methods).

Empirically, the main challenge in establishing a causal link between the construction of SSRF and the production of science, and particularly high-quality publications by all China-based scholars, is the potential endogeneity of SSRF. As mentioned, our main strategy is to exploit the fact that only certain scientific disciplines would use synchrotron light for experimentation and hence are affected by the construction of SSRF. This cross-discipline heterogeneity allows us to use the following DID equation for estimating the causal influence of SSRF:

$$y_{i,t+1} = \beta_1 Treatment_i * After_t + \beta_2 After_t + \beta_3 Control + \delta_i + \delta_t + \varepsilon_{i,t} \quad (1)$$

$y_{i,t+1}$ are measures of indigenous science for discipline *i* in year *t+1* (i.e. *Numb_highquality_indigenous, Percent_highquality_indigenous, Mean_IF_indigenous*). $Treatment_i$ is a dummy variable that equals 1 if discipline *i* utilizes experiments at SSRF, and it equals 0 if otherwise. Although there could be potential spillover from those affected disciplines to ones that do not use SSRF[1], our estimated coefficient may only under-estimate the effect of SSRF on affected disciplines. $After_t$ is a dummy variable that equals 1 after the year 2009 when SSRF was first opened to users, and it equals 0 before or in 2009. *Control* are control variables that include *NSFC fund*, *National key discipline*, and *Overseas returnee*. The model also includes discipline fixed effects δ_i and year fixed effects δ_t. Our interested coefficient is β_1 of the interaction term $Treatment_i * After_t$, which captures the effect of SSRF on affected disciplines compared to disciplines that do not utilize SSRF, since the facility's opening in 2009.

Results

We first look at the data pattern by aggregating the raw data for the two groups: 'treatment' disciplines that benefit from SSRF for experiments and 'control' disciplines that do not rely on synchrotron light for experiments. As we are particularly interested in the quality of indigenous science and the degree to which domestic scholars produce high-quality research on their own, we first focus on the quality of indigenous publications (*Mean_IF_indigenous*) and the percentage of high-quality indigenous publications to total publications (*Percent_highquality_indigenous*). Figure 1 illustrates the time trend of the concerned variables for the two groups. The figures clearly show that before SSRF was put into use in 2009, there is little systematic difference between the two groups. Even though the treatment group's indigenous publications seem to always have a higher mean IF, the difference between the treatment group and the control group is 'parallel' in the sense that the difference does not change much with time. However, since 2009, the disciplines that rely on SSRF for experimentation have experienced higher growth in both *Mean_IF_indigenous* and *Percent_highquality_indigenous*, compared to the disciplines that are not associated with SSRF, and the differences between the two groups hence become more substantial.

We further quantify the above-mentioned effect by estimating the DID model in equation (1) and Table 2 reports the estimation results. As shown in Model (1), the coefficient β_1 is 0.435 and significant at the $p<0.01$ level. It indicates that the establishment of SSRF has caused a 43.5% increase in the number of indigenous publications that appeared in high-quality journals (*Numb_highquality_indigenous*) for the affected disciplines, compared to the control disciplines that do not utilize SSRF. Model (2) shows that the estimated β_1 is 0.177 with $p<0.01$. That is, for disciplines that utilize SSRF, the percentage of indigenous publications that appeared in high-quality journals (*Percent_highquality_indigenous*) is increased by 17.7%. These two results suggest that SSRF has led to a significant and substantial increase in China's high-quality indigenous publications. Finally, Model (3) explores the effect of SSRF on the overall quality of indigenous publications and shows that SSRF leads to a 0.208 increase in the impact factor of indigenous publications for affected disciplines (*Mean_IF_indigenous*), which corresponds to a 14% increase over the sample mean. Overall, the results show that the establishment of SSRF

[1] For instance, the development of pediatrics (a control discipline) may benefit from discoveries from hematology or virology (two treatment disciplines) that utilizes synchrotron facility. In this case, the DID model would under-estimate the treatment effect.

has significantly increased the production of high-quality and impactful indigenous science by domestic scholars. The results remain robust when we performed various robustness tests such as placebo treatment test, randomization of treatment and control group and so on (appendix C).

Our final set of results explored the discipline-level heterogeneity of the SSRF effect. We first split the 117 treatment disciplines into six broad categories: Physics, Chemistry, Material Science, Engineering, Life and Medical Science, and Environmental Science. We then re-estimated our baseline DID model for six times with each broad category being the treatment group. The control group is the same as that in the baseline model (the 135 disciplines that do not use synchrotron facility for experiments). The results are summarized in Table 2 with a focus on the overall quality of indigenous publications (*Mean_IF_indigenous*). Chemistry seems to benefit most from SSRF while Material Science, Engineering and Environmental Science all have significantly increased their quality of indigenous publications since the opening of SSRF. However, SSRF does not seem to have a significant effect on Physics and Life and Medical Science, at least in our sample period of 1998-2015 (i.e. 6 years after SSRF was put into use). The insignificant result on Life and Medical Science is quite counter-intuitive, since this field has been allocated much of the beamtimes at SSRF, and future research may further investigate the resource allocation efficiency of RIs.

