Research Evaluation, 26(2), 2017, 132–143
doi: 10.1093/reseval/rvx011
Advance Access Publication Date: 16 February 2017
Article

The role of the Chinese Key Labs in the international and national scientific arena revisited

Per Ahlgren[1,2,*], Ting Yue[2], Ronald Rousseau[3,4] and Liying Yang[2]

[1]KTH Royal Institute of Technology, School of Education and Communication in Engineering Sciences (ECE), Osquars backe 25, Stockholm 10044, Sweden, [2]Library of Chinese Academy of Sciences, 33 Beisihuan Xilu, Zhongguancun, Beijing P.R. China, 100190, [3]Faculty of Social Sciences, Middelheimlaan 1, University of Antwerp (UA), Antwerp B-2000, Belgium and [4]Department of Mathematics, Celestijnenlaan 200B, KU Leuven, Leuven B-3000, Belgium

*Corresponding author. Email: perahl@kth.se

Abstract

In this contribution, which builds on and develops a study that was published more than 10 years ago, we address the role of the Chinese Key Labs (KLs) in the international and national scientific arena. We give a short overview of the position of KLs in China, including their budget and manpower. Based on large numbers of Chinese publications obtained from the Web of Science (WoS) and the Chinese Science Citation Database (CSCD), the KLs are compared across publication years to the rest of China (ChRest) with respect to publication output and citation impact. We also look at collaboration in terms of co-publishing between the KLs and the ChRest. As to publications in the WoS, we found that the contribution of KLs compared with the ChRest is slightly and irregularly increasing (using full counting as well as fractional counting), whereas a stronger increasing trend is observed for the corresponding contribution in the CSCD. We observed an increase in the number of collaborations between KLs and Chinese colleagues, regardless of database. For WoS and field normalized citation indicators, we obtained the expected results that researchers at KLs perform considerably better than other Chinese colleagues and, moreover, perform clearly better than database average. As such we may conclude that KLs have lived up to their promise and made real impact on the international arena.

Key words: Chinese S&T system; Key Labs; publication volume; citation impact; internationalization; Chinese and foreign journals.

1. Introduction: Key Labs in China

A series of programs extending over a period of more than 30 years has resulted in the introduction, construction, and upgrading of a system of Key Labs (KLs) in Mainland China. As part of the national S&T innovation system, State KLs are a major force in basic research. They have the following two main functions: attracting outstanding talents and achieving high-level research. They, moreover, fulfill significant duties for the Chinese government. To reach these goals, they are endowed with high-quality equipment. As such, these labs aim to make China's research more visible to the global scientific community, and upgrading it to the level of the best western countries. By bringing the best scientists together in a specialized and well-equipped environment, China is convinced that it will be able to lead the way in ever more scientific fields. Scientists recruited to work in KLs do not only come from China but also from abroad. Since 2008, the Ministry of Finance of China has set up special funds to maintain the basic operation of the country's KLs. In 2014, these special funds approached 3.045 billion Yuan, which is 0.23% of the R&D budget. Besides these special funds from the central government, KLs are also supported by project money. In 2014 KLs received project funding for an amount of 16.78 billion Yuan, which is 1.29% of the R&D budget. KLs accept management by a support institution and regular evaluation by the government (Tsou 1989). In 2014, there were 22,124 full-time and 9,545 part-time staff members in KLs. The number 22,124 refers to all full-time staff (scientific and administrative staff). Unfortunately, a number that pertains to full-time research staff only is not available to us. Regarding the development of full-time staff, we only have statistics for the years 2013 and 2104. From the former to the latter year, the full-time staff increased by 3.9%. The number of KLs itself has been relatively stable during recent years.

According to the latest data (Ministry of Science and Technology of China 2016), China has 258 KLs divided over different scientific fields as shown in Table 1.

The objective of this work is to shed light on the role of the KLs in the international and national scientific arena. We use data obtained from the Web of Science (WoS) and the Chinese Science Citation Database (CSCD) to compare the KLs across publication years to the rest of China (ChRest) with respect to publication output and citation impact. We also deal with collaboration in terms of co-publishing between KLs and the ChRest.

The rest of this article is organized as follows. In Section 2, we report previous research on the KLs. Section 3 describes the data and methods of the study, whereas the findings are given, and discussed, in Section 4. In the last section, conclusions are put forward.

2. Previous research

In 2005–6, Jin, Rousseau and Sun (2005, 2006) published two articles about the role of State KLs and departmental Open Labs in the Chinese S&T system. In the first of these articles (Jin, Rousseau and Sun 2005), the authors explained the evaluation system applied to these labs. We briefly recall the main points, incorporating some changes that have been introduced since then. The evaluation procedure is mainly qualitative and is performed by experts, some of which come from abroad, according to a set of predefined quality indicators. These quality indicators are subdivided into three groups:

- Research level and contributions
- Increasing scientific capacity (including training)
- Open communication and management structure

Each of these aspects includes several sub-aspects. Under 'Research level and contributions', the following four points are considered: orientation, tasks, key results and collaboration and independent research, explained below.

Orientation and task: the lab is classified as a basic research lab or an applied-basic research lab (KLs are not supposed to make products directly aimed at the market and hence are never of the purely applied type). It has a clear statement of goals and means and knows its priorities. It is capable of taking on major tasks for the government and has high productivity.

Table 1. KLs per field

Field	Number of KLs
Geosciences	46
Biology	44
Engineering	43
Information Science and Technology	32
Medical Science	31
Chemistry	25
Material Science	21
Mathematics and Physics	16

Key results: the lab presents the best research results obtained over the latest 5 years and which fall within the scope of the lab's (scientific and/or technological) orientation and goals. Yet, for obvious reasons, basic research and applied-basic research labs are evaluated according to different standards.

For basic research, requirements include are

- doing research on the frontier of science[1];
- publishing original articles in international journals with high impact;
- publishing monographs; and
- giving keynote speeches at important international conferences.

For applied research, the corresponding requirements are

- developing new methods and ideas of importance for the national economy, social development, or national safety;
- making considerable progress in experimentation;
- doing innovative work, especially in key technologies;
- obtaining patents; and
- building up a repertoire of new techniques with high potential for industrial applications.

Moreover, both types of labs must show that they obtained fundamental scientific data and information, and that based on these data and information, they provided an open service in a spirit of resource sharing. All these activities must support scientific research and policy decisions.

Collaboration and independent research: On the one hand, the lab must have a research plan oriented versus international collaboration and actively plans international activities. The most important indicator here is publishing papers through international collaboration. On the other hand, the lab must be able to develop independent research and act as a true representative of its research field.

The point 'Increasing scientific capacity' refers to the research capacity of individual scientists as well as the lab as a whole. It is subdivided into three sets of requirements. The first relates to the lab director and upper-level researchers. Here the CVs of the top persons are taken into account. The reason for including these in the evaluation is that labs must be directed by top-level scientists and academic leaders, not by mediocre scientists, administrators, or politicians. Directors must have the intellectual capacity and the time to work in their lab as scientists (not just as administrators), and play a central scientific role. Upper-level researchers must be researchers with a known reputation in the field. The second point relates to the personnel structure and aspects of team building. Here the internal structure and collaboration (teamwork) are evaluated. This structure should be such that it leads to the best possible scientific research. The lab should neither be too fragmented into small units nor be a huge mastodon with little internal structure. Researchers are encouraged to take upon them leading duties in academic organizations at the national and international level. Leading scientists from all over the world should be invited to participate in the lab's work. Teamwork in a spirit of cooperation should be encouraged. Moreover, teams should consist of scientists of all ages, but be dominated by younger and middle-aged colleagues. The lab is to be characterized by an intellectual and science-friendly atmosphere. Finally, the lab must provide training for young scientists.

