Web of Science Group

A Clarivate Analytics company

Research in Brazil: Funding excellence

Analysis prepared on behalf of CAPES by the Web of Science Group

Contents

This analysis has been prepared for CAPES by the Web of Science Group.

It provides a high-level overview of Brazilian research performance describing recent trends.

We are grateful to Professor Carlos Henrique de Brito Cruz, the scientific director of the São Paulo Research Foundation (FAPESP) and Mr. Abel Packer, director and co-founder of the Scientific Electronic Library Online (SciELO) for their contributions.

Executive summary

This analysis provides a snapshot of Brazilian research in a global context and highlights the significance of international and industry collaborations on research impact and visibility. Using bibliometrics to analyze data on Brazilian research papers published between 2013 and 2018, illustrative examples of analyses and evidence describe the research landscape and provide an indication of the potential impact of Brazilian research and science policy. The data are drawn from the *Web of Science* – the world's most trusted and largest publisher neutral citation index.

Productivity
Brazil ranks 13th in the world in terms of its output of research articles and reviews indexed in the *Web of Science*.

In 2018 alone, Brazilian researchers published more than 50,000 papers.

The growth in output is 30% over this six year period observed and is twice the global average.

There are pockets of excellence, in terms of citation impact, in Brazilian research, in the life sciences, physical sciences and engineering.

Citation Impact
The percentage of Brazilian papers in the world's top one percent of most highly cited papers is consistently more than one percent during the six-year period, 2013-2018.

The Category Normalized Citation Impact (CNCI) of this research is below the world average, but steadily improving.

Peer Comparisons
Compared to other BRICS countries, Brazil's research output, in terms of the number of papers indexed in the *Web of Science* is average.

Compared to neighboring countries in Latin America, Brazil's research output has a lower CNCI.

Collaboration: International
Between 2013 and 2018, Brazilian researchers collaborated with researchers from 205 countries.

Approximately a third of Brazil's research papers are co-authored with researchers from other countries.

Growth in the % of papers with an international co-author has slowed from and increase of 17.5% between 2013 and 2015 to an increase of 1.8% between 2016 to 2018.

Collaboration: Industry
Strategies to promote university-industry research collaborations are bearing fruit in terms of the number of papers published.

81% of the joint university industry publications for the period 2015-2017 where collaborations between public universities and industry.

Universities and Research Institutes
Public universities are the main source of research publications in Brazil.

The 15 universities with the highest research output, all public, produce over 60% of the total research output.

Background

In the last decade, Brazil has seen sizable growth in its research enterprise, maintaining its status as one of the BRICS countries, alongside Russia, India, China, and South Africa, and as an emerging economy demonstrating rapidly growing scientific productivity and influence measured through bibliometrics. Brazil is the fifth most populous country in the world, and remains in the top 10 world economies, despite recent economic turmoil.

In this updated analysis the 2017 research profile of Brazil is revisited[1] to examine where Brazilian researchers continue to be active and to identify areas of excellence. Over the last decade there has been a growing interest in Brazil's research output focused on both the national level[2] and on specific fields of research.[3] Remarking on the rapidly changing research landscape in Brazil, Leta, Thijs, and Glänzel (2013) note that "The huge growth in the Brazilian production of publications constitutes the greatest potential that goes far beyond the Latin American region."

The 2017 report documented the state of research output indexed in the *Web of Science*. In this analysis we explore *Web of Science* data[4] with a focus on international and university-industry collaborations and the research output of universities and research institutes to offer illustrative examples of how different analyses can generate evidence to help inform policy and decision-making.

We also examine some results from the Scientific Electronic Library Online (SciELO), which was created and implemented in Brazil in 1998 to strengthen research infrastructure by developing scholarly communication capacity. SciELO currently includes 296 Brazilian journals, of which 99 are high quality, influential journals indexed in the *Web of Science*.

When studying the scientific output of a country such as Brazil, it is advisable to focus bibliometric analysis on documents within the *Web of Science* that are classified as articles or reviews as these are peer-reviewed publications that reflect mature research findings.

[1] https://www.capes.gov.br/images/stories/download/diversos/17012018-CAPES-InCitesReport-Final.pdf accessed August1, 2019. Adams J. and King C. (2009) Global Research Report: Brazil - Research and Collaboration in the New Geography of Science. https://www.slideshare.net/nielsleidecker/grr-brazil-jun09-1

[2] Glänzel, W., Leta, J. Thijs, B. (2006) Science in Brazil Part 1: A macro-level comparative study. *Scientometrics,* 67 (1), 67–86. Leta, J, Thijs, B, & Glänzel, W. (2013) A macro level study of science in Brazil: seven years later. *Bibli Encounters: Electronic Journal of Library and Information Science* [Online], accessed 25 Aug. 2019

[3] Leta, J., Glänzel, W., Thijs, B. (2006), Science in Brazil. Part 2: Sectorial and institutional research profile. *Scientometrics*, 61 (1), 87-105. Andrade Vargas, R.; de Souza Vanz, S.A; Chittó Stumpf, I.R. (2015) Brazilian agricultural research in the *Web of Science*: a bibliometric study of scientific output and collaboration (2000-2011). *Em Questão,* 21 (3), 296-318.

[4] Throughout this document, the analysis is restricted to articles and reviews, indexed in the Science Citation Index Expanded (SCIE), the Social Sciences Citation Index (SSCI) and the Arts & Humanities Citation Index (AHCI) of the *Web of Science*. Whenever mentioned, additional analyses are based on the Brazilian collection of SciELO Citation Index, hosted on the *Web of Science* platform.

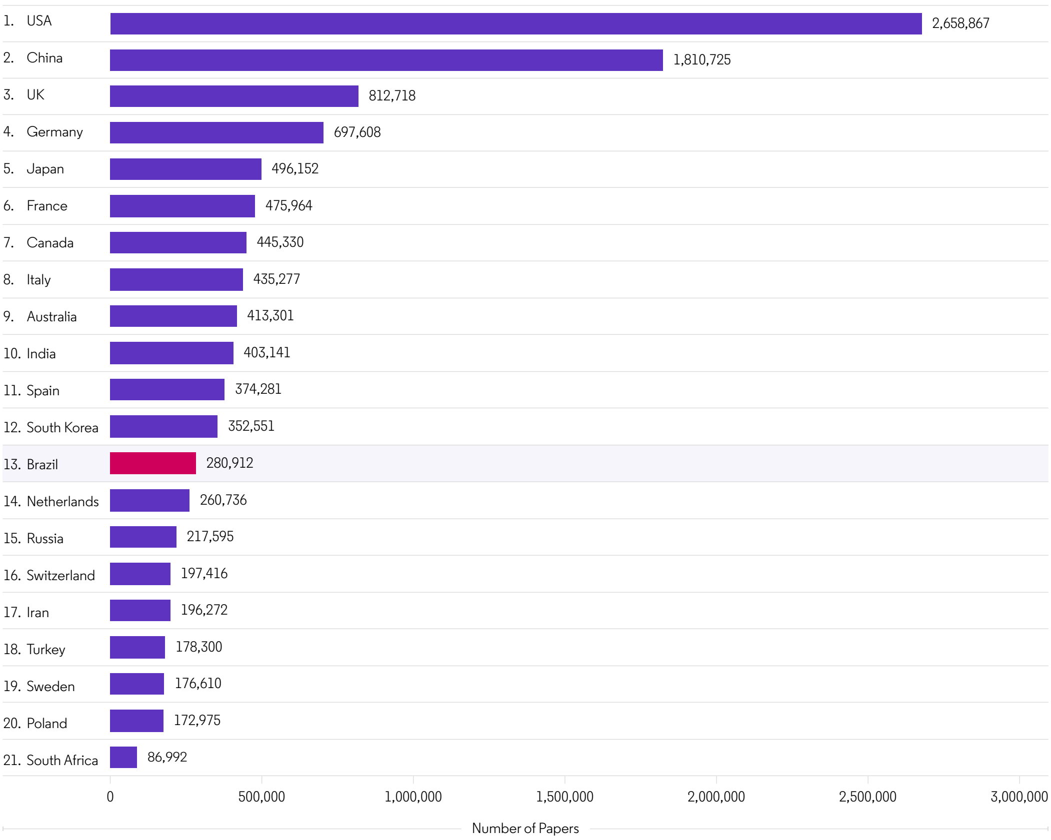
How does Brazilian research perform?

An overview of Brazilian research performance in recent years

How many research papers does Brazil produce?

Brazil was ranked 13th in the world in terms of its output of research papers between 2013 - 2018, just behind India (10th) and South Korea (12th), and ahead of Russia (15th) and South Africa (21st) (Figure 1). It is worth noting that despite changing economic conditions, the growth in Brazilian research output has been consistently strong over this six-year period.