Discussion

Our research builds on and extends previous work in several ways. Using a novel DID design to quantify the causal impact of a major RI in China, this study addresses the question that whether big-science RIs can improve the quality of scientific research in emerging countries. Our analysis shows that SSRF has significantly enhanced the indigenous scientific publications by all China-based scholars. Such findings have important implications for public policy design and a better understanding of the current global technological competition. The study also addresses the concern that China lacks originality and indigenous innovation capability (18). Indeed, the recent US-China tension essentially is one on technology with denied access to many state-of-the-art technologies and high-end products such as microchips that severely crippled China's ambition to become an innovation-oriented nation and world's leading scientific power. The construction of big-science RIs may provide a domestically accessible platform that better enables and facilitates domestic scholars to conduct novel research to shift the knowledge frontier of many scientific disciplines.

Several challenging questions arising from our discoveries require further research. Above all, what is the delicate balance between self-strengthening in science and international collaboration? Collaboration is one of the most remarkable characteristics of contemporary research and international collaboration often results in greater visibility and higher impact (19, 20). In fact, international co-publication has played a considerable role in the rise of China's scientific publications in recent decades (21). However, with the current anti-globalization headwind and intensified global technological competition, the Chinese government has taken self-strengthening in science and technology as the strategic underpinning for national development. Our study suggests that government fully funded big-science RIs can increase China's indigenous science. But we do not mean to exaggerate that having such RIs has to be at the expense of international collaboration as China was and still is a latecomer in scientific research and even in the construction and utilization of big-science RIs. Quantitatively, we know that since the opening of SSRF, the share of publications that use other similar RIs has declined. How will this shift affect the social networks of individual Chinese scientists? What are the implications for the long-term development of China and the global science community? These future research directions promise further contributions to the literature on the economics of science and science policy-making.

Finally, in this study, we measure the impact of big-science RIs through the lens of scientific publications. Potential byproducts created by the establishment of a basic research facility such as patents or the formation of scientific networks are not considered. However, these byproducts are certainly important when indigenous innovation and national competitiveness are concerned. Moreover, the technological and industrial developments required in building RIs such as sophisticated detectors and data processing and analysis systems could generate substantial payoffs over the long run. During an interview, the manager of SSRF mentioned to us that "companies can also apply for experimenting opportunities at SSRF and many scientific discoveries from SSRF have turned into patentable technologies." What roles can big-science RIs play in the process of transforming scientific discoveries into productivity gains? What's the appropriate form of cooperation between SSRF, scientists, and firms? We encourage future research to collect more nuanced data to study the spillovers of big-science RIs.

Materials and Methods

Our initial sample included 1,908,610 English journal articles with at least one author based in China (including Hong Kong, Macau and Taiwan) from 1970-2015. The data on publications came from the Web of Science, one of the world's largest databases of academic publications. The data on the journal impact factor (IF) came from Journal Citation Reports which are also maintained by the Web of Science. We collected the information on the national natural science funds, national key discipline, and the number of overseas returnees from the China Science and Technology Statistic Yearbooks and the website of the Ministry of Education of the People's Republic of China.

Acknowledgments

We are grateful to Douglas Almond, Yi Chen, Hanming Fang, Xiaoou Liu, Ronald Rousseau, Wei Si, Richard Suttmeier for comments. We thank Haolin Shao and Yang Lu for their research assistance. This research is supported by the National Natural Science Foundation of China grant 71774091 and 72102145.

References

1. OECD, Report on the impacts of large research infrastructures on economic innovation and on society: Case studies at CERN. (2014).
2. O. Hallonsten, How expensive is Big Science? Consequences of using simple publication counts in performance assessment of large scientific facilities. *Scientometrics* **100**, 483-496 (2014).
3. S. Traweek, *Beamtimes and Lifetimes: The World of High Energy Physicists* (Harvard University Press, 1992).
4. M. Low, *Science and the Building of a New Japan* (Springer, 2005).
5. K. C. Cramer, O. Hallonsten, I. K. Bolliger, A. Griffiths, "Big Science and Research Infrastructures in Europe" in Big Science and Research Infrastructures in Europe: History and current trends. (Edward Elgar Publishing, 2020), https://doi.org/10.4337/9781839100017.00007.
6. S. Porter, J. Wastl, D. Hook (2019) Japanese Collaboration in the Global Research Landscape. (Digital Science).
7. H. P. Beck, P. Charitos, *The Economics of Big Science: Essays by Leading Scientists and Policymakers* (Springer Nature, 2021).
8. C. Helmers, H. G. Overman, My Precious! The Location and Diffusion of Scientific Research: Evidence from the Synchrotron Diamond Light Source. *The Economic Journal* **127**, 2006-2040 (2017).
9. R. Heidler, O. Hallonsten, Qualifying the performance evaluation of Big Science beyond productivity, impact and costs. *Scientometrics* **104**, 295-312 (2015).