Finally, ‘Open communication and management’ consists, like ‘Research level and contributions’, of four parts.

1. Publicly shared facilities and instruments. Does the lab have all the research facilities and instruments necessary for research at the forefront of science? The lab should develop its own instruments for leading-edge experiments and share the use of large expensive instruments with other labs.
2. Open academic communication. The lab must have an open communication structure, internally as well as with the outside world, in particular with fellow scientists all over the world. Open communication also implies sharing the use of costly instruments (see above). Its openness is also characterized by participation in international events, and the organization of international, national, regional and local conferences, and symposia. Moreover, an Academic Degrees Committee must play an important role in the selection of research fields and the evaluation of the lab.
3. Management. Clear management guidelines must exist, so that each member knows their duties and all operations are performed smoothly. The lab has access to all necessary research materials. Financial conditions are adequate and the lab receives the necessary support from an affiliated institute.
4. Open research and scientific knowledge dissemination. Rules are provided about temporary and visiting scholars, and research projects exist making it possible to invite skilled colleagues and to perform collaborative research. Finally, the lab should disseminate scientific knowledge to the public and especially to students.

As with all evaluation procedures also, these indicators have a prescriptive function: they tell lab directors what they should certainly do, namely, activities that are explicitly mentioned in the evaluation procedures and what is of less importance, namely, activities that are not evaluated.

Each research-intensive institution has a unique and characteristic profile with respect to the research that is performed in this institution. This is taken into consideration during institutional evaluation exercises. If the outcome of such an evaluation considers a KL to be unqualified, then it will lose its KL status.

The second article (Jin, Rousseau and Sun 2006) dealt with the role played by these labs within the national research system. Citation data were retrieved from CSCD (Jin and Wang 1999) and the *Science Citation Index* (SCI). It was observed that at that time, almost one quarter of all internationally oriented Chinese publications originated from these labs. A similar remark held for citations was received by Chinese scientists in the SCI. Comparisons between SCI-based and CSCD-based performance results showed that the relative academic impact of KLs and Open Labs was more international than domestic. KLs had a higher total production and received more citations than Open Labs. Yet their impact, measured as citation per publication (CPP), was very similar. Moreover, scientists working outside these labs obtained a similar impact as those working at these prestigious labs. At that time, it was concluded that when it came to impact on the international scene, these labs had not yet led to a big step forward for Chinese science as a whole.

Besides these two articles, we do not know any other studies dealing with KLs and involving bibliometrics. Yet, several studies deal with the bibliometric performance of Chinese universities or university departments (Zhu et al. 2004; Zhou and Leydesdorff 2011; Fu and Ho 2013; Zhu et al. 2014). Other studies focus, using bibliometric approaches, on the internationalization of Chinese journals (Wang, Wang and Weldon 2007; He and Liu 2009), or on the production and citation impact of Chinese authors within a single journal (Yan, Rousseau and Huang 2016). The bibliometric performance of China within different scientific fields has also been studied (Huang et al. 2015; Sun and Hua 2015; Yan and Sun 2015; Yang et al. 2015), as well as university–industry co-publishing (Fan, Yang and Chen 2015) and university–industry–research institute co-publishing (Gao, Guo and Guan 2014) in China. Chinese science policy, including the role of KLs, is discussed in (Benner, Liu and Serger 2012).

In this contribution, more than 10 years after the original investigation, we have a new look at the results obtained by KLs and compare them with the rest of Chinese research. We note that the importance of so-called Open Labs has been reduced (many do not exist anymore as Open Labs); hence, they are not anymore taken into account. Moreover, we focus on publication results as in (Jin, Rousseau and Sun 2006) and do not reconsider the evaluation procedure (Jin, Rousseau and Sun 2005).

The article considers two main aspects: publications and citations. As usual these are used as proxies for production and impact, respectively.

3. Data and methods

Our main data sources are Bibmet, the bibliometric version of the WoS at KTH Royal Institute of Technology (Sweden), used for all international aspects of our investigation, and the CSCD, used for aspects specifically related to China. The main reason for using the CSCD, and not one of the other Chinese databases, is that the original investigation too used the CSCD. CSCD was established in the year 1989. It is provided by the National Science Library of CAS, Beijing. Currently, CSCD includes approximately 1,000 journals (20% of them are in English). Each journal belongs to at least one of the following fields: agriculture, astronomy, biology, chemistry, engineering, environmental science, forestry, geography, health, mathematics, medicine, and physics and technology. Rousseau 2015 estimates that there are about 6,000 academic journals in China, among which somewhat more than 2,000 belong to the social sciences and humanities. This indicates that inclusion in the CSCD, which does not include the social sciences and humanities, is rather restrictive. Not all Chinese journals are included in CSCD. Journals are first evaluated based on their performance with regard to a set of citation indicators. Journals with a performance above a certain threshold qualify as candidates for inclusion in CSCD. The candidates are then assessed by subject experts affiliated to Chinese universities and institutes, and some of the candidates are selected for inclusion. In the year 2007, CSCD became the first non-English language database in the Web of Knowledge platform (provided—at that time—by Thomson Reuters).

For WoS data, the publications of the study are articles (WoS document type: Article) and reviews (WoS document type: Review) with at least one Chinese address, recorded in the three journal indexes of the WoS (SCI-EXPANDED, SSCI, and A&HC) and published in the period 2001–14. Citations to these publications were collected at the end of the year 2015.

We use (address) fractionalization, which leads to fractional counting, and compare KLs to the ChRest. Publications with at least one KL address were identified with a regular expression approach, involving three steps. Such an expression is a sequence of characters that specifies a search pattern (Hopcroft, Motwani and Ullman 2007). We constructed several regular expressions, involving wild cards and (parts of) expressions known to refer to KL, to search the address field of the publications for matching text segments. For instance, the following regular expression (within single quotation marks) was used: 'state key labs?|state labs?|natl labs?|natl key labs?', where '?' is a wild card (zero or one occurrence of the character that immediately precedes '?') and '|' the alternation operator. The expression searches for text parts that matches any of the subexpressions separated by '|'. The many text operands used in the regular expressions, together with the expressions themselves, can be delivered to the reader upon request.

A potential problem in the retrieval of KL addresses concerns false negatives, i.e. KL addresses that are not retrieved by the used regular expression approach. We believe, however, that the extent to which false negatives occur is negligible. We base this belief on the following. The bibliometric group at the National Science Library of CAS, Beijing, has manually standardized all WoS Chinese addresses (with an annual update for addresses in new bibliographical records). In the standardization process, a large number of variant organization names are identified. In our regular expression approach, the identified organization name variants for KL are used. Also regarding CSCD, address standardization is used. In the light of these facts, the issue of false negatives is a minor one.[2]

Let p be a publication, T the total number of addresses in p, m the number of addresses in p, where the country name is 'China' (referring to an address in China), and n ($n \leq m$) the number of KL addresses in p. Then:

1. The *China address fraction* of p and of the citations received by p is defined as m/T.
2. The *KL address fraction* of p and of the citations received by p is defined as n/T.
3. The *ChRest address fraction* of p and of the citations received by p is defined as $(m - n)/T$.