Figure 1

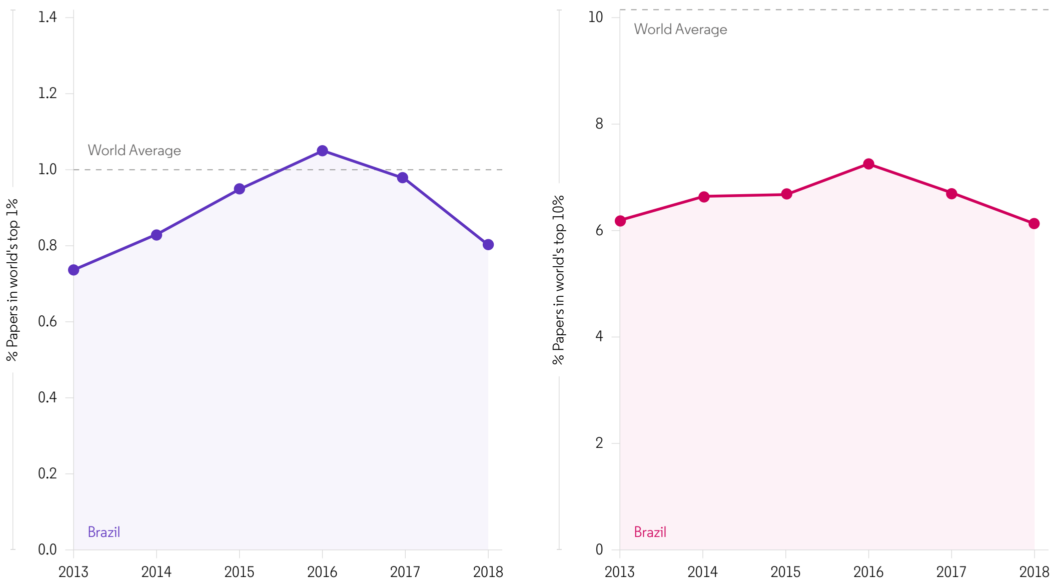


Papers added to *Web of Science*, 2013 – 2018.

This growth, 30% over this six-year period, has been twice that of the global average (15%). In 2018 alone, Brazilian researchers published over 50,000 papers indexed in the *Web of Science*.

In addition to the 280,912 papers indexed in the *Web of Science* (2013-2018), Brazilian authors published more than sixty thousand articles and reviews indexed in SciELO journals that are not covered in the *Web of Science*. Bibliometric studies often use the world's top one and ten percent of most highly cited papers as proxy indicators for the production of excellent research. In Figure 2, the left panel shows the percentage of papers published by Brazil that are in the world's top one percent of most highly cited papers, and the right panel shows the percentage of papers published by Brazil that are in the world's top ten percent. When looking at the top ten percent of highly cited work, Brazil is below the world average and the decrease in recent years seems more prominent. This seeming drop is not unique to Brazil as larger drops in this measure are seen in recent data from countries that have extensive research outputs. It is important to emphasize that care be exercised in the interpretation of recent citation counts from 2018, which tend to be volatile and the apparent deviations from the norm may be artefacts of recency of the data rather than dramatic departures from persistent trends.[5]

Figure 2



Percentage of Brazilian papers in the top one percent and top ten percent of most highly cited papers, 2013 – 2018 (world average denoted by the dotted line).

[5] Adams, J. (2018) Information and misinformation in bibliometric time-trend analysis. *Journal of Informetrics* 12(4):1063-1071

International comparisons

What is the citation impact of Brazilian research?

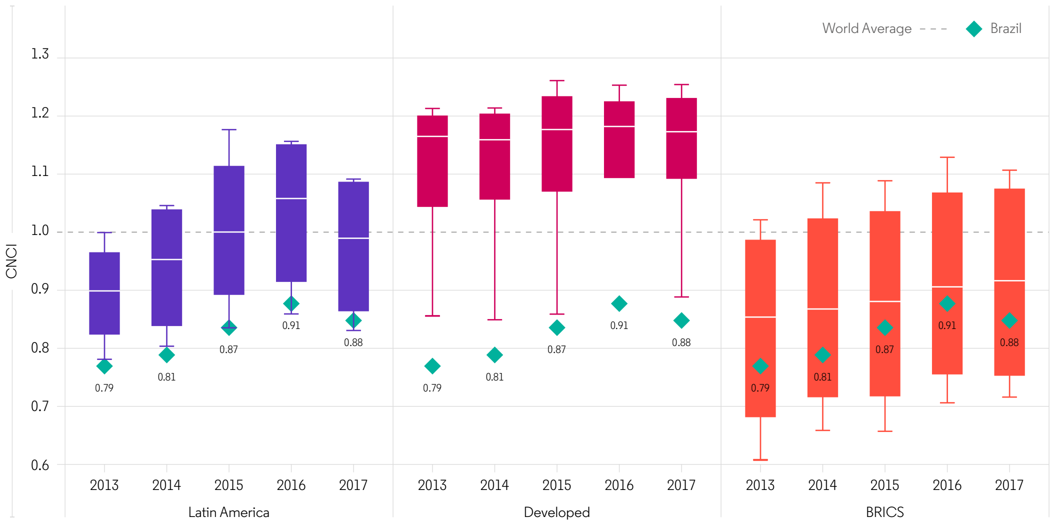
As Brazil's research output grew between 2013-2016 the impact of its research, as indicated by the number of citations received, also grew steadily. Figure 3 presents a selective global perspective on the distribution of the Category Normalized Citation Impact (CNCI)[6].

Brazil's annual CNCI reached a peak in 2016 at 0.91, compared to a world average of 1.0. There is an observed drop however, in Brazil's CNCI in 2017.

Other Latin American[7] countries also exhibit this trend. This trends it is not replicated in developed[8] countries. This trends it is not replicated in the BRICS[9] as the dramatic proportional increase in publications and high impact publications from China masks effects in other countries.

This downturn in 2017 is not due to a change in the coverage of Brazilian journals in the *Web of Science*. One potential explanation for the decrease in CNCI is the reduced growth in the number of papers with one or more international co-authors. It is widely recognized that international collaborations enhance the citation impact of an article[10]. In the three years, 2013 to 2015, the percentage of papers with an international co-author increased by 17.5%, in the most recent three years 2016 to 2018 there has only been 1.8% growth. As mentioned above, a more likely explanation of this seeming downturn in Brazilian CNCI is that it is an artefact of the volatility of recent data.

Figure 3



Category Normalized Citation Impact (CNCI) for Brazil[11] and comparator countries by groups, 2013 – 2017.

[6] See the appendix for additional details regarding the computation of the CNCI.

[7] Argentina, Chile, Colombia, Mexico

[8] Canada, France, Germany, Japan, Spain

[9] Russia, India, China, South Africa

[10] Katz, J.S. & Hicks, D. *Scientometrics* (1997) 40: 541. https://doi.org/10.1007/BF02459299, accessed August 21, 2019. Adams, J. & Loach, T. (2015). Comment: A well-connected world. *Nature*. 527. S58-S59.

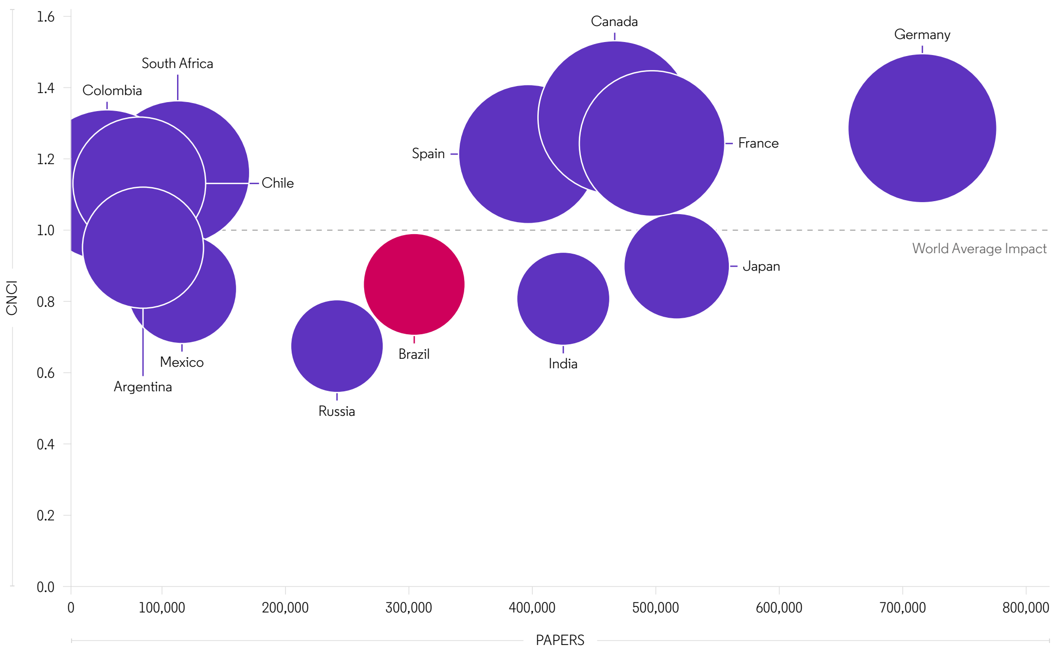
[11] The diamonds and the numbers above the diamond indicate Brazil's CNCI

How does Brazilian research compare internationally?

Several factors, such as the strength of the economy, level and focus of research funding, international collaborations and the caliber of educational and research institutions, can affect the quantity and influence of a country's research output. Different countries vary in their approaches to science and technology policy, so although it is difficult to tease out the individual contribution of any of these factors, it could be instructive to compare productivity and research impact across countries to gain insight into the relationship between these factors and research outputs. For instance, China's recent investments in higher education and research have led to exponential growth in the quantity and quality, as measured by the number of citations, of research output.

In this analysis of Brazil and its comparator countries, China, which has published at least three times more papers than Brazil or any other comparator country, is an extreme outlier, which distorts the visualization of the data. China data is included in the analysis but excluded from the visualisation Figure 4.