10. W. Huang, Advancing basic research towards making China a world leader in science and technology. *Natl Sci Rev* **5**, 126-128 (2018).
11. R. P. Suttmeier, Inventing the future in Chinese labs: How does China do science today?
12. Y. Xie, C. Zhang, Q. Lai, China's rise as a major contributor to science and technology. *Proc Natl Acad Sci U S A* **111**, 9437-9442 (2014).
13. Q. Xie, R. B. Freeman, Bigger Than You Thought: China's Contribution to Scientific Publications and Its Impact on the Global Economy. *China & World Economy* **27**, 1-27 (2019).
14. B. Guarino, E. Rauhala, W. Wan (2018) China increasingly challenges American dominance of science. in *The Washington Post* (Fred Ryan, Washington, D.C.).
15. Y. Wang, Y. Bai, Developing Mega-science Facility to Lead the Innovation Globally. *Management World* **5**, 172-188 (2020).
16. D. Cyranoski, China joins world-class synchrotron club. *Nature* **459**, 16-17 (2009).
17. F. Huang (2018) Low quality studies belie hype about research boom in China. in *Sci Am* (Springer Nature).
18. R. M. Abrami, W. C. Kirby, F. W. McFarlan, Why China can't innovate. *Harv Bus Rev* **92**, 107-111 (2014).
19. J. Adams, The rise of research networks. *Nature* **490**, 335-336 (2012).
20. A. Witze, Research gets increasingly international. *Nature* 10.1038/nature.2016.19198 (2016).
21. C. Cao, J. Baas, C. S. Wagner, K. Jonkers, Returning scientists and the emergence of China's science system. *Science and Public Policy* **47**, 172-183 (2020).

Figures and Tables

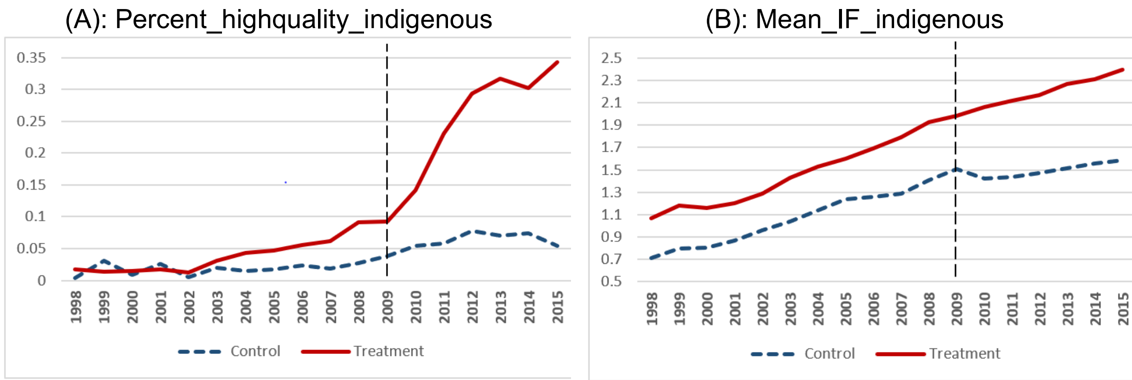


Figure 1. Time trend of indigenous publications of the treated and control disciplines (1998-2015). The figure plots the time trend of indigenous publications for the treatment group (disciplines that utilize SSRF for experimentation) and control group (disciplines that are not affected by the opening of SSRF), by taking average of the respective raw data. It shows that the two groups have similar ‘pre-trend’ before the opening of SSRF in 2009 when either the percentage of high-quality indigenous publications (Panel A) or the mean IF of indigenous publications (Panel B) is concerned.

Table 1. Effect of SSRF on China's Indigenous Scientific Publications.

	(1)	(2)	(3)
	Numb_highquality _indigenous	Percent_highquality _indigenous	Mean_IF _indigenous
Treatment*After	0.435***	0.177**	0.208***
	(0.120)	(0.077)	(0.058)
After	-0.584*	-0.176	-1.036***
	(0.331)	(0.294)	(0.332)
Controls	Yes	Yes	Yes
Discipline FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Obs	4047	3990	3659
R square	.1262	.0582	.4902

The table reports estimation results of the baseline DID model for measures of discipline's indigenous scientific output. Standard errors are clustered at the discipline level, and reported in parentheses. ***, **, * denotes significance at the 1, 5 and 10% levels respectively.

Table 2. Heterogeneous Effect of SSRF across Broad Disciplines.