In the study, we use (2) and (3).[3] We note that research performed in Hong Kong and Macao belongs to ChRest, as these regions (with a few exceptions) do not have KL, but are part of China. To investigate the effect of research performed in Hong Kong and Macao on the outcome of the analyses, the analyses were performed also on the basis of the data set obtained from the full data set by removing all Hong Kong and Macao addresses. We, however, do not report the results for this latter data set, since these are very similar to the corresponding results for the full data set, and they do not provide new information.

We further note that by 2002 there were 162 KL, while now (2016) there are already 258 ones.

Our study is divided into two parts. In the first part, raw citations, i.e., non-normalized ones, are used (cf. Jin, Rousseau and Sun 2006), whereas the second part makes use of field normalization of citations. As the structure of the CSCD does not allow for such normalization, the domestic aspect of our investigation based on the CSCD belongs completely to the first part.

3.1 Part 1

In this part, we mainly use fractional counting. However, we also use full counting, for instance regarding collaboration between KL and ChRest. When the full counting method is used, a publication is fully assigned to each unit of analysis (ChRest and KL in this work) that is represented in the publication. Two of the indicators used in this part of the study are the number of KL publications (citations) divided by the number of ChRest publications (citations), multiplied by 100. The first indicator refers to production and the second one to impact. In the remainder of this work, we refer to these two indicators as 'Ratio pub.' and 'Ratio cit.', respectively. The publication period for Ratio pub. is 2001–14. We used a 4-year citation window: $[Y, Y+3]$, where Y is a publication year. Therefore, the publication period for Ratio cit. is 2001–11 (recall that the citations were collected at the end of 2015).[4] Publications as well as citations are weighted by address fractions. For instance, if the KL fraction of publication p is 1/4, and the citation rate of p is 8, then p is represented in the numerators of the two indicators with 1/4 (production indicator) and $1/4 \times 8 = 2$ (impact indicator).

KL performance compared to ChRest over time is analyzed with regard to both the CSCD and the WoS as in (Jin, Rousseau and Sun 2006). The total number of WoS publications for the period 2001–14 is 1,638,170, whereas the corresponding number of CSCD publications is 2,428,007. For WoS data, we analyzed citation impact under two scenarios: inclusion and exclusion of Chinese self-citations. In this work, a citation from a publication p to a publication p', which has at least one China address, is a *China self-citation* if also p has at least one China address. Also for the CSCD data, we analyzed citation impact under inclusion/exclusion of self-citations. However, since it is not meaningful to use exclude *China* self-citations in a Chinese database, we instead excluded *author* self-citations. In this work, a citation from a publication p to a publication p' is an *author self-citation* if p and p' have at least one common author name.

3.2 Part 2

In this part of the study, where the focus is on the normalized citation impact performance of KL and ChRest, 1,382,922 publications from the WoS are used. These publications are published in the period 2001–13. An open citation window, hence with variable length, is used. For instance, the windows for the publication years 2001 and 2013 are [2001, 2015] and [2013, 2015], respectively. The year 2013, and not 2014, is used as the last publication year to avoid an improperly short citation window.

We analyze the development over time of the normalized citation impact of KL and ChRest, where the indicators mean field normalized citation rate (mncr) and proportion of publications among the 10% most cited (Top10) are used. Basically, the mncr is the same indicator as the mean normalized citation score (Waltman et al. 2011a, 2011b). Given a set of publications, say P, of a unit of analysis, mncr measures the average normalized citation rate across the publications in P. To obtain the normalized citation rate for a given publication p in P, the citation rate of p is divided by a field reference value with regard to the field to which p belongs. The reference value is the mean citation rate across all publications, published in the same year and belonging to the same document type as p, in the field of p. These publications are the reference publications of p (one of them is p itself). Top10, also a ratio, is a percentile-based indicator which is, compared to mncr, less sensitive to extreme values. For this indicator, the publication p is compared to the same reference publications as in the mncr case. By using an approach proposed by Waltman and Schreiber (2013), and used in the CWTS Leiden Ranking 2016,[5] we calculate the fraction of p with which p is assigned to the 10% most cited publications regarding the distribution. By doing this for all publications in P, we obtain the proportion of publications in *P* among the 10% most cited. The used approach ensures, for a given field, that exactly 10% of the publications of the field belong to the top 10% publications. In Appendix 1, an example, based on scenarios put forward by Waltman and Schreiber (2013), illustrates the approach. For the mathematical framework of the approach, we refer the reader to Waltman and Schreiber (2013).

Traditionally, and in this study, a classification scheme is used for field normalization (Braun and Glänzel 1990; Moed, De Bruin and van Leeuwen 1995; van Raan 1996). Each considered publication is assigned to one or more of the fields of the scheme. Typically, and in our case, the fields used for normalization are the WoS subject categories. These categories are such that a publication can belong to more than one category (if its journal has been assigned to more than one category). Let p be a publication associated with more than one category. The normalized citation rate of p was calculated as the arithmetic mean across the normalized citation rates of p with respect to the categories to which p belongs. Note that the arithmetic mean across the normalized citation rates for a publication p, with c citations, is equal to c over the harmonic mean of the field reference values for p (see Waltman et al. 2011b for an illustration of the equivalence of the calculation methods). For the fraction of p with which p is assigned to the 10% most cited publications, we obtained this fraction by calculating the arithmetic mean across the corresponding fractions of p with respect to the categories to which p belongs. For instance, assume that p belongs to three WoS subject categories. For each of these three categories, we first calculate the fraction of p with which p is assigned to the 10% most cited publications, using the Waltman–Schreiber approach. Then the three resulting fractions, all of them multiplied by 1/3, are summed, which gives the arithmetic mean across the three fractions.[6]

Regarding the calculation of field reference values, which are ratios and used in the calculation of mncr, a publication p that belongs to a subject category *SC* contributes to the denominator of the field reference value of SC with $1/k$, where k is the number of subject categories of (the journal of) p, and with $1/k \times c_p$, where c_p is the citation rate of p, to the numerator of the reference value.

Both normalized citation rates and fractions with which publications are assigned to the 10% most cited publications were weighted by address fractions. For instance, if the KL fraction of publication p is 1/4, both the normalized citation rate of p and the fraction of p with which p is assigned to the 10% most cited publications are multiplied by 1/4. Further, p is represented in the denominators of mncr and Top10 by 1/4 (and not by 1).

Note that mncr = 1 and Top10 = 0.1 can be interpreted as database averages. For formal definitions of mncr and Top10, see Ahlgren and Sjögårde (2015), Sections 1.1, and 1.2, respectively. The definition of Top10 involves a description of the Waltman–Schreiber assignment of publication fractions to the top 10 segment of a citation distribution.

The normalized citation impact was studied separately for contributions in Chinese and non-Chinese journals, and for contributions regardless of the country of journals.

4. Results and discussion

4.1 Part 1A. Results based on the WoS

Table 2 provides results based on full counting. Let us first explain the notation: '# pub.' denotes the total number of publications (articles or reviews) in the WoS with a least one China address, '# KL pub.' denotes the number of publications with a least one KL address, and '# ChRest pub.' denotes the number of publications with at least one Chinese address minus the number of publications in which all Chinese addresses are KL addresses, in other words: these publications contain at least one Chinese organization, which is not a KL.