Figure 4



Output and Category Normalized Citation Impact (CNCI) of all comparator countries excluding China.
The diameter of the circles is proportional to the percentage of papers in the world's one percent most highly-cited papers, 2013 - 2018.

Compared to the other BRICS countries Brazil's research output is average, with more publications than those from South Africa and Russia and a higher citation impact than Russia and India. China has higher citation impact, above the world average, and vastly greater output of peer reviewed papers. Compared to developed countries Brazil has a lower citation impact. Brazil's research impact is only slightly behind Japan's which is the only comparator developed country with a below world average citation impact (Figures 3, on page 9 and figure 4, on page 10).

Figure 4 also shows that Brazil publishes many more papers than other Latin American countries, though the other countries have higher citation impact and proportion of papers in the top one percent of papers worldwide. The higher citation impact is probably derived from international collaboration. Chile and Colombia produce most of their research, 67% and 67.5% respectively, in collaboration with researchers from other nations while this percentage for Brazil is 36% over the same period. Additional analyses of trends would provide useful insights into Brazil's relative position and how it has evolved over time with respect to other countries.

International collaboration and its citation impact

Who does Brazil collaborate with internationally?

Brazilian researchers collaborate with scholars from all over the world. Over the six-year period, 2013-2018, they co-authored papers with researchers from 205 countries, which represented approximately a third of all the papers in the *Web of Science* with Brazilian authors. These collaborations included established research-intensive nations such as the G7, neighboring countries in Latin America, as well as fellow BRICS countries. As mentioned earlier, international collaborations usually have a higher citation impact.

The highest number of international co-authored papers are published with authors in the United States, although these papers do not deliver the highest citation impact. Some of the highest citation impacts are associated with publications with partners in fellow BRICS countries, such as China and India, where research output has been growing rapidly in recent years. Here again, a more detailed analysis of the papers with multiple co-authors suggests a qualitative difference between collaborations that arise out of work conducted at hyper-collaborative research institutions such as CERN where papers with hundreds of co-authors garner thousands of citations versus papers that are the outcome of small groups of researchers working collaboratively on relatively small projects.

Brazil is also developing more regional collaborations within Latin America as noted in the Global Research Report on Brazil (Adams & King, 2009)[12]. Investments in projects such as the Brazilian synchrotron light laboratory, the only source of synchrotron light in Latin America, enhance opportunities for international collaborations.

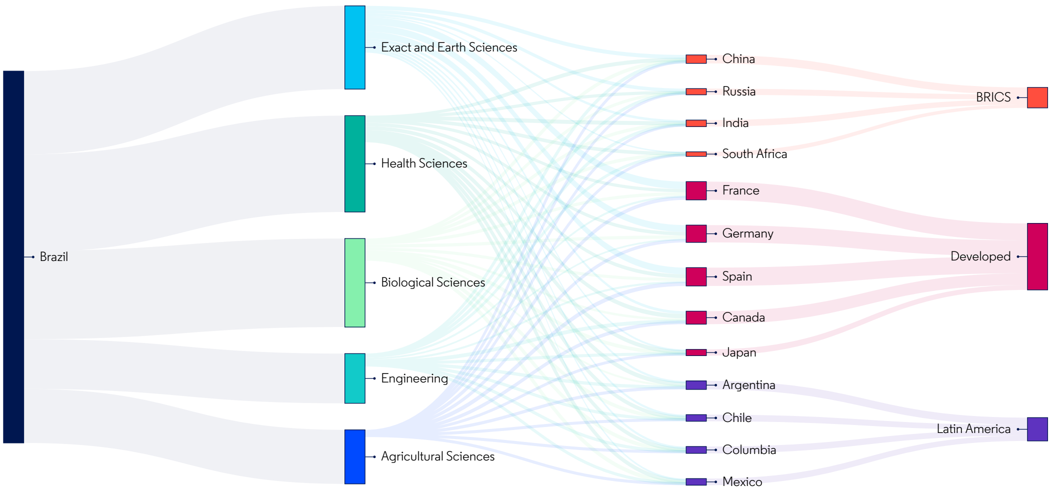
[12] Adams J. and King C. (2009) Global Research Report: Brazil - Research and Collaboration in the New Geography of Science. https://www.slideshare.net/nielsleidecker/grr-brazil-jun09-1

This laboratory is the research home for approximately 1,200 Brazilian and foreign researchers involved in hundreds of projects and co-authored publications[13].

Detailed analyses of Brazilian research output from international collaborations is presented in our 2017 CAPES report[14]. An alternative analysis and visualization of international collaborations is demonstrated in Figure 5, showing the proportion of papers published betweeen 2013 and 2018 in a select subset of the nine CAPES research categories relative to international co-authors countries of origin.

The bar on the left represents all papers published in this six-year period in five of the nine CAPES research categories with the largest numbers of publications. Linguistics, Literature and Arts, Humanities and Social Sciences, Applied Social Sciences and Multidisciplinary are excluded from this analysis. The distribution of the number of papers in each of the remaining five categories is shown by the size of the bars in the middle. The flows from the middle bars to the countries on the right are proportional to the number of papers with international co-authors from the countries on the right-hand side. There is some double counting in these flows because several papers have co-authors from multiple countries. The size of the rectangles is proportional to the number of co-authored papers, showing Spain with the most co-authored papers, which is notable as it has a smaller overall output of papers compared to the other developed countries and some of the BRICS.

Figure 5



International collaboration with select countries by different research area

[13] https://www.lnls.cnpem.br/the-lnls/about/ - accessed August 17, 2019.

[14] https://www.capes.gov.br/images/stories/download/diversos/17012018-CAPES-InCitesReport-Final.pdf - accessed August 17, 2019

Among these comparator countries, developed countries have the largest number of collaborations with Brazil. However, the number of collaborative papers does not seem proportional to the partner country's total publication output. If this were so, it would be expected that China would have the most collaborative papers. It appears that the number of collaborations depends upon existing research networks and with researchers in partner countries with a history of excellence in science.

In proportional terms, the fewest collaborations are in Agricultural Sciences and the most in Exact and Earth Sciences. For each partner country, Exact and Earth Sciences is usually the most productive research area, with at least 1,000 papers per partnership. The highest number of papers are with European countries, with France collaborating on 6,381 papers in Exact and Earth Sciences, the most papers in a single discipline.

A large proportion of collaborations in Health Sciences and Biological Sciences tend to be with the US (40 to 50%) followed by European countries and Argentina, with BRICS countries playing a much smaller role.

Across all research areas, Japan has the lowest number of papers in collaboration with Brazil. These low numbers are consistent with Japan's lower overall percentage of papers with international co-authors (31.8%) compared to the other developed partners which have 50% or more.

Latin American countries have lower numbers of collaborative papers compared to Brazil's partner countries from other parts of the world. Argentina collaborates most frequently with Brazil in Agricultural Sciences (439 papers) and Biological Sciences (1,762). Across the region, the most collaborations are in Exact and Earth Sciences and Health Sciences and the least in Engineering.

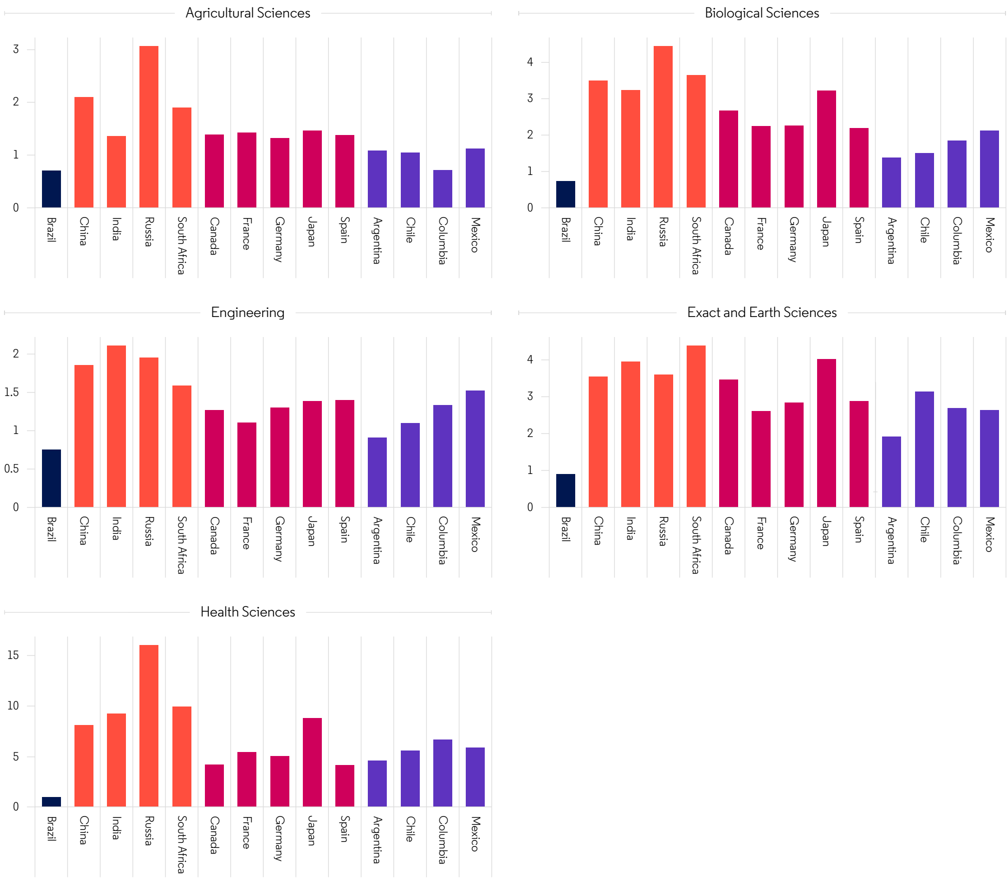
In SciELO, international collaborations have grown steadily over the past few years and are close to 10% of all publications in 2018. The most frequent collaborations, indexed in SciELO, are with researchers in USA, Portugal, China, Argentina, Spain and Turkey.