DV=Mean_IF_indigenous	(1)	(2)	(3)	(4)	(5)	(6)
Physics*After	0.101					
	(0.092)					
Chemistry*After		0.748***				
		(0.167)				
Material Science*After			0.468***			
			(0.155)			
Engineering*After				0.388**		
				(0.156)		
Life and Medical Science*After					0.017	
					(0.060)	
Environmental Science*After						0.215**
						(0.107)
After	-0.137	-0.130	-0.216	-0.267*	-0.104	-0.155
	(0.132)	(0.137)	(0.142)	(0.138)	(0.122)	(0.137)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Discipline FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Obs	1794	1746	1762	1835	2302	1744
R square	.3961	.4664	.4358	.4190	.4206	.4098

The table reports estimation results of the baseline DID model for six broad scientific disciplines. The dependent variable is the mean IF of indigenous publications. Standard errors are clustered at the discipline level, and reported in parentheses. ***, **, * denotes significance at the 1, 5 and 10% levels respectively.

Appendix A. Institutional Background of Big-science Research Infrastructures (RIs) in China

Big-Science RIs and Their Objectives in China

The concept of ‘big-science’ was first introduced in a 1961 *Science* article to describe large-scale research projects that utilize large infrastructures and usually involve inter-disciplinary collaboration to address grand challenges of human society (1). Since then, a new trend has emerged in the development of science and technology that the discoveries in many scientific fields, especially frontier breakthroughs, are inseparable from big-science RIs. For instance, the Brookhaven National Laboratory in the United States, with such advanced RIs as synchrotron radiation sources, imaging facilities, relativistic heavy-ion colliders, and free-electron lasers, has spawned at least seven Nobel Prizes since 1947. Similarly, high-energy physics and cosmology discoveries have brought numeral Nobel Prizes to the European Organization for Nuclear Research (CERN) in Geneva, where some of the world’s largest and most complex RIs are located.

China began to lay out its own big-science RIs in as early as the 1980s (see Table S1 for a summary of major RIs in China) (2). In 1988, in celebrating the successful construction of the Beijing Electron Positron Collider (BEPC), Deng Xiaoping, then China’s paramount leader and a supporter of the BEPC project, made an important remark that “China must develop its own advanced technology and take its place in the global arena of high technology” (3). This remark reveals three critical objectives of big-science RIs in China in addition to their political manifestation. First, RIs function as an enabler for Chinese scientists to make important breakthroughs in scientific frontiers. This characteristic is also resonated by former President Jiang Zemin in his interview with *Science* Magazine in 2000: “The construction and operation of these (big-science) facilities have enhanced the capabilities of China in scientific research and the exploration of the unknown world” (4).

In addition to scientific significance, China’s RIs are designed to serve the nation’s developmental strategy (5). With increasing engagement in the global technological competition, China urgently needs to enhance its basic research capability to provide a basis for the development of new technologies. Particularly for some bottleneck technologies, China needs to reduce its reliance on foreign suppliers. To this end, some large RIs, such as the National Center for Protein Science and the Micro-Satellite Center, respond to the nation’s inspiration of developing strategic high-tech industries such as the biopharmaceutical and navigating systems. A third objective of big-science RIs in China is to boost international collaboration. Both the construction and implementation of RIs involve substantial resources and knowledge transfer from other countries. The Daya Bay Reactor Neutrino Experiment is a good example to illustrate how a China-based RI assembled a multinational particle physics research team in studying neutrinos. The multinational collaboration includes researchers from China, Chile, the United States, Russia, and the Czech Republic. The Chinese director, Yifang Wang from the Chinese Academy of Sciences (CAS), gave high praise to this international collaboration: “The collaboration is truly international, and lessons we learned here are invaluable” (6). More recently, taking advantage of the Belt and Road Initiative since 2014, China has increasingly collaborated with Asian and European countries in big-science RI projects such as the Earth data-sharing platform (7).

Of course, China’s emphasis on big-science RIs may not be completely without practical considerations. For example, developing large research facilities could bring about other ancillary benefits—the increased knowledge of the large-scale mobilization of resources, the organization of advanced technologies, and the administration of a large number of active scientists—that China needs to become a modern nation (8).

Big-Science RIs and Indigenous Science Growth in China

After about 30 years' intensive investment, China has achieved remarkable progress in science and technology at least in the quantitative sense (9, 10). To further enhance the country's global competitiveness and ensure control over core technological areas, the Chinese central government starts to emphasize the importance of strengthening indigenous innovation capability (*zizhu chuangxin nengli*, in Chinese) to showcase the nation's self-reliance in science (11). To this end, China needs more original breakthroughs in basic research as there is constant criticism concerning the low quality of Chinese scientific publications (12). The criticism centers around two main issues. First, although the number of China's English-language publications has been skyrocketing, a large proportion of them are published in poor-quality journals with low impact factors. Second, China's high-quality publications are often jointly done with coauthors from Western developed countries, implying that China's own scientists may lack certain indigenous innovation capacity.

Yet whether the construction of big-science RIs can help counter the criticism and improve China's indigenous innovation is an empirical question to be answered. In this study, we aim to fill this gap by studying the causal impact of the Shanghai Synchrotron Radiation Facility (SSRF) on China's production of indigenous science, measured by English publications by China-based scientists in journals catalogued by the Web of Science.