As a consequence of using full counting in Table 2, it is possible to calculate the number of publications for which KL scientists have collaborated with other Chinese scientists. Indeed, this number is the sum of KL publications and ChRest publications from which the number of all publications has been subtracted. It then becomes also possible to count the relative number of publications for which this collaboration has occurred. Besides an absolute increase for all types of publications, we also note an increase in the relative number of Chinese publications originating (at least in part) from KL, namely, from 17.8 to 23% and an increase of the percentage of publications resulting from a collaboration between KL and other Chinese research units, namely, from 8.7 to 15.6%.

In Table 3 we use a form of country-based fractional counting. It is country-based, as we start from the number of publications with at least one Chinese address (see Table 2, second column). Then we used fractional counting as follows: if there are two different KL

Table 2. Number of Chinese publications by publication year: full counting

Publication year	# pub.	# KL pub.	% KL pub.	# ChRest pub.	% Collaboration
2001	35,295	6,286	17.8	32,094	8.7
2002	39,596	7,451	18.8	35,734	9.1
2003	47,822	9,189	19.2	43,230	9.6
2004	59,404	11,093	18.7	54,001	9.6
2005	71,910	14,232	19.8	65,424	10.8
2006	85,943	17,468	20.3	78,269	11.4
2007	95,393	19,882	20.8	86,940	12.0
2008	109,264	23,680	21.7	99,487	12.7
2009	126,277	27,966	22.1	115,057	13.3
2010	139,328	31,116	22.3	127,168	13.6
2011	162,796	37,303	22.9	149,106	14.5
2012	188,137	44,845	23.8	172,168	15.3
2013	221,757	51,992	23.4	203,876	15.4
2014	255,248	58,749	23.0	236,409	15.6

Table 3. Number of Chinese publications by publication year: fractional counting

Publication year	KL pub. frac	ChRest pub. frac	foreign frac	% KL frac	% foreign	Ratio pub.
2001	4,178.8	26,797.8	4,318.4	11.8	12.2	15.6
2002	5,015.7	29,614.6	4,965.6	12.7	12.5	16.9
2003	6,055.1	35,892.7	5,874.2	12.7	12.3	16.9
2004	7,272.9	45,035.4	7,095.7	12.2	11.9	16.2
2005	9,074.0	54,666.6	8,169.4	12.6	11.4	16.6
2006	10,885.3	65,419.0	9,638.6	12.7	11.2	16.6
2007	12,205.2	72,424.7	10,763.2	12.8	11.3	16.9
2008	14,253.3	82,178.1	12,832.6	13.0	11.7	17.3
2009	16,501.9	94,728.1	15,047.1	13.1	11.9	17.4
2010	17,898.2	103,756.4	17,673.4	12.8	12.7	17.3
2011	20,907.4	121,185.7	20,702.9	12.8	12.7	17.3
2012	24,778.4	139,705.7	23,652.9	13.2	12.6	17.7
2013	28,224.7	165,802.4	27,729.9	12.7	12.5	17.0
2014	30,968.5	192,641.4	31,638.1	12.1	12.4	16.1

addresses, four other (different) ChRest addresses and four different foreign addresses mentioned in a publication, then KL is assigned a publication fraction of 2/10, ChRest a fraction of 4/10, and foreign a fraction of 4/10. The number of authors belonging to these addresses is not taken into account because before 2008 this information was not available in the WoS.

The percentage contribution of KL fluctuates between 11.8 and 13.2% (Column 5 of Table 3) and is slightly increasing. The same observation holds for foreign units. The seventh column concerns the indicator Ratio pub. (defined in the section 'Data and methods'): the ratio (in %) between KL publication fractions and ChRest publication fractions (values in Column 2 divided by values in Column 3). This ratio behaves irregularly but has a slightly increasing trend (Fig. 1; $r = 0.45$).

Table 4 shows the number of fractionated citations for KL and ChRest with China self-citations (as defined in the section 'Data and methods') included and excluded. The next last column concerns the indicator Ratio cit. (defined in the section 'Data and methods'): the ratio (in %) between fractionated KL citations and fractioned ChRest citations (values in Column 2 divided by values in Column 4). The last column shows the same data but omitting China self-citations (values in Column 3 divided by values in Column 5). The data of the two columns are visualized in Fig. 1.

Table 4 shows that for KL, relative to ChRest, the number of fractionated citations is increasing over the years. The same observation holds when self-citations are removed. Here we note that, as the ratio without China self-citations is always smaller than that when self-citations are included, KLs have relatively more China self-citations than other research units. This result seems contrary to the international mission of KL. We have not yet been able to find a clear explanation for this observation.

As there are much more non-Chinese than Chinese journals in the WoS, we expect the ratio of publications in non-Chinese journals over publications in Chinese journals to be (much) larger than one. We moreover expect this ratio to increase, especially for KL, as they have an international mission. Finally, and for the same reason, we expect this ratio to be larger for KL than for ChRest. Results are reported in Table 5 (Columns 4 and 7), and graphically shown in Fig. 2.

Table 5 yields a surprise. As expected, all ratios are larger than one, and as expected, they increase over the years. Yet, for 6 of the 14 considered years, the ratio for KL is smaller than for ChRest. For

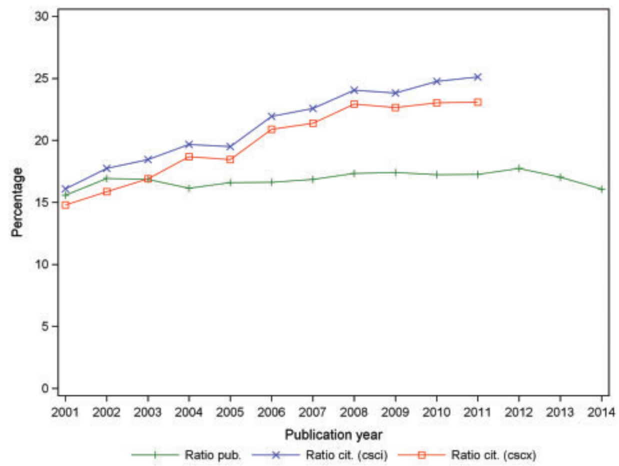

Figure 1. Percentage KL publications and citations relative to ChRest. 'csci(x)' stands for 'China self-citations included (excluded)'. Fractional counting. 4-year citation window.

instance, the KL ratio is smaller for the latest 3 years. This would point to the surprising conclusion that the ChRest is more internationally oriented than KL (this point was already made in relation to Table 4, regarding citations received from China). One possible explanation is that most Chinese journals included in the WoS are hosted by KL, and it might be the case that scientists affiliated to these labs often prefer to publish in KL journals.[7]

4.2 Part 1B: Results based on the CSCD

In this section we report results based on the Chinese database CSCD. Tables 6–8 are counterparts to Tables 2–4, respectively. The notation is explained in the first paragraph of the section 'Part 1A. Results based on the WoS'.

Table 6 gives results based on full counting. The number of KL publications in CSCD increased from 7,587 in 2001 to 24,532 in 2014. We note an increase in the relative number of Chinese publications originating from KL, from 5.1 to 8.8%. This outcome agrees with the corresponding outcome for WoS data (Table 2). However, these relative numbers are consistently lower for CSCD data than for WoS data. The percentage of publications co-published by KL and ChRest increased from 1.9% in 2001 to 5.5% in 2014.