BRICS countries collaborate infrequently in Agricultural Sciences and Biological Science. There is greater collaboration by China and Russia in Exact and Earth Sciences and with China and India in Health Sciences.

What is the citation impact of international collaborations?

The citation impact of papers with international co-authors is generally higher than that of papers published by only Brazilian authors.[15] For papers with Brazilian and international co-authors this number is consistently high (Figure 6), and above the world average of 1.0. The few exceptions where the CNCI is below 1.0 are for collaborations with Latin American countries in Agricultural Sciences (Colombia, 0.72) and Engineering (Argentina, 0.91). The lowest CNCI numbers are in Agricultural Sciences where they range between 0.72 and 3.17 (Russia). Even in Agricultural Sciences, CNCI is extremely high for BRICS collaborations. Papers published in Health Sciences have the highest citation impact ranging from 4.13 (Spain) to the remarkably high score of 15.51 (Russia).

Figure 6



Category Normalized Citation Impact (CNCI) of Brazilian papers with co-authors from select countries by research area

These high citation impact numbers should be treated with caution. The publications underlying these numbers include several hyper-collaborative papers with many authors and many affiliated countries reporting on topics such as particle physics experiments, astronomy, medical guidelines or global impact of diseases. These papers are often cited thousands of times. Brazilian citation impact tends not to be driven by regional collaboration; rather it is concentrated in papers with co-authors from further afield, such as the United States, Europe, and other parts of the world.

[15] The link between co-authorship and citation impact is a phenomenon that has been widely observed, and is discussed in Moed, H. (2005) Citation Analysis in Research Evaluation. Dordrecht: Springer pp285-290.

University – industry collaboration

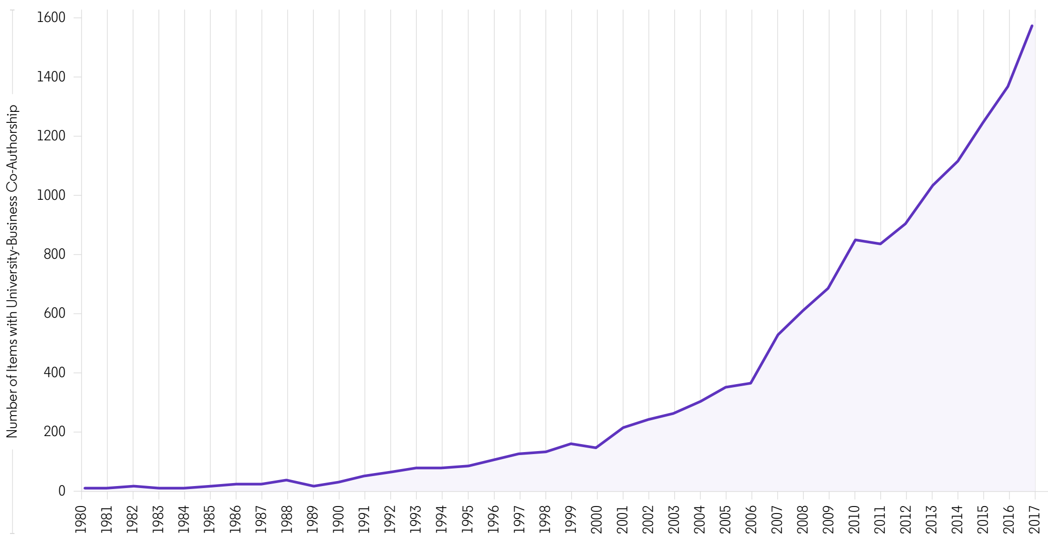
How has university-industry research collaboration in Brazil changed in 37 years?

University-industry interaction has been a central objective of Brazilian Science and Technology policy for many decades. Paradoxically, very few measurements of the quantity and intensity of this interaction have been performed. Here we contribute an analysis based on the number of scientific and technical publications that display co-authorship between researchers in universities in Brazil and researchers in industry[16] (for a discussion of the methods, see the Appendix 2). In order to provide a broader context for these university-industry collaborations, a longer-term view, going back to 1980 is informative, rather than the six-year window used in the rest of this anlysis.

The number of publications co-authored by researchers in universities and industry offers a window into ideas that were jointly created and developed by researchers in the two sectors indicating a higher level of engagement than mere consultation, contract research and development or research support and donations. Figure 7 shows the evolution of these research interactions since 1980, by counting the number of *Web of Science* items (all categories) that have at least one author from a university in Brazil and a co-author from industry, anywhere in the world.

Figure 8 shows the evolution of co-authorship between universities in Brazil and industrial sector authors since 1980 for the ten universities with the largest number of joint publications. The ten universities shown - all public - account for 81% of the joint publications for the period 2015-2017.

Figure 7

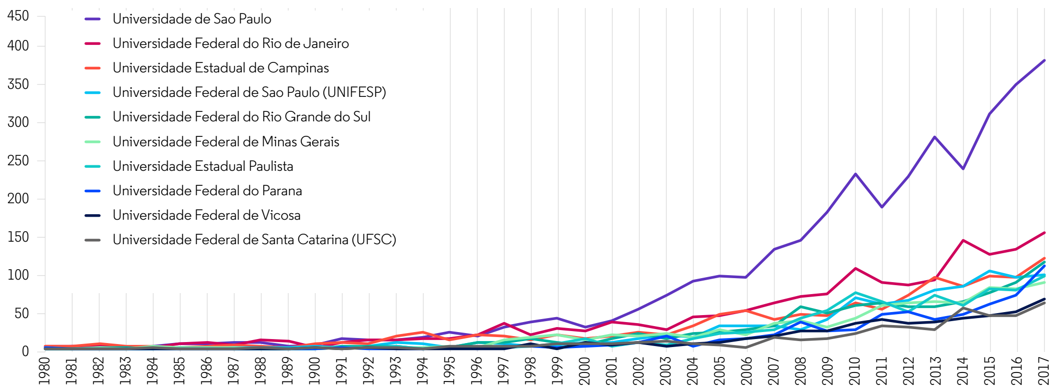


Number of *Web of Science* items with at least one author in a university in Brazil and at least one co-author from industry.

[16] The data presented here on university-industry collaboration includes a set of corporate entities which is much larger than that presently considered at the *Web of Science* or Incites. These data were obtained by performing an enhanced search procedure which included reclassifying the nature of entities in the list of publications with authors in Brazil and also searching for entities that have suffixes associated with industry.

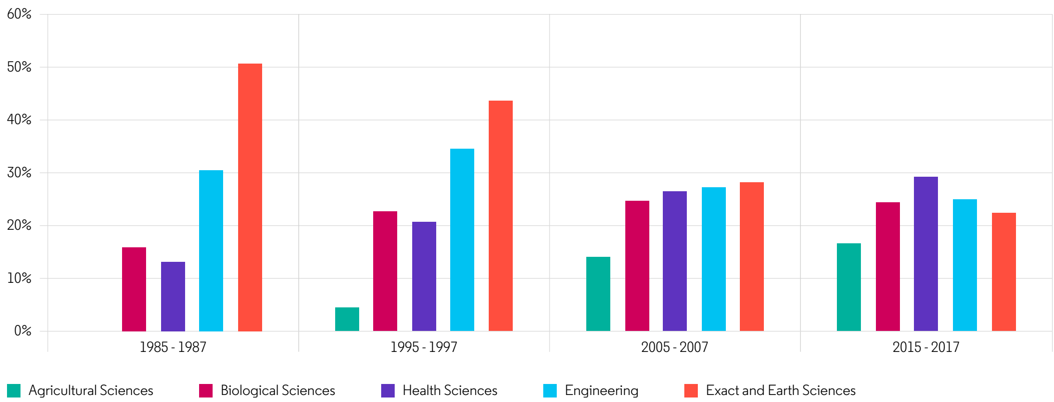
Which universities co-author more with industry?

Figure 8



Number of *Web of Science* items with at least one author in a university in Brazil and at least one co-author from industry for the ten universities with the largest number of items in 2017.

Figure 9



Classification of the publications in co-authorship between universities in Brazil and companies according to the 'Fields of Research' defined by CAPES.

The distribution of the university-industry co-authored publications among five of the nine fields of research defined by CAPES is shown in Figure 9, for four three-year periods: 1985-87, 1995-97, 2005-07, and 2015-17. Only five of the nine CAPES categories are shown here because of the small number of collaborations in Linguistics, Literature and Arts, Humanities and Social Sciences, Applied Social Sciences and Multidisciplinary categories. The distribution which was strongly dominated by the Exact and Earth Sciences 30 years ago has evened out over time across the five categories, especially, between the life and physical sciences for the 2015-17 period.

Which companies co-publish with universities in Brazil?