Appendix B. Primer on SSRF

SSRF is the first and only 3rd generation synchrotron facility in Mainland China. Synchrotron facilities are circular particle accelerators that produce beams of x-rays, infrared light and ultraviolet light. The facility usually consists of a large ring-shaped tube into which charged particles are fired from accelerators and in which they are accelerated further. The ring is enclosed by magnets that keep the particles in the tube 'on orbit'. The continuous movement of the electrons results in electromagnetic waves, so-called synchrotron radiation. This radiation is then captured in beamlines which are used for scientific experiments at beamline stations (see Figure S1).

Synchrotron beamline is particularly useful to study small objects, such as molecules and atoms, whose visualization requires light with shorter wavelengths than what is available in microscopes. Therefore, since its first discovery in a General Electric lab in 1947 (13), it has become indispensable for scientific breakthroughs in areas such as condensed matter physics, atomic and molecular physics, chemistry, materials science, and life science. Yet the construction of synchrotron facilities requires huge financial investment and when it opened, SSRF was China's costliest single research infrastructure (14).

The decision to construct SSRF in China can be traced back to 1993 when three honorific members of the CAS, Shouxian Fang, Dazhao Ding and Dingchang Xian, suggested to the central government that "China needs a 3rd generation synchrotron facility" (15). In 1995, the Shanghai Municipal Government and the CAS agreed to co-invest in the project and began a feasibility study. Their proposal was however disapproved by the central government in 2001 due to concerns for insufficient demand for the facility, thus leading to the necessity to have a second-phase feasibility study. Until January 2004, the State Council formally approved the project and SSRF started its construction in December 2004 and opened to the public in May 2009.
Since then the facility has steadily operated for more than ten years. Over 26,000 scientists have conducted experiments there and produced more than 6,000 academic publications, including 109 in *Nature*, *Science*, or *Cell* (these statistics are self-reported from managers at SSRF, as of May 2020). Our conversations with scientists conducting experiments at SSRF confirmed its

indispensability: they regard SSRF as crucial for their research and report that before its existence, they either had to use similar synchrotron facilities abroad, for example, the SPring-8 in Japan, or they could not pursue their specific lines of research. Furthermore, some of the scientists mentioned that applying for beam times at overseas facilities is much more difficult than at SSRF. 'Non-star' scientists either had to work with a local team or would have little chance to be accepted. Hence, we expect the establishment of SSRF to substantially lower the barrier for Chinese domestic scientists to conduct novel experiments that increase the indigenous science growth of related disciplines.

Appendix C: Robustness Checks and Additional Results

C1. Verifying the parallel trend assumption

The validity of our baseline DID regression results depends on whether the treatment and control disciplines have similar time trends before SSRF's opening in 2009. To verify this assumption, we replace $Treatment_i*After_t$ in the baseline model with a set of dummy variables $Treatment_i*After_{t\,k}$ that indicates the kth year before or after SSRF was put into service. We then test whether the coefficients $\beta_{1\,k}$ are statistically equal to 0 for k<0 and statistically larger than 0 for k>0. The results are plotted in Figure S2. We focus on the percentage of high-quality publications and the mean IF of indigenous publications of each discipline. The horizontal axis measures the number of years before or after SSRF's operation, and the vertical axis measures the estimated $\beta_{1\,k}$ in regressions with dependent variable equals *Percent_highquality_indigenous* (Panel (A)) or *Mean_IF_indigenous* (Panel (B)). Both panels show similar trends for treated and control disciplines before SSRF's operation: the estimated coefficients $\beta_{1\,k}$ (k<0) are close to zero and almost all insignificant, confirming that the estimates in Table 1 are not affected by unobserved omitted variables. Only after SSRF's operation have the estimated coefficients $\beta_{1\,k}$ (k>0) become significantly larger than 0.

C2. Test with placebo treatments prior to the real SSRF shock

To further verify that SSRF started to impact China's indigenous science only after 2009, we have conducted a robustness test with placebo treatments prior to the real SSRF shock. Specifically, we created a set of placebos shocks before 2009. For instance, '1 year' indicates that the placebo shock dummy equals 1 for the one year prior to the real opening of SSRF in 2009, and '5 years' indicates that the placebo shock dummy equals 1 for the five year prior to the real opening of SSRF. We then re-estimate our baseline model by replacing the original $After_t$ dummy, which equals 1 after 2009, with the placebo dummy. The results are reported in Table S2 and the estimated β_1 is almost all insignificant or negative and significant, confirming that the positive effect of SSRF indeed occurs since 2009.