Table 4. Citation data for KL and ChRest by publication year

Publication year	KL cit. frac (csci)	KL cit. frac (cscx)	ChRest cit. frac (csci)	ChRest cit. frac (cscx)	Ratio cit. (csci)	Ratio cit. (cscx)
2001	14,811.4	6,193.4	91,973.8	41,870.2	16.1	14.8
2002	20,727.7	8,345.7	116,811.4	52,617.1	17.7	15.9
2003	29,337.2	12,194.6	158,855.3	72,120.0	18.5	16.9
2004	40,977.7	17,295.1	208,230.1	92,528.4	19.7	18.7
2005	51,919.2	21,577.5	266,193.7	116,961.2	19.5	18.5
2006	71,329.9	30,253.6	325,058.8	144,793.0	21.9	20.9
2007	89,517.6	38,229.8	396,434.6	178,703.6	22.6	21.4
2008	115,872.1	49,455.2	481,613.3	215,659.9	24.1	22.9
2009	141,777.7	59,375.4	595,252.0	262,098.8	23.8	22.7
2010	172,374.4	68,666.5	695,913.0	298,174.2	24.8	23.0
2011	214,416.0	81,083.4	853,894.2	351,042.1	25.1	23.1

Note: 'csci(x)' stands for 'China self-citations included (excluded)'. Fractional counting. 4-year citation window.

Table 5. Ratio of publications in non-Chinese journals over publications in Chinese journals (Ratio non-ChJ/ChJ) by publication year

Publication year	KL			ChRest		
	pub. frac (ch)	pub. frac (nch)	Ratio non-ChJ/ChJ	pub frac (ch)	pub. frac (nch)	Ratio non-ChJ/ChJ
2001	1,286.2	2,892.6	2.25	7,359.5	19,438.3	2.64
2002	1,431.4	3,584.3	2.50	7,999.3	21,615.3	2.70
2003	1,635.7	4,419.4	2.70	9,172.7	26,720.0	2.91
2004	1,776.8	5,496.2	3.09	10,621.6	34,413.8	3.24
2005	1,911.5	7,162.5	3.75	11,815.8	42,850.8	3.63
2006	2,015.9	8,869.4	4.40	12,850.4	52,568.6	4.09
2007	2,210.1	9,995.0	4.52	14,625.8	57,798.8	3.95
2008	2,512.5	11,740.8	4.67	16,328.1	65,850.0	4.03
2009	2,717.6	13,784.3	5.07	17,110.3	77,617.8	4.54
2010	3,112.8	14,785.5	4.75	18,775.8	84,980.6	4.53
2011	3,181.3	17,726.2	5.57	18,928.0	102,257.7	5.40
2012	3,626.0	21,152.4	5.83	19,967.1	119,738.6	6.00
2013	3,638.2	24,586.5	6.76	19,485.7	146,316.6	7.51
2014	3,204.9	27,763.6	8.66	19,324.2	173,317.2	8.97

Note: Fractional counting. 'ch' stands for 'Chinese journals' and 'nch' for 'non-Chinese journals'.

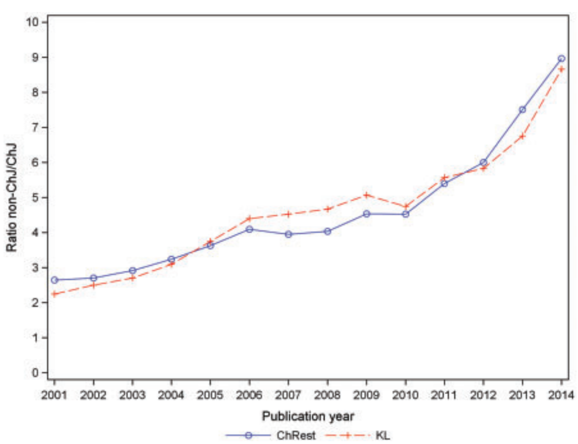

Figure 2. Ratio of publications in non-Chinese journals over publications in Chinese journals (Ratio non-ChJ/ChJ) by publication year. Fractional counting.

In Table 7, country-based fractional counting is used.[8] The percentage contribution of KL fluctuates between 3.9 and 6.2% and is increasing. The ratio between KL publication fractions by ChRest publication fractions, Ratio pub., has an increasing trend (Fig. 3; $r = 0.93$). This is in agreement with the corresponding outcome for WoS data (Table 3), but the correlation is stronger for CSCD data.

Table 8 shows the number of fractionated citations for KL and ChRest with author self-citations (as defined in the section 'Data and methods') included and excluded. The next to last column concerns the indicator Ratio cit.: the ratio (in %) between fractionated KL citations and fractioned ChRest citations. The last column shows the same data but omitting author self-citations. The data of the two columns are visualized in Fig. 3.

Table 8 and Fig. 3 show that for KL, relative to ChRest, the number of fractionated citations is increasing over the years, regardless of whether author self-citations are excluded. For the latest three publication years, the ratio when author self-citations are excluded is approximately equal to the ratio when author self-citations are included.

4.3 Part 2: A study using field normalized citation indicators

Table 9 uses field normalization to study the research performance (in terms of received citations) of KL and ChRest. Values of the indicators mncr and Top10 by publication year and journal set are shown. Besides the two journal sets used for Table 5—the set of Chinese journals and the set of non-Chinese journals—we here use the union of these two sets, i.e. the set of all journals of the study. The results are illustrated in Fig. 4 (mncr) and Fig. 5 (Top10). KLs perform consistently better than ChRest when the set of all journals or the set of non-Chinese journals is considered, regardless of indicator. For KL and for these two journal sets, mncr and Top10 values are increasing and become considerably larger than the corresponding database averages, 1 and 0.1, respectively. The performance for ChRest, regarding the two journal sets, is increasing, but has not yet reached world average. We note, though, a performance drop for KL (both indicators) from the year 2004 to the year 2005. We did not find any science policy explanation for this phenomenon. However, we have some support in our data for a descriptive explanation. The largest relative increase in KL publication fraction output with regard to year $Y-1$ to year Y occurs for 2004 and 2005 (Table 3, second column). The increase is about 25%, considerably larger than for any other pair of considered years. Further, the indicator CPP with regard to KL, China self-citations being included or not, shows the smallest relative increase for 2004 and 2005, 1.55 and 0%, respectively (Table 3, second column; Table 4, second and third columns). This is considerably smaller than for any other year pair. The relatively higher growth in publication output might have led to the observed KL decrease in normalized citation impact: proportionally more low-impact publications, relative to their fields, could have been published 2005 compared to 2004.[9]

Results for publications in Chinese journals, although increasing, are considerably below their field's database averages, as expected. We clearly see that when we control for differences in citation volumes between fields, KLs perform considerably better than other research organizations in China. We also studied these data when excluding author self-citations, but results are not reported, as they are very similar to the reported ones.