Petrobras, which has a strong program for developing research collaborations with academia, dominates the landscape of university-industry collaboration in Brazil (Figure 10). In the period 2015-2017, with 543 publications, the company participated in 14% of all university-industry collaborative output.

14%

Petrobas participated in 14% of all university-industry collaborative output.

Figure 10

The 25 companies that had the largest number of publications in co-authorship with universities in Brazil from 2015-2017.

Of the 50 companies with the highest levels of co-authorship with academia, 17 are of Brazilian origin and 33 are multinationals. There is a predominance of companies from the pharmaceutical sector (18 of 50), followed by the agricultural sector (12 of 50).

The creation of the Science and Technology Sector Funds, to finance domestic research, development, and innovation spurred green economy developments and university-industry bio-ethanol and bio-pharma research collaboration.[17]

[17] Financial Times Confidential Research, (2015) Brazil – Biopharma seeking a pick-me-up, http://www.scienceforbrazil.com/wp-content/uploads/2015/12/Brazil-%E2%80%93-Biopharma-seeking-a-pick-me-up.pdf accessed August 22, 2019. Corteza, LAB, Nogueirab, LAH, Lealc, MRVL, Junior, RB. 40 Years of the Brazilian Ethanol Program (Proálcool): Relevant Public Policies and Events Throughout Its Trajectory and Future Perspectives http://bioenfapesp.org/gsb/lacaf/documents/papers/05_ISAF_2016_Cortez_et_al.pdf accessed August 22, 2019.

Identifying excellence in Brazilian research

How does Brazilian research in different research areas compare?

To focus on research excellence, between 2013-2018, analysis of the nine categories defined by CAPES is insightful. Excellence in this context is measured by the Category Normalized Citation Impact (CNCI).

Among these nine CAPES categories, most papers by Brazilian researchers are published in life and physical sciences and engineering. The main mode for the dissemination of research results in these fields are journals, which are well represented in the *Web of Science* database. In contrast, in the arts, humanities and social sciences researchers are more likely to write books or policy documents, which are less well represented in this database.

Partly because of the small numbers, Brazilian research output has a high citation impact, above world average, in Applied Social Sciences and Humanities and Social Sciences (Figure 11). Because the CNCI is a weighted average, the highly cited papers in sub-categories of Theology (145 papers with an average CNCI of 2.37), Archaeology (224 papers with a CNCI of 1.24) and Anthropology (393 papers with a CNCI of 1.36) raise the overall research impact above the world benchmark of 1.0. Theology is notable in another sense in that only 18.6% of the papers have an international co-author even though its CNCI is the highest of any of the CAPES 121 sub-categories. Thus, the excellence is driven by domestic strength in the field.

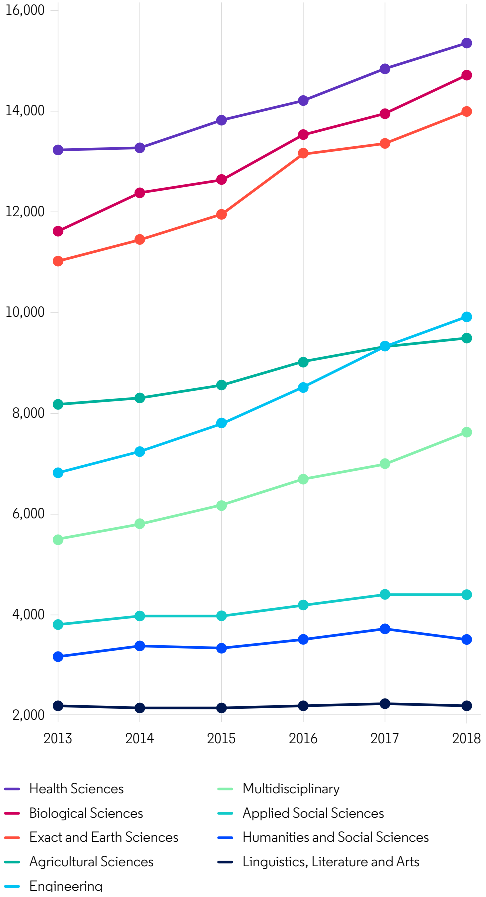
As shown in Figure 11, more than a third of the papers in Biological Sciences and almost half in Exact and Earth Sciences have foreign collaborators, which, as previously noted, contributes to a higher CNCI.

Figure 11

Research Area	Papers	CNCI	% International Collaborations
Health Sciences	82,406	0.96	34.7
Biological Sciences	75,717	0.74	37.0
Exact and Earth Sciences	71,214	0.90	45.1
Agricultural Sciences	46,222	0.71	21.7
Engineering	42,506	0.76	37.1
Multidisciplinary	30,190	0.82	39.5
Applied Social Sciences	14,229	1.03	30.1
Humanities and Social Sciences	9,581	1.00	30.5
Linguistics, Literature and Arts	953	0.68	19.1

Output and Category Normalized Citation Impact (CNCI) of Brazilian papers published between 2013 and 2018 in nine CAPES research categories

Figure 12



Trends in the number of papers published annually in different CAPES research categories between 2013 and 2018.

Brazilian research focus has evolved over six years (Figure 12). There has been increased output in all research areas except for Linguistic, Literature and Arts. The smallest increases have been in Health Sciences, (18.7%) and Agricultural Sciences (21.9%). Although not insubstantial, these increases seem small compared to Engineering, which has grown by nearly two-thirds, surpassing the output of Agricultural Sciences in 2018.

Delving deeper into the CAPES 121 sub-categories provides some insight into how this growth came about. For a long time within Engineering, Materials Sciences and Metallurgy has been a highly productive area that has seen rapid growth in recent years. There has also been considerable growth in the closely aligned area of Exact and Earth Sciences (chemistry, physics, etc.). Brazil's wide range of expertise and skills in these two related areas, and multidisciplinary research that bridges the two seem to have fueled this rapid growth.

Bibliometric analysis can be used to determine research areas that demonstrate strength, weakness, potential for opportunity, or areas that are under threat in a research portfolio. Figure 13 maps Brazil's research in the nine CAPES categories along two indices benchmarked on global output and citation impact.

The horizontal axis measures relative proportional output which is a ratio representing the proportion of a nation's research output relative to that research area's proportion of the world output. For instance, if Biology's share of research output in Brazil is 4.5% and the global share of Biology research indexed in *Web of Science* is 1.5% then Biology's score along the horizontal axis would be 3. The vertical axis is the Category Normalized Citation Impact (CNCI). Thus, research areas located in the top right-hand quadrant where the relative output and citation impact are both above 1.0 are considered areas of strength. Research areas located in the top left-hand quadrant where the output is relatively low, but citation impact is high are deemed potential opportunities. A research area is considered weak if it is in the bottom left quadrant where both relative output and citation impact are below world averages. Research areas in the bottom right-hand quadrant with relatively large output but low citation impact are under threat.

Areas of strength and opportunity

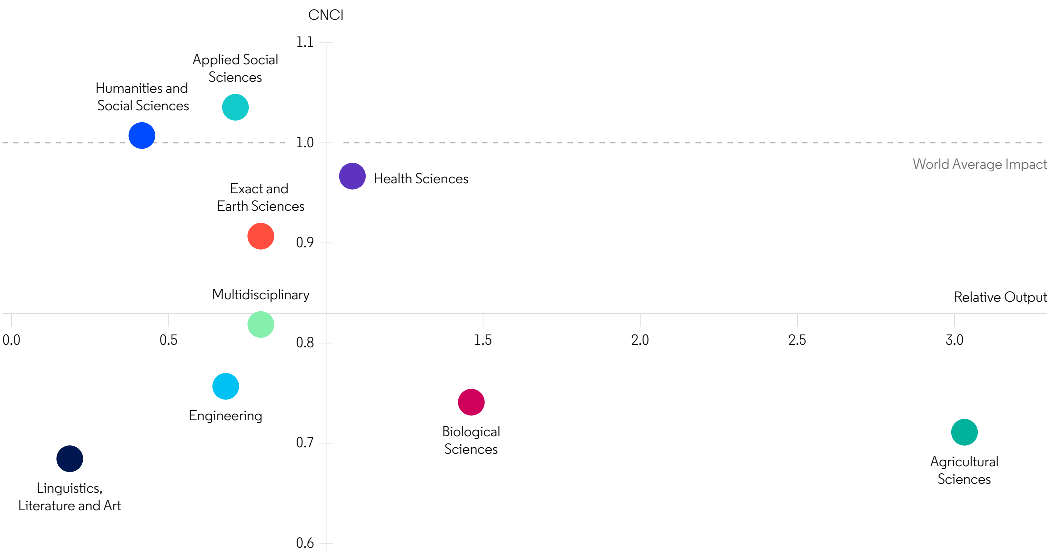
Health Science with its six-year relative output slightly above 1.0 and CNCI at 0.91 can be an area of potential strength, especially because papers published since 2015 have had above world average citation impact. In the bottom right, threatened areas are Biological Sciences and Agricultural Sciences, as Brazil does proportionately more research in these areas than the global average. Brazil is 3.5 times more productive than the world average in Agricultural Sciences though these papers have below average citation impact for the world and Brazil. However, agricultural research which is often of tremendous local importance does not always garner international attention. A low CNCI, as a measure of global citation impact, does not account for the impact of regionally relevant research.