C3. Falsification test with randomized treatment and control group

To verify that the effect of SSRF is unlikely driven by unobservable characteristic of the affected disciplines, we have run a falsification test with randomized treatment and control disciplines. If underlying characteristics of the disciplines drive our results, we would expect to find similar results using the randomized treatment group. In the randomization test, we use a uniform distribution to generate 1,000 random placebo treatment groups. We then estimate our baseline model for 1,000 times using the corresponding randomized treatment groups. Results are reported in Table S3. When either *Percent_highquality_indigenous* or *Mean_IF_indigenous* is concerned, the coefficients based on the random data are close to zero and differ statistically and economically from the coefficients estimated using the actual data. Thus, these results provide additional evidence ruling out the alternative explanation of unobservable discipline characteristics driving our results. In another robustness test, we have created some placebo treatment disciplines in the areas of economics, management, and social sciences. Th estimations based on the placebo treatment also suggest that SSRF have no significantly positive effect on the scientific publications of these disciplines.

C4. Account for disciplinary heterogeneity in initial domain of indigenous scholars

Finally, if SSRF is particularly useful for Chinese's indigenous science growth by enabling domestic scholars to conduct cutting-edge experiments, we expect the above-mentioned effect to be strengthened for disciplines that were initially dominated by domestic scholars compared to disciplines that already involved much international collaboration. We test such conjecture by adding two more interaction terms in the DID model. First, we include a triple interaction between $Treatment_i$, $After_t$ and $Domestic\ instit\ share_{2009}$, which is the share of domestic author institutions to total author institutions for all the publications of discipline *i* in the year 2009. Second, we include the interaction term between $Domestic\ instit\ share_{2009}$ and $After_t$. The effect from the interaction term between $Treatment_i$ and $Domestic\ instit\ share_{2009}$, as well as $Domestic\ instit\ share_{2009}$ itself have been controlled given that the model includes δ_i which are discipline fixed effects. The estimation results are reported in Table S4. Indeed, for all the three measurements of discipline's indigenous publications, the triple interaction term is positive and significant, confirming that disciplines that were initially populated with domestic scholars benefit more from SSRF in increasing high-quality indigenous publications.

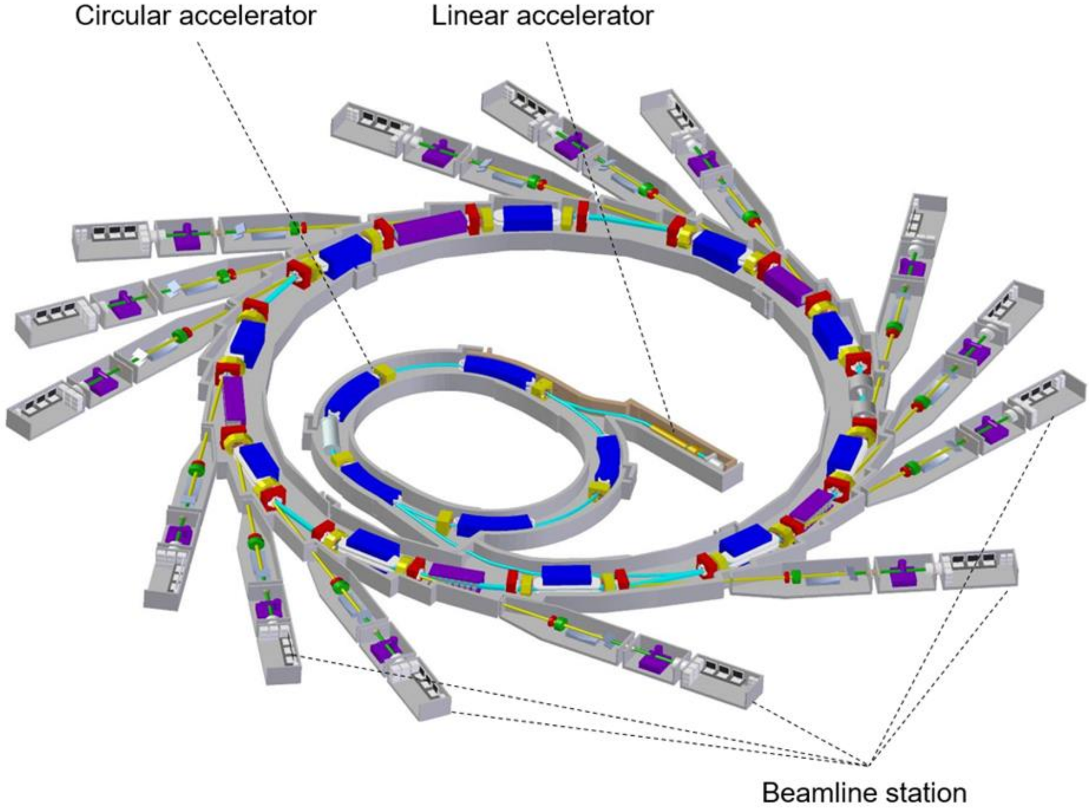
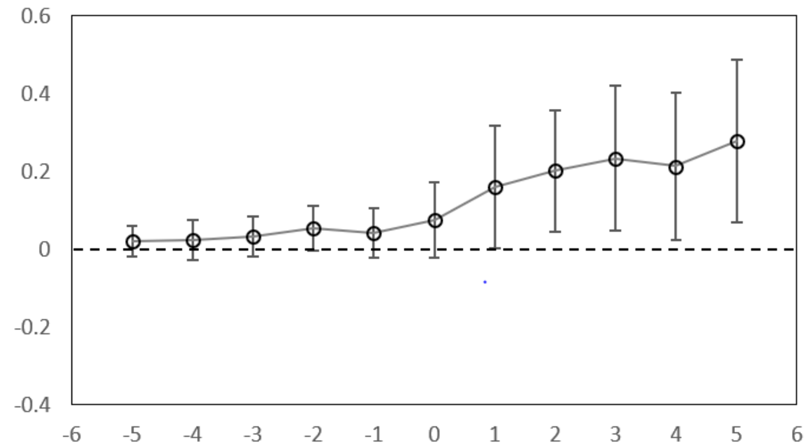