Table 6. Number of Chinese publications by publication year: full counting

Publication year	# pub.	# KL pub.	% KL pub.	# ChRest pub.	% Collaboration
2001	149,829	7,587	5.1	145,103	1.9
2002	164,668	8,448	5.1	160,334	2.5
2003	189,416	9,384	5.0	184,301	2.3
2004	202,427	10,705	5.3	197,223	2.7
2005	219,886	12,284	5.6	214,435	3.1
2006	236,295	13,922	5.9	229,901	3.2
2007	266,596	16,077	6.0	260,036	3.6
2008	272,680	18,339	6.7	266,280	4.4
2009	291,519	18,277	6.3	289,642	5.6
2010	293,239	20,217	6.9	286,907	4.7
2011	290,544	22,883	7.9	281,762	4.9
2012	288,777	23,889	8.3	279,670	5.1
2013	282,559	24,780	8.8	273,188	5.5
2014	279,572	24,532	8.8	270,531	5.5

Table 7. Number of Chinese publications by publication year: fractional counting

Publication year	KL pub. frac	ChRest pub. frac	% KL frac	Ratio pub.
2001	6,061.4	143,767.6	4.0	4.2
2002	6,724.6	157,943.4	4.1	4.3
2003	7,464.8	181,951.2	3.9	4.1
2004	8,399.0	194,028.0	4.1	4.3
2005	9,462.9	210,423.1	4.3	4.5
2006	10,526.3	225,768.7	4.5	4.7
2007	11,421.2	255,174.8	4.3	4.5
2008	12,323.5	260,356.6	4.5	4.7
2009	13,614.9	277,904.1	4.7	4.9
2010	14,807.1	278,431.9	5.0	5.3
2011	16,611.7	273,932.3	5.7	6.1
2012	16,937.6	271,839.4	5.9	6.2
2013	17,488.2	265,070.8	6.2	6.6
2014	17,136.3	262,435.7	6.1	6.5

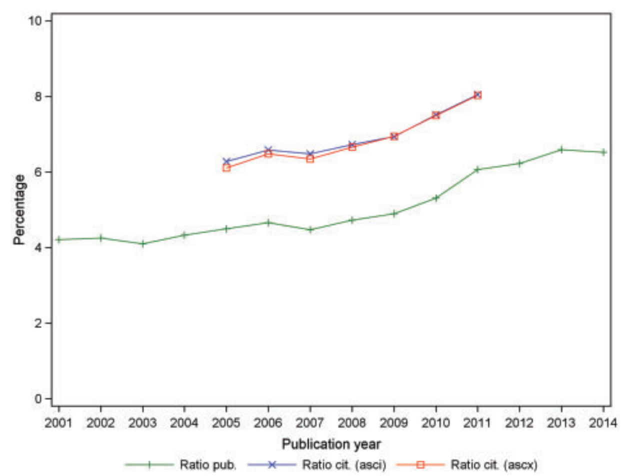

Figure 3. Percentage KL publications and citations relative to ChRest. 'asci(x)' stands for 'author self-citations included (excluded)'. Fractional counting. 4-year citation window.

Table 8. Citation data for KL and ChRest by publication year

Publication year	KL cit. frac (asci)	KL cit. frac (ascx)	ChRest cit. frac (asci)	ChRest cit. frac (ascx)	Ratio cit. (asci)	Ratio cit. (ascx)
2005	12,597.9	6,250.9	200,507.1	102,325.1	6.3	6.1
2006	15,047.1	8,670.4	228,591.0	133,790.7	6.6	6.5
2007	15,856.4	9,997.1	244,538.6	157,468.9	6.5	6.4
2008	17,230.1	11,700.1	255,897.9	175,700.9	6.7	6.7
2009	18,485.3	13,343.6	266,214.7	191,966.4	6.9	7.0
2010	21,033.5	15,822.2	279,887.5	210,853.8	7.5	7.5
2011	15,863.1	12,100.9	196,954.9	150,662.1	8.1	8.0

Note: 'asci(x)' stands for 'author self-citations included (excluded)'. Fractional counting. 4-year citation window.

Table 9. mncr and Top10 for KL and ChRest by publication year and journal set

Publication year	Journals	mncr-KL	mncr-ChRest	Top10-KL	Top10-ChRest
2001	nch	0.93	0.88	0.094	0.085
	ch	0.18	0.19	0.002	0.003
	all	0.70	0.69	0.066	0.063
2002	nch	1.01	0.96	0.107	0.095
	ch	0.19	0.21	0.003	0.005
	all	0.77	0.76	0.078	0.071
2003	nch	1.08	0.96	0.119	0.095
	ch	0.20	0.22	0.004	0.004
	all	0.84	0.77	0.088	0.072
2004	nch	1.13	0.93	0.127	0.093
	ch	0.21	0.22	0.003	0.004
	all	0.91	0.77	0.097	0.072
2005	nch	1.05	0.93	0.113	0.090
	ch	0.23	0.23	0.003	0.004
	all	0.87	0.78	0.090	0.072
2006	nch	1.15	0.91	0.127	0.090
	ch	0.25	0.26	0.005	0.006
	all	0.99	0.78	0.105	0.073
2007	nch	1.22	0.97	0.142	0.097
	ch	0.27	0.26	0.008	0.005
	all	1.05	0.83	0.117	0.078
2008	nch	1.23	0.99	0.140	0.102
	ch	0.34	0.27	0.011	0.007
	all	1.07	0.85	0.117	0.083
2009	nch	1.26	1.00	0.143	0.101
	ch	0.33	0.29	0.011	0.008
	all	1.10	0.88	0.121	0.085
2010	nch	1.29	1.00	0.146	0.100
	ch	0.37	0.32	0.013	0.009
	all	1.13	0.88	0.123	0.083
2011	nch	1.32	1.00	0.151	0.101
	ch	0.37	0.32	0.012	0.009
	all	1.18	0.89	0.130	0.086
2012	nch	1.32	1.00	0.151	0.102
	ch	0.40	0.35	0.015	0.011
	all	1.18	0.91	0.131	0.089
2013	nch	1.30	0.98	0.148	0.099
	ch	0.43	0.38	0.016	0.015
	all	1.19	0.91	0.131	0.089

Note: 'all' stands for 'all journals', 'ch' for 'Chinese journals', and 'nch' for 'non-Chinese journals'.

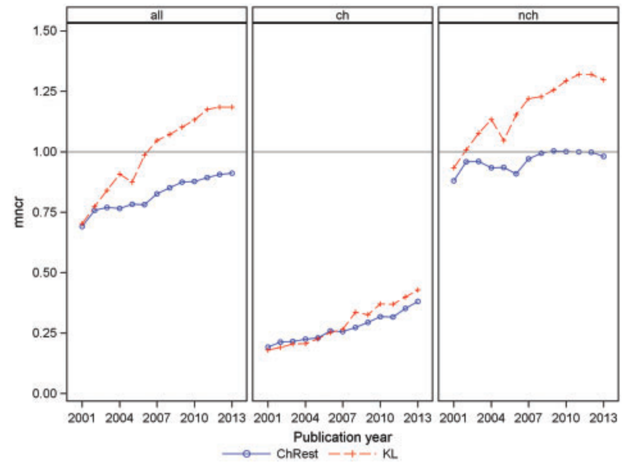

Figure 4. mncr for KL and ChRest by publication year and journal set.

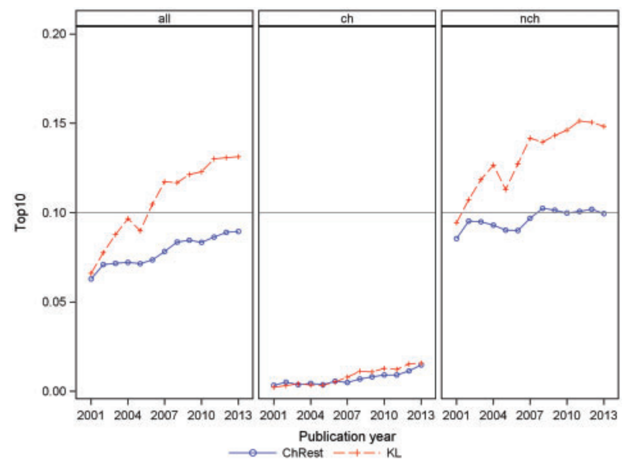

Figure 5. Top10 for KL and ChRest by publication year and journal set.