Linguistic, Literature and Arts, and Engineering in the bottom left might be considered weak research areas. However, as already noted, Linguistics, Literature, and Arts its location in the bottom left might be an artefact of how "counting" is done in the *Web of Science*. Much of this research may be in Portuguese with a limited global readership and therefore, low coverage in the *Web of Science*. Thus, in global terms, this area is likely to appear "weak" because of its specialized nature and small research community.

With Sirius[18], the new Brazilian synchrotron light source, coming on line, more opportunities for Brazilian research exist in the Exact and Earth Sciences, where citation impact is above average for Brazil, but internationally it does proportionally less well than other research areas.

On the other hand, additional analysis must be conducted to ascertain whether the scores associated with Applied Social Sciences and Humanities and Social Sciences are artefacts of their small size.

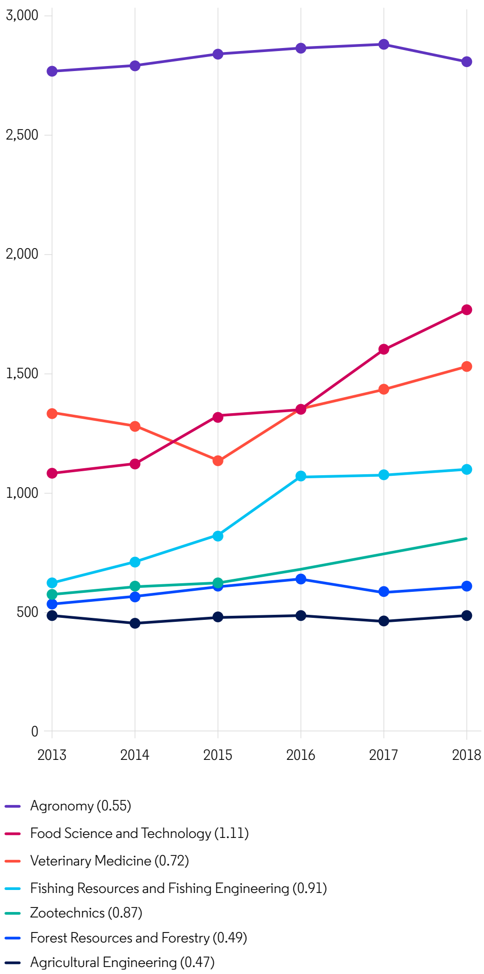
Figure 13



Strength, weakness, opportunity and threat analysis in Brazilian research aggregated by CAPES research area.

[18] https://www.lnls.cnpem.br/sirius-en/ - accessed August 19, 2019.

Figure 14



Six-year trend in the number of papers published annually in each CAPES Agriculture sub-category. Category Normalized Citation Impact (CNCI) included in parentheses.

Further analyses of CAPES sub-category data could suggest strategies for enhancing research growth, citation impact or some combination of both. For instance, the proportion of Agricultural Sciences research in Brazil is over three times the global proportion and continues to be a highly productive area. Its citation impact is persistently below world averages, locating it under "threat" in the bottom right of Figure 13. However, a detailed analysis of the sub-categories that make up the field could suggest a different story and different strategies for attaining Brazil's research objectives for this field.

The Agricultural Sciences' low CNCI can be attributed to Agronomy (CNCI of 0.46), which accounts for over a third of all papers in this research category (Figure 14). But the field is changing. There has been rapid growth in historically less productive research areas that have achieved or are approaching world average citation impact. The number of papers has grown by 60% in Food Science and Technology, which has a CNCI of 1.11. There has also been a 78% growth in Fishing Resources and Fishing Engineering papers whose citation impact has recently exceeded the world average. The level of international collaborations in Agricultural Sciences is low and below average for Brazil at 21.7%. Efforts to increase international collaborations could enhance both international visibility and potentially, the CNCI.

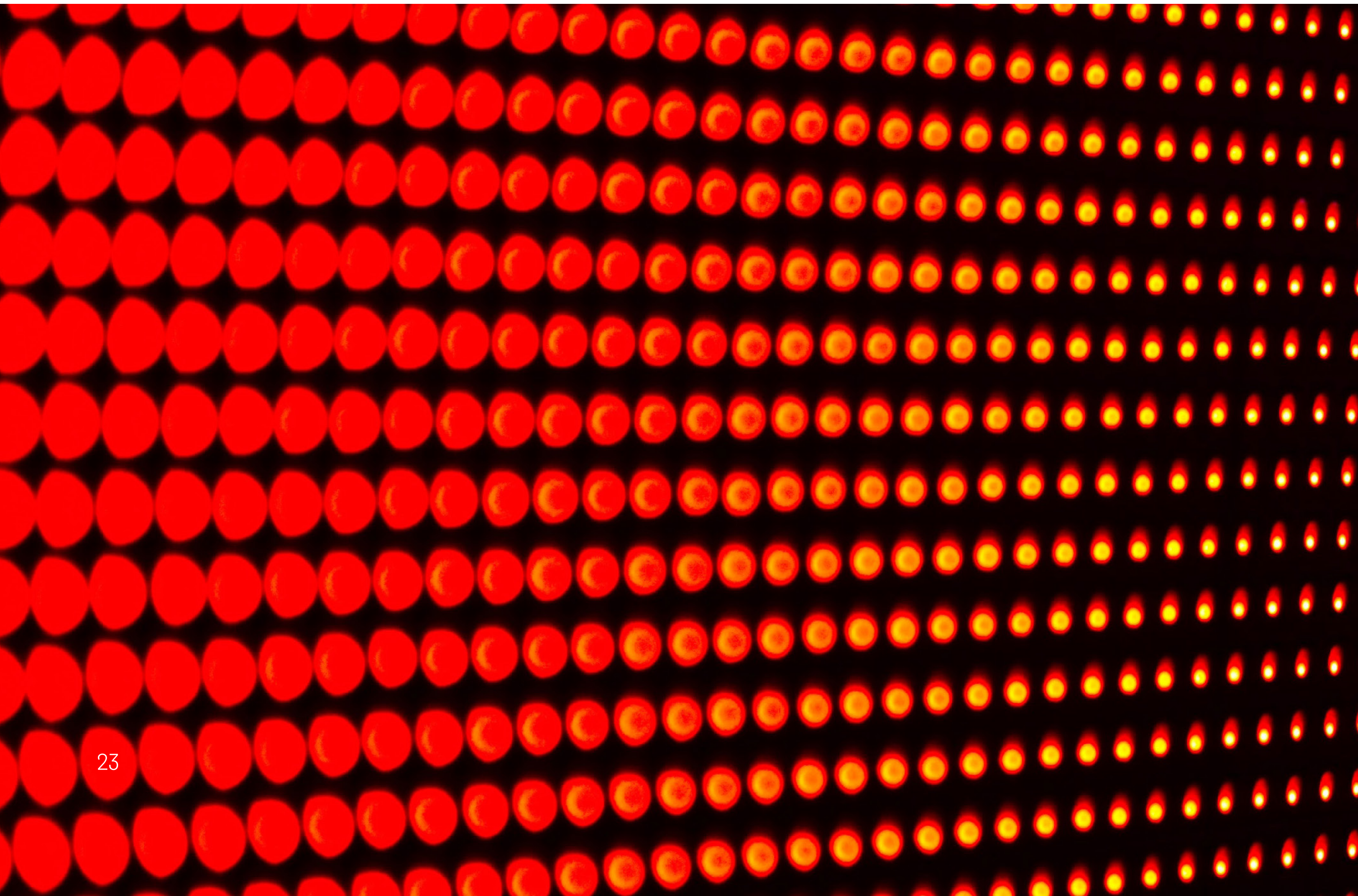
Where is this research conducted?

Which are the leading research organizations in Brazil and where do they excel?

Examining the five most productive CAPES research categories comparisons can be drawn between the institutional output and the citation impact of papers published between 2013 and 2018. The analysis is divided into two organizational categories: universities and specialized research institutes. The most productive 15 universities and 10 research institutes are examined. While the universities have areas of emphasis, most of them conduct research in each of five CAPES research categories. In contrast, the research institutions, with two notable exceptions, were established to focus on specific research areas, which are evident from the distribution of their research output across the five CAPES categories.

The relative intensity of research in different disciplines is similar across Brazil and most universities. Usually the most productive areas of research for a university are the areas with the highest output across Brazil; that is, Health Sciences, Biological Sciences, and Exact and Earth Sciences. However, there appear to be universities that do proportionally more Agriculture Sciences and Engineering research.

Figure 15 shows the numbers of papers in five CAPES categories published from these 15 Brazilian universities with the largest output in the *Web of Science* between 2013 and 2018. The shading indicates level of research output, that is, darker shades indicate institutions with higher output within a CAPES research category. These 15 universities produce more than half of all research output in Brazil.



Universidad de São Paulo dominates among the universities, producing more than twice the total number of publications compared to Universidade Estadual Paulista, which is the second most productive university in terms of research output. Only in Agricultural Science, and perhaps to a lesser extent in Engineering is the distribution of research output more evenly distributed across other universities producing papers in numbers that are comparable to the number of papers published by researchers affiliated with Universidade de São Paulo.