Fig. S1. Sketch-map of SSRF.
Source: Internal materials from the Shanghai Synchrotron Radiation Facility

Panel (A): DV=Percent_highquality_indigenous



Panel (B): DV=Mean_IF_indigenous

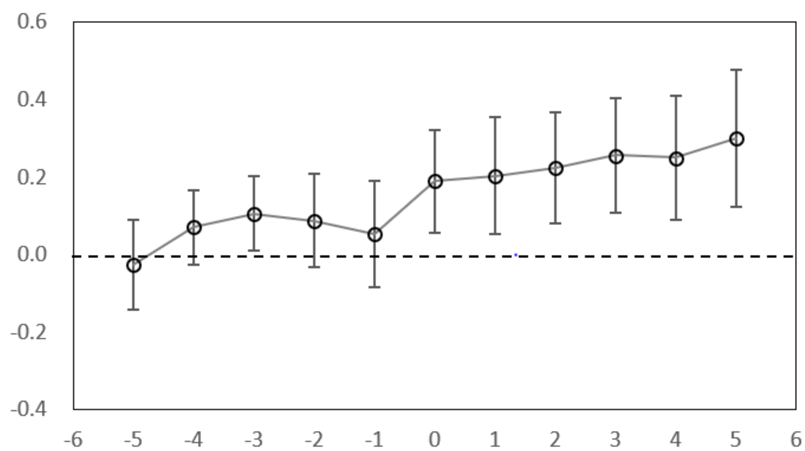


Fig. S2. Dynamic Impacts of SSRF on Treatment Group's Indigenous Publications.

Note: (i) the horizontal axis represents the time relative to the reference year, which is 2009 when SSRF was first put into service. "-5" is the fifth year prior to the reference year, and "5" is the fifth year after the reference year; (ii) the dots represent the estimated coefficients for the effect of SSRF on treatment group (disciplines that utilize SSRF for experimentation) compared to the control group (disciplines that are not affected by the opening of SSRF) relative to the reference year; (iii) the solid line shows the 95% confidence intervals, and standard errors are clustered at the discipline level

Table S1. List of Major Big-science RIs in China.

Name	Year of construction	Cost (in million RMB)	Location
Beijing Electron Positron Collider (BEPC)	1984	240	Beijing
Shenguang-II High Power Laser	1994	.	Shanghai
Heavy Ion Research Facility (HIRFL)	2000	293.5	Lanzhou
Experimental Advanced Superconducting Tokamak (EAST)	2000	200	Hefei
Shanghai Synchrotron Radiation Facility (SSRF)	2004	1,200	Shanghai
China Spallation Neutron Source (CSNC)	2007	1,200	Dongguan
Steady High Magnetic Field Facility (SHMFF)	2008	250	Hefei
Daya Bay Reactor Neutrino Experiment	2008	.	Shenzhen
Five-hundred-meters Aperture Spherical Telescope (FAST)	2009	.	Guizhou
Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST)	2009	235	Beijing
National Center for Protein Science	2010	756	Shanghai
Soft X-ray FEL	2011	200	Shanghai
Hefei Advanced Light Source (HALS)	2013	.	Hefei
Dalian Coherent Light Source, DCLS	2014	103	Dalian
Large High Altitude Air Shower Observatory (LHAASO)	2016	.	Daocheng
Earth System Numerical Simulation Facility (EarthLab)	2018	1,255	Beijing
Shanghai High repetition-rate XFEL and Extreme light facility (SHINE)	2018	10,000	Shanghai

Table S2. Robustness Tests with Placebo Treatments Prior to the Real SSRF Shock.