5. Conclusions

In this work, we revisited the role of the Chinese KLs in the international and national scientific arena. An obvious difference between now (2016) and then (2002) is the increase in the number of KLs: from 162 to 258.

We began this study with a short overview of the position of KLs in China, including their budget and manpower. We also briefly recalled the procedures applied during national evaluation exercises. As these have changed only in a minor way since 2004, we did not go deeply into this aspect.

As in the earlier investigation, data sources were the WoS for international aspects and the CSCD for local Chinese aspect. For the WoS, we used Bibmet, the bibliometric version of the WoS at KTH Royal Institute of Technology (Sweden), which allowed, among other things, field normalization of citations.

As to publications in the WoS, we found that the contribution of KLs compared with the ChRest is slightly and irregularly increasing (using full as well as fractional counting). When using full counting, KLs contribute to about 23% of the publications (in 2014); this means that in about 23% of all 2014 publications with at least one Chinese address in the WoS, at least one of these addresses is a KL address. When using fractional counting, KLs contribute about 12% of China's publications in the WoS for the same year. These Chinese contributions contain about the same percentage of non-Chinese collaborations over the considered years. We observed an increase in the number of collaborations between KLs and Chinese colleagues. Compared with the previous investigation (Jin, Rousseau and Sun 2006), we observed a fivefold increase in the absolute number of KL publications and an increase from 18.7% (in 2004) till 23% (in 2014) in full counting. A limitation of our study, however, is the possibility of database artifacts with regard to publication output. More and more journals are added to WoS over time. For instance, if several Chinese journals are included in year *Y*, then this gives rise to an artificial increase in publication output for China from year $Y - 1$ to *Y*. We compared, however, the relative increase in the number of Chinese WoS publications from year $Y-1$ to year Y (Table 2) with the corresponding increase in the number of publications in Chinese WoS journals. The outcome is that that the number of publications from China generally increase substantially more than the number of publications in Chinese journals. This suggest that the increase of Chinese journals in WoS is not a major cause of the observed increase of Chinese WoS publications.[10]

We made the surprising observations that non-KLs are more internationally oriented in terms of publications as well as received citations. One possible explanation is that most Chinese journals included in the WoS are hosted by KLs, and it might be the case that scientists affiliated to these labs often prefer to publish in KLs journals. Yet, we have not investigated this in detail and leave this as a topic for further research.

Regarding publications in CSCD during the period 2001–14, the relative number of KL publications increased, irrespective of the counting method. The full counting value reached 8.8% in 2014, whereas the corresponding fractional counting value reached 6.1%. Further, the percentage of collaboration publications also increased over the years. KLs have a responsibility to promote the development of Chinese research within their fields, and the number of publications performed in collaboration with other institutions is used in China for their evaluation.

Comparing indicator values for KL across WoS and CSCD, we find that the values are generally, and substantially, higher when the underlying data are obtained from the WoS. For instance, the Ratio cit. values (including self-citations) in 2014 are 25.11 for the WoS and 8.05 for CSCD. The standing of KL relative to the ChRest concerning publication output and received (raw) citations is therefore stronger in the international database WoS compared to the domestic database CSCD.

In the older investigation we made the observation that in terms of citations KLs, Open Labs and the other institutes in China were roughly performing on par, although KLs are supposed to harbor the best scientists. Now, we had a new look at this observation using field normalized citation indicators. This led to the expected results that researchers at KLs perform considerably better than other Chinese colleagues and, moreover, perform clearly better than database average. For instance, we note that in 2013, 13.1% of KL publications belong to the 10% best of their field. As such, we may conclude that, finally, KLs have lived up to their promise and made real impact on the international arena.

For practical reasons, we use an open citation window in the second part of the study. Assume that a unit of analysis has a constant citation impact over time, relative to the publications of its reference groups, but that the short-time impact is lower than the long-term impact. It might then look like the impact for publications (from the unit) published in the later years is less than the impact for publications published in the earlier years. This could occur *even if* field normalization is used. For instance, it could occur if citation peaks vary considerably across countries. However, we did not find any reason to believe that the results are significantly affected by this limitation of the study.

In this work, a classification scheme is used for field normalization of citations, and the fields used are the journal subject categories of the WoS. However, this traditional approach to field normalization has several drawbacks (Waltman and van Eck 2013b). For instance, it is possible that the subfields of a certain field, where the fields are defined at a given level of granularity, differ substantially from each other in terms of citation volumes. The bibliometric group at KTH Royal Institute of Technology (Sweden) has recently, on the basis of a methodology put forward by Waltman and van Eck (Waltman and van Eck 2012, 2013a), implemented a (hierarchical) publication-level classification system. About 28 million publications have been algorithmically grouped into classes based on direct citation links between them. Concerning future research, it would be interesting to use this publication-level classification system for field normalization to gain further insight on the citation impact of KLs compared to the rest of Chinese research.

Acknowledgements

P.A. and R.R. thank their colleagues at the National Science Library of CAS for their hospitality during their visits. The authors wish to thank two anonymous reviewers for giving valuable comments on an earlier version of the article. Work of R.R. is partially supported by a grant of the National Natural Science Foundation of China (grant number 71573225).

Notes

1. Potentially, the objective of doing research on frontiers of science might conflict with the objective of taking major tasks for the Chinese government. However, we have no evidence that this is a serious problem.
2. Similarly, false positives, i.e. addresses that are retrieved by the used regular expression approach and that refer to non-KL units, are hardly likely to occur.
3. If the used regular expression approach, which is applied to the full addresses of the publications, detects two or more KL addresses in a publication, it is not likely that these addresses are spelling variants that refer to the same underlying KL unit (which would overestimate the KL address fraction for the publication, given that there is at least one non-KL full address in the publication), admitting that this is theoretically possible. However, such variants occur across publications, but this is not a problem for our approach.
4. However, due to incomplete cited reference data in CSCD with regard to publications published before 2005, we restrict the publication period used to calculate Ratio cit. to 2005–11.
5. http://www.leidenranking.com/
6. Note that weights are used for the citation distributions considered in the calculation of Top10 values. Each citation value in a given distribution is assigned the weight $1/k$, where k is the number of subject categories of the corresponding publication. The weight is the fraction with which the publication contributes to each of its subject categories. The proportion publications with less than c citations are then the sum of the weights for the citation values that are less than c, divided by the sum of weights for all the citation values in the distribution.
7. As mentioned in the section 'Data and methods', analyses were performed also on the basis of the data set obtained from the full data set by removing all Hong Kong and Macao addresses. For this reduced data set, the pattern of ratios of publications in non-Chinese journals over

publications in Chinese journals turned out to be similar to the corresponding pattern for the full data set.

8. Compared to Table 3, we omit the columns 'foreign frac' and '% foreign'. The rationale behind this omission is that few publications in the CSCD involve foreign addresses.
9. A possible explanation of the drop could be that the subject profiles of KL for 2004 and 2005 are considerably different. We measured the Pearson correlations between publication years with regard to the distributions of KL publication fractions over WoS subject categories. No support for the indicated hypothesis was found, though.
10. For an outline of the changed inclusion policy regarding WoS and regional journals from 2005 onward, see Testa (2011). This publication reports that the number of Chinese journals in WoS increased by 62 from 2005 to 2010.