Figure 15

	Health Sciences	Biological Sciences	Exact and Earth Sciences	Agricultural Sciences	Engineering	All Research
Universidade de Sao Paulo	21,912	17,025	14,536	6,476	6,819	58,899
Universidade Estadual Paulista	5,283	6,948	5,336	5,908	2,914	22,868
Universidade Estadual de Campinas	5,719	4,416	6,571	1,989	3,941	19,317
Universidade Federal do Rio de Janeiro	4,672	5,351	5,503	981	3,038	17,484
Universidade Federal do Rio Grande do Sul	5,199	4,009	3,960	2,168	2,599	15,860
Universidade Federal de Minas Gerais	5,233	4,349	3,293	1,809	2,108	14,904
Universidade Federal de Sao Paulo (UNIFESP)	7,372	3,186	1,212	358	724	11,228
Universidade Federal do Parana	2,133	3,333	2,486	2,190	1,628	9,995
Universidade Federal de Santa Catarina (UFSC)	2,473	1,974	2,468	1,358	2,284	9,162
Universidade Federal de Pernambuco	1,778	2,302	2,391	662	1,082	7,098
Universidade de Brasilia	1,756	2,039	2,023	895	892	7,056
Universidade do Estado do Rio de Janeiro	2,110	1,315	3,046	281	1,030	7,039
Universidade Federal de Sao Carlos	977	1,727	2,643	670	2,072	6,980
Universidade Federal de Vicosa	602	2,726	940	3,064	441	6,893
Universidade Federal de Santa Maria (UFSM)	1,247	1,809	1,425	2,522	782	6,670

Research output of 15 top universities in five CAPES categories and All Research accross the 9 categories.

Figure 16

	Health Sciences	Biological Sciences	Exact and Earth Sciences	Agricultural Sciences	Engineering	All Research
Universidade Federal do ABC (UFABC)	1.06	0.74	1.95	-	0.95	1.68
Universidade Federal de Sao Joao del-Rei	0.65	0.61	2.53	1.29	0.95	1.54
Universidade Federal de Juiz de Fora	0.96	0.66	1.89	1.03	0.69	1.30
Universidade Federal de Sergipe	2.68	0.74	0.70	0.71	0.73	1.28
Universidade Federal de Pelotas	1.72	0.68	1.59	0.58	0.85	1.15
Universidade Federal de Santa Catarina (UFSC)	1.98	0.84	0.87	0.98	0.72	1.13
Universidade do Estado do Rio de Janeiro	0.91	0.72	1.37	0.63	1.09	1.06
Universidade Federal de Sao Paulo (UNIFESP)	1.17	0.94	0.73	0.68	0.87	1.06
Universidade Federal de Minas Gerais	1.56	0.90	0.80	0.93	0.79	1.03
Universidade Estadual de Campinas	0.86	0.94	1.23	0.87	0.74	1.03
Universidade de Sao Paulo	1.18	0.91	1.10	0.59	0.67	1.02
Universidade Federal do Rio Grande do Sul	1.50	0.78	0.97	0.88	0.82	1.02
Universidade Federal do Rio Grande do Norte	0.81	0.81	1.43	0.87	0.85	1.02
Universidade Federal do Rio de Janeiro	0.96	0.87	1.24	0.55	0.81	0.98
Universidade de Brasilia	1.50	0.86	0.64	0.59	0.87	0.90

Category Normalized Citation Impact of the 15 universities with highest impact in 5 subject categories and All Research Impact accross the 9 CAPES categories.

Figure 16 with citation impact shows that the most productive universities do not always produce the most highly cited research. Darker shades in Figure 16 indicate a field with higher CNCI and lighter shades indicate lower CNCI at that university. A university with a negligible number of papers in an area is indicated by a "–". Universidade Federal do ABC (UFABC) is the only one in this list that has too few publications in Agricultural Sciences to compute a CNCI.

Generally, universities have one or two areas of high citation impact that is above the Brazilian or world average. Most frequently these areas are Health Science and Exact and Earth Sciences which are the highest impact fields across Brazil. The exceptions are above world average impact in Agricultural Sciences at Universidade Federal de São Joao del-Rei and Engineering at Universidade do Estado do Rio de Janeiro.

Smaller to medium sized universities are doing well in terms of citation impact. Universities in São Paulo State do not dominate this list of universities ranked by overall CNCI shown on the right in Figure 16. As mentioned above, high research output is not associated with high citation impact. The top three universities in research output (Figure 15) are in the bottom half of the universities in this citation impact ranking (Figure 16).

Figure 17

	Health Sciences	Biological Sciences	Exact and Earth Sciences	Agricultural Sciences	Engineering	All Research
Empresa Brasileira de Pesquisa Agropecuaria (Embrapa)	325	3,737	1,089	5,451	504	9,598
Fundacao Oswaldo Cruz	5,264	4,983	663	560	138	9,195
Centro Brasileiro de Pesquisas Fisicas	16	26	1,969	4	233	2,097
Institute Nacional de Pesquisas da Amazonia	84	1,438	205	387	28	1,898
Instituto Nacional de Pesquisas Espaciais (INPE)	25	161	1,406	95	338	1,817
Comissao Nacional de Energia Nuclear (CNEN)	272	153	738	37	801	1,435
Hospital Israelita Albert Einstein	1,142	278	23	19	28	1,345
Instituto Butantan	450	1,046	90	99	31	1,294
Comando-Geral de Tecnologia Aeroespacial (CTA)	30	17	656	4	570	1,063
Instituto Tecnologico de Aeronautica (ITA)	25	14	648	4	555	1,039

Research output of 10 Research Institutes in five CAPES categories and All Research accross 9 categories.

Figure 17 shows the numbers of papers in five CAPES categories obtained from the *Web of Science* published by researchers affiliated with these 10 research institutes between 2013 and 2018. In spite of their specialization in one or at most two research areas, the research capacity of the research institutes is much smaller than that of the universities. The research output of the most productive research institution, Empresa Brasileira de Pesquisa Agropecuaria (Embrapa), is less than one sixth the research output of Universidade de São Paulo.

As might be expected, physics dominates the research output of Centro Brasileiro de Pesquisas Fisicas and Instituto Nacional de Pesquisas Espaciais (INPE). Physics and engineering are the focus of ComisSão Nacional de Energia Nuclear (CNEN), Comando-Geral de Tecnologia Aeroespacial (CTA), and Instituto Tecnologico de Aeronautica (ITA) (Figure 17). While Agricultural Sciences papers account for more than half the output of Empresa Brasileira de Pesquisa Agropecuaria (Embrapa), it is also a substantial producer of research in Biological Sciences and to a lesser extent in Earth and Exact Sciences. No other research institute seems to have a major focus Agricultural Sciences, except for the Instituto Nacional de Pesquisas da Amazonia, where it is the second largest research category.

Figure 18

	Health Sciences	Biological Sciences	Exact and Earth Sciences	Agricultural Sciences	Engineering	All Research
Centro Brasileiro de Pesquisas Fisicas	-	-	2.23	-	0.80	2.21
Instituto Nacional de Pesquisas Espaciais (INPE)	-	1.49	2.52	0.54	0.95	2.19
Hospital Israelita Albert Einstein	1.75	1.09	-	-	-	1.62
Fundacao Oswaldo Cruz	1.21	0.98	0.75	0.68	0.86	1.07
Instituto Nacional de Pesquisas da Amazonia	0.70	0.83	1.62	0.71	-	0.91
Empresa Brasileira de Pesquisa Agropecuaria (Embrapa)	1.17	0.75	1.06	0.62	1.08	0.72
Instituto Butantan	0.80	0.70	0.58	0.66	-	0.68
Instituto Tecnologico de Aeronautica (ITA)	-	-	0.78	-	0.60	0.65
Comando-Geral de Tecnologia Aeroespacial (CTA)	-	-	0.78	-	0.60	0.65
Comissao Nacional de Energia Nuclear (CNEN)	0.47	0.67	0.62	-	0.87	0.58

Category Normalized Citation Impact (CNCI) of 10 top research institutions in five CAPES categories

In Figure 18 the research institutions are ranked by their overall CNCI. Both of the two most productive research institutes, Centro Brasileiro de Pesquisas Fisicas and Instituto Nacional de Pesquisas Espaciais (INPE), focus on Exact and Earth Sciences with CNCI of 2.23 and 2.52 respectively. Their overall CNCI of 2.24 and 2.19[19] is significantly higher than that of any of the universities, where the highest CNCI is 1.68. However, in terms of citation impact, the universities have greater depth, where 13 of the 15 universities have overall CNCI above the world average of 1.0, while only four of the ten research institutions score above 1.0. In common with the universities, the research institutes with the highest number of publications do not have the highest citation impacts.

[19] It is likely that these high citation impacts are distorted because they include several massively multi-authored papers. For instance, there are 95 papers published with authors from INPE whose CNCI is 24.

How does Brazil support research infrastructure?

How does Brazil enhance research communication?

In addition to contributing to the expansion of the arts, humanities, engineering and scientific knowledge base and training and educating the workforce for the next generation, institution building and supporting the scientific infrastructure are also valuable contributions to the scientific enterprise. In 1998, Brazil launched the Scientific Electronic Library Online (SciELO)[20], developed by São Paulo Research Foundation (FAPESP) in partnership with BIREME, the Latin American and Caribbean Center on Health Sciences Information. It continues to flourish with additional support from the Coordination for the Improvement of Higher Education Personnel (CAPES) and the National Council for Scientific and Technological Development (CNPq).

With multilingualism[21] as a key feature, the SciELO network covers and indexes research output in Portuguese, Spanish, and English from Latin America, Portugal, Spain, and South Africa. A distinguishing feature of SciELO is its emphasis on open access and that that all journals in its collection be indexed in the Directory of Open Access Journals (DOAJ).