	(1)	(2)	(3)
Placebo treatment prior to the SSRF shock	Numb_highquality _indigenous	Percent_highquality _indigenous	Mean_IF_indigenous
5 years	-0.130***	-0.064*	-0.071**
	(0.045)	(0.036)	(0.032)
4 years	-0.092**	-0.055*	-0.039
	(0.041)	(0.032)	(0.032)
3 years	-0.057*	-0.046	-0.033
	(0.033)	(0.029)	(0.033)
2 years	-0.036	-0.037	-0.045
	(0.032)	(0.026)	(0.037)
1 year	-0.111**	-0.042	-0.061
	(0.056)	(0.035)	(0.048)

Notes: (i) the table reports estimated coefficients (β_1 in equation (1)) and standard errors (in parentheses) of tests that utilize pseudo treatments prior to the real opening of SSRF in 2009; (ii) each placebo dummy equals 1 when the placebo treatment is introduced, and equals 0 otherwise. For instance, "2 years" indicates that the placebo dummy equals 1 for the two years prior to the real opening of SSRF in 2009; (iii) coefficients and standard errors are obtained from separate regressions, each containing all the control variables in the baseline model in equation (1); (iv) *, **, *** indicate significance at the 10, 5, and 1% levels, respectively.

Table S3. Falsification Tests with Randomized Treatment and Control Groups.

	$\hat{\beta}$ Actual data	β_i Random data	$H_0: \hat{\beta} > \beta$ [p-value]
DV=Percent_highquality_indigenous	0.177	-0.0002	[p<0.000]
	(0.077)	(0.077)	
DV=Mean_ IF_indigenous	0.208	0.004	[p<0.000]
	(0.058)	(0.062)	

Note: (i) The randomization procedure is as follows: we use a uniform distribution to randomize whether any disciplines from our sample falls into the treatment group. The randomization procedure takes 1,000 random draws. We then estimate our baseline model in equation (1) for 1,000 times using the corresponding randomized treatment groups; (ii) We then compare the estimated coefficient using actual data ($\hat{\beta}$) and the mean of the estimated coefficients using the randomized data (β_i). The standard error of $\hat{\beta}$ and the standard deviation of β_i is reported in parentheses; (iii) p-values (in brackets) reflect the probability that the coefficient estimated using the actual data based on Table 2, Column 2 ($\hat{\beta}$=0.208) and Column 4 ($\hat{\beta}$=0.177) is greater than the mean of the coefficients estimated using the randomized data.

Table S4. Effect from Discipline's Initial Level of Researcher Composition

	(1) Mean_IF _domestic	(2) Percent_highquality _domestic
Treatment_After_Share of domestic institutions 2009	1.430***	1.128*
	(0.474)	(0.656)
Share of domestic institutions 2009_After	0.159	0.179
	(0.182)	(0.111)
Treatment_After	-0.950***	-0.753*
	(0.345)	(0.448)
After	-1.122***	-0.269
	(0.351)	(0.283)
Controls	Yes	Yes
Discipline FE	Yes	Yes
Year FE	Yes	Yes
Obs	3659	3983
R^2	.5010216	.0744085

Notes: (i) standard errors are clustered at the discipline level, and reported in parentheses; (ii) ***, **, * denotes significance at the 1, 5 and 10% levels respectively.

Appendix References

1. A. M. Weinberg, Impact of Large-Scale Science on the United States. *Science* **134**, 161-164 (1961).
2. L. Qiao, R. Mu, K. Chen, Scientific effects of large research infrastructures in China. *Technol Forecast Soc Change* **112**, 102-112 (2016).
3. X. Deng, "China must take its place in the field of high technology" in Selected Works of Deng Xiaoping Volume III. (People's Publishing House, Beijing, 1988), pp. 232.
4. H. Chen, *Large Research Infrastructures Development in China: A Roadmap to 2050* (Springer, Berlin, Heidelberg, 2011), https://doi.org/10.1007/978-3-642-19368-2.
5. Y. Wang, Y. Bai, Developing Mega-science Facility to Lead the Innovation Globally. *Management World* **5**, 172-188 (2020).
6. BerkeleyLab (2020) Scientists Say Farewell to Daya Bay Site, Proceed with Final Data Analysis.
7. H. Jia, Scientific collaborations shine on Belt and Road. *Natl Sci Rev* **4**, 652-657 (2017).
8. L. C. Yuan, A technical note on high energy physics. *Science in Contemporary China (Stanford: Stanford University Press, 1980)*, 111-119 (1980).
9. R. P. Appelbaum, C. Cao, X. Han, R. Parker, D. Simon, *Innovation in China: Challenging the global science and technology system* (John Wiley & Sons, 2018).
10. J. Tollefson, China declared world's largest producer of scientific articles. *Nature* **553**, 390 (2018).
11. E. Baark, "Global Implications of China's Policies on Domestic Innovation" in Innovation and China's Global Emergence, B. Hofman, J. Qian, E. Baark, Eds. (National University of Singapore Press, Singapore, 2021).
12. F. Huang (2018) Low quality studies belie hype about research boom in China. in *Sci Am* (Springer Nature).
13. H. C. Pollock, The discovery of synchrotron radiation. *Am J Phys* **51**, 278-280 (1983).
14. D. Cyranoski, China joins world-class synchrotron club. *Nature* **459**, 16-17 (2009).
15. Y. Sun, Shanghai synchrotron radiation facility unveils to the public. *Liao Wang* **18**, 38-39 (2009).