References

Ahlgren, P., and Sjögårde, P. (2015) *Formal Definitions of Field Normalized Citation Indicators and their Implementation at KTH Royal Institute of Technology* <http://www.kth.se/polopoly_fs/1.544479!/Formal definitions of field normalized citation indicators at KTH.pdf> accessed 1 June 2016.

Benner, M., Liu, L., and Serger, S. S. (2012) 'Head in the Clouds and Feet on the Ground: Research Priority Setting in China', *Science and Public Policy*, 28/2: 258–70.

Braun, T., and Glänzel, W. (1990) 'United Germany - the New Scientific Superpower', *Scientometrics*, 19/5–6: 513–21.

Fan, X., Yang, X. W., and Chen, L. M. (2015) 'Diversified Resources and Academic Influence: Patterns of University-Industry Collaboration in Chinese Research-Oriented Universities', *Scientometrics*, 104/2: 489–509.

Fu, H. Z., and Ho, Y. S. (2013) 'Comparison of Independent Research of China's Top Universities Using Bibliometric Indicators', *Scientometrics*, 96/1: 259–76.

Gao, X., Guo, X., and Guan, J. C. (2014) 'An Analysis of the Patenting Activities and Collaboration Among Industry-University-Research Institutes in the Chinese ICT Sector', *Scientometrics*, 98/1: 247–63.

He, T. W., and Liu, W. (2009) 'The Internationalization of Chinese Scientific Journals: A Quantitative Comparison of Three Chemical Journals from China, England and Japan', *Scientometrics*, 80/3: 583–93.

Hopcroft, J. E., Motwani, R., and Ullman, J. D. (2007) *Introduction to Automata Theory, Languages, and Computation*, 3rd edn. Boston: Pearson/Addison Wesley.

Huang, Y. L. et al. (2015) 'Bibliometric Analysis for the Literature of Traditional Chinese Medicine in PubMed', *Scientometrics*, 105/1: 557–66.

Jin, B. H., Rousseau, R., and Sun, X. X. (2005) 'Key Labs and Open Labs in the Chinese Scientific Research System: Qualitative and Quantitative Evaluation Indicators', *Research Evaluation*, 14/2: 103–9.

Jin, B. H., Rousseau, R., and Sun, X. X. (2006) 'Key Labs and Open Labs in the Chinese Scientific Research System: Their Role in the National and International Scientific Arena', *Scientometrics*, 67/1: 3–14.

Jin, B. H., and Wang, B. (1999) 'Chinese Science Citation Database: Its Construction and Application', *Scientometrics*, 45/2: 325–32.

Ministry of Science and Technology of China. (2016) *Annual Report of Basic Research for China*, in press.

Moed, H. F., De Bruin, R. E., and van Leeuwen, T. N. (1995) 'New Bibliometric Tools for the Assessment of National Research Performance: Database Description, Overview of Indicators and First Applications', *Scientometrics*, 33/3: 381–422.

Rousseau, R. (2015) 'The Tip of the Chinese Publication Iceberg', *ISSI Newsletter*, 11: 100–2.

Sun, Z. M., and Hua, W. N. (2015) 'A Comparative Study of Ocean Engineering Research Between China and the World', *Scientometrics*, 105/1: 51–63.

Testa, J. (2011) The Globalization of Web of Science: 2005-2010 <http://wokinfo.com/products_tools/multidisciplinary/webofscience/contentexp/expansionessay> accessed 9 Oct. 2016.

Tsou, C. L. (1989) 'The National Laboratories of China', *Faseb Journal*, 3/12: 2443–4.

van Raan, A. F. J. (1996) 'Advanced Bibliometric Methods as Quantitative Core of Peer Review Based Evaluation and Foresight Exercises', *Scientometrics*, 36/3: 397–420.

Waltman, L., and Schreiber, M. (2013) 'On the Calculation of Percentile-Based Bibliometric Indicators', *Journal of the American Society for Information Science and Technology*, 64/2: 372–9.

Waltman, L., and van Eck, N. J. (2012) 'A New Methodology for Constructing a Publication-Level Classification System of Science', *Journal of the American Society for Information Science and Technology*, 63/12: 2378–92.

Waltman, L., and van Eck, N. J. (2013a) 'A Smart Local Moving Algorithm for Large-Scale Modularity-Based Community Detection', *European Physical Journal B*, 86/11: 471.

Waltman, L., and van Eck, N. J. (2013b) 'Source Normalized Indicators of Citation Impact: An Overview of Different Approaches and an Empirical Comparison', *Scientometrics*, 96/3: 699–716.

Waltman, L. et al. (2011a) 'Towards a New Crown Indicator: An Empirical Analysis', *Scientometrics*, 87/3: 467–81.

Waltman, L. et al. (2011b) 'Towards a New Crown Indicator: Some Theoretical Considerations', *Journal of Informetrics*, 5/1: 37–47.

Wang, S. H., Wang, H. J., and Weldon, P. R. (2007) 'Bibliometric Analysis of English-Language Academic Journals of China and their Internationalization', *Scientometrics*, 73/3: 331–43.

Yan, S. L., Rousseau, R., and Huang, S. Q. (2016) 'Contributions of Chinese Authors in PLOS ONE', *Journal of the Association for Information Science and Technology*, 67/3: 543–9.

Yan, S. M., and Sun, J. Q. (2015) 'Assessing China's Salt Lake Resources R&D Based on Bibliometrics Analysis', *Scientometrics*, 105/2: 1141–55.

Yang, L. et al. (2015) 'Growing Trend of China's Contribution to Haze Research', *Scientometrics*, 105/1: 525–35.

Zhou, P., and Leydesdorff, L. (2011) 'Fractional Counting of Citations in Research Evaluation: A Cross- and Interdisciplinary Assessment of the Tsinghua University in Beijing', *Journal of Informetrics*, 5/3: 360–8.

Zhu, J. et al. (2014) 'Measuring Recent Research Performance for Chinese Universities Using Bibliometric Methods', *Scientometrics*, 101/1: 429–43.

Zhu, X. et al. (2004) 'Highly Cited Research Papers and the Evaluation of a Research University: A Case Study: Peking University 1974-2003', *Scientometrics*, 60/2: 237–47.

Appendix

Appendix 1: The Waltman–Schreiber approach to calculating percentile-based indicators

Example 1. Assume that, for a citation distribution of 105 publications belonging to a given field, publications of the same document type and published the same year, a given unit of analysis, say U, has 10 publications: nine publications with zero citations and one publication with 10 citations. We further assume that five of the 105 publications have 20 citations, two publications have 18 citations, 10 publications have 10 citations, and the remaining 88 publications have zero citations.

Each of the 105 publications represents $1/105 = 0.952\%$ of the citation distribution of the field. Then the five publications with 20 citations, together with the two publications with 18 citations, represent $7 \times 0.952\% = 6.664\%$ of the top 10% of the distribution.

Thus, $10\% - 6.664\% = 3.336\%$ needs to be covered by the 10 publications with 10 citations.

The 10 publications with 10 citations represent $10 \times 0.952\% = 9.52\%$ of the distribution. Now, each of these 10 publications is assigned to the top 10% with the fraction $3.336\%/9.52\% = 0.35$. This yields that 10% of the 105 publications, 10.5 publications, belongs to the top 10%, since $(10 \times 0.35 + 7)/105 = 10\%$. Finally, with regard to the unit U, the proportion of U publications among the Top10 is equal to $0.35/10 = 0.035$.