While there is some overlap between SciELO and the *Web of Science Core Collection,* access to SciELO can also be obtained through the *Web of Science* regional hosted collections.

[20] https://scielo.org/, http://www.scielo.br/ accessed August 21, 2019

[21] Meneghini, R, Packer, A.L. (2007). Is there science beyond English? Initiatives to increase the quality and visibility of non-English publications might help to break down language barriers in scientific communication. EMBO Reports, 8(2): 112–116. Available at http://bit.ly/2Mn43I5, accessed August 21, 2019

Findings and conclusions

During the six-year window, 2013-2018, the volume of Brazil's research output has continued to grow. Brazil has maintained its position as the 13th largest producer of research publications globally. Among the BRICS countries, this level of output is below that of China and India and above that of Russia and South Africa. Gross Domestic Expenditures on Research and Development (GERD) as a percent of Brazil's Gross Domestic Product (GDP) has ranged between 1.1 and 1.4% according to the most recent (2012-2016) data[22] available, which has not kept up with expectations of a decade ago that it might reach two percent. Except for the potentially volatile numbers for the most recent two years, the percentage of Brazilian papers among the top one percent of cited papers in the world has exceeded one percent.

The make-up of Brazilian research reveals activity and excellence concentrated in fields which have received targeted sector investment. For instance, although the Agricultural Sciences as a research field has low citation impact, disaggregating that into its constituent components shows rapid growth in sub-categories that have above world-average citation impact. Additional investments in such areas would enhance international visibility and citation impact. Similarly, investments in projects such as Sirius, the new Brazilian synchrotron light source, not only promote international collaboration but can also attract international research funding.

Policies that foster university-industry collaborations have borne fruit, resulting in exponential growth in papers co-authored with researchers from industry. A third of these collaborations are with Brazilian companies and the rest with multinationals. Petrobras is the strongest collaborator with universities and the pharmaceutical industry also has a long history of collaboration. Public universities are at the forefront of these collaborations with industry.

Public organizations are also doing well in terms of research output and its citation impact (CNCI). The 15 most productive - all of them public - organizations include 13 universities and two specialized research institutes. Of these 15, 11 have a CNCI above the Brazil average and eight have a CNCI above the world average.

In addition to promoting scholarly research and education, Brazilian public policy also supports scholarly communication, which has yielded SciELO, the electronic library that has grown beyond Brazil to 13 other countries and will soon include scholarly publications from a total of 17 countries.

[22] http://data.uis.unesco.org/index.aspx?queryid=74 accessed on August 21, 2019

Appendix 1

Bibliometrics and citation data

Bibliometrics are about publications and their citations. The academic field emerged from 'information science' and now usually refers to the methods used to study and index texts and information.

Publications cite other publications. These citation links grow into networks, and their numbers are likely to be related to the significance or impact of the publication. The meaning of the publication is determined from keywords and content. Citation analysis and content analysis have therefore become a common part of bibliometric methodology. Historically, bibliometric methods were used to trace relationships amongst academic journal citations. Now, bibliometrics are important in indexing research performance.

Bibliometric data have particular characteristics of which the user should be aware, and these are considered here.

Journal papers (publications, sources) report research work. Papers refer to or 'cite' earlier work relevant to the material being reported. New papers are cited in their turn. Papers that accumulate more citations are thought of as having greater 'impact', which is interpreted as significance or influence on their field. Citation counts are therefore recognized as a measure of impact, which can be used to index the excellence of the research from a particular group, institution or country.

The origins of citation analysis as a tool that could be applied to research performance can be traced to the mid-1950s, when Eugene Garfield proposed the concept of citation indexing and introduced the Science Citation Index, the *Social Sciences Citation Index* and the *Arts & Humanities Citation Index,* produced by the Institute of Scientific Information (now part of Clarivate Analytics).

We can count citations, but they are only 'indicators' of impact or quality – not metrics. Most impact indicators use average citation counts from groups of papers, because some individual papers may have unusual or misleading citation profiles. These outliers are diluted in larger samples.

Data source

The data used in this analysis came from the *Web of Science* database which give access not only to journals but also to conference proceedings, books, patents, websites, and chemical structures, compounds and reactions. *Web of Science* has a unified structure that integrates all data and search terms together and therefore provides a level of comparability not found in other databases. It is widely acknowledged to be the world's leading source of citation and bibliometric data. The *Web of Science* focuses on research published in journals, conferences and books in science, medicine, arts, humanities and social sciences.

The *Web of Science* was originally created as an awareness and information retrieval tool, but it has acquired an important primary use as a tool for research evaluation, using citation analysis and bibliometrics. Data coverage is both current and retrospective in the sciences, social sciences, arts and humanities, in some cases back to 1900. Within the research community this data source was previously referred to by the acronym 'ISI'.

Unlike other databases, the *Web of Science* and underlying databases are selective, that is: the journals abstracted are selected using rigorous editorial and quality criteria. The authoritative, multidisciplinary content covers over 18,000 of the highest impact journals worldwide, including Open Access journals, and over 180,000 conference proceedings. The abstracted journals encompass the majority of significant, frequently cited scientific reports and, more importantly, an even greater proportion of the scientific research output which is cited. This selective process ensures that the citation counts remain relatively stable in given research fields and do not fluctuate unduly from year to year, which increases the usability of such data for performance evaluation.

The Web of Science Group has extensive experience with databases on research inputs, activities and outputs and has developed innovative analytical approaches for benchmarking and interpreting international, national and institutional research impact.

22

There are 22 fields in *Essential Science Indicators* and 254 fields in *Web of Science*.

Database categories

The source data can be grouped in various classification systems. Most of these are based on groups of journals that have a relatively high cross-citation linkage and naturally cluster together. Custom classifications use subject maps in third-party data such as the OECD categories set out in the Frascati manual.

The Web of Science Group frequently uses the broader field categories in the *Essential Science Indicators* system and the finer journal categories in the *Web of Science*. There are 22 fields in *Essential Science Indicators* and 254 fields in *Web of Science*. In either case, our bibliometric analyses draw on the full range of data available in the underlying database, so analyses in our reports will differ slightly from anything created 'on the fly' from data in the web interface.

Most analyses start with an overall view across the data, then move to a view across broad categories and only then focus in at a finer level in the areas of greatest interest to policy, program or institutional purpose.

Assigning papers to addresses

A paper is assigned to each country and each institution whose address appears at least once for any author on that paper. One paper counts once and only once for each assignment, however many address variants may occur for the country or institution. No weighting is applied.

For example, a paper has five authors, thus:

Author	Institution	Country		
Gurney, KA	Univ Leeds	UK	Counts for Univ Leeds	Counts for UK
Adams, J	Univ Leeds	UK	No gain for Univ Leeds	No gain for UK
Kochalko, D	Univ C San Diego	USA	Counts for UCSD	Counts for USA
Munshi, S	Gujarat Univ	India	Counts for Gujarat Univ	Counts for India
Pendlebury, D	Univ Oregon	USA	Counts for Univ Oregon	No gain for USA

So this one paper with five authors would be included once in the tallies for each of four universities and once in the tallies for each of three countries.

Work carried out within the Web of Science Group, and research published elsewhere, indicates that fractional weighting based on the balance of authors by institution and country makes little difference to the conclusions of an analysis at an aggregate level. Such fractional analysis can introduce unforeseen errors in the attempt to create a detailed but uncertain assignment. Partitioning credit would make a greater difference at a detailed, group level but the analysis can then be manually validated.

Citation counts

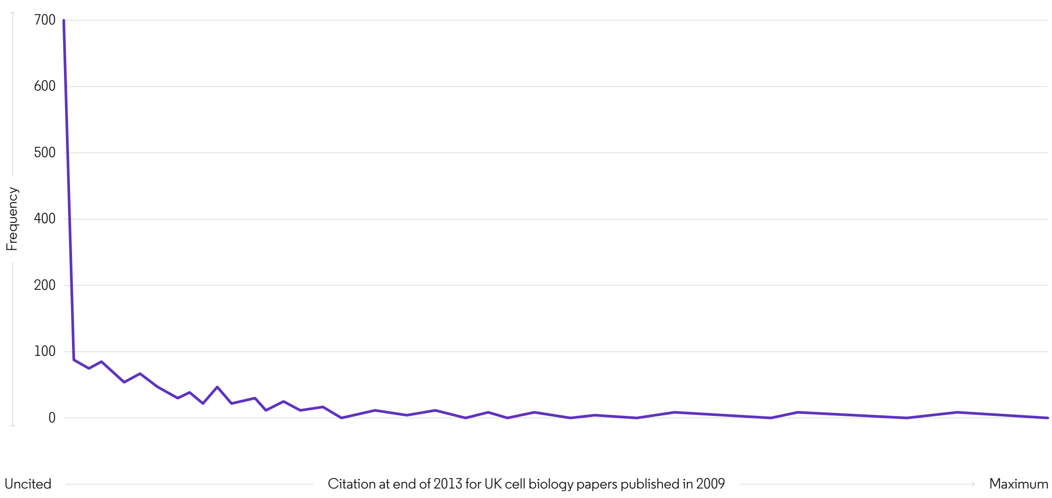
A publication accumulates citation counts when it is referred to by more recent publications. Some papers get cited frequently and many get cited rarely or never, so the distribution of citations is highly skewed.

Why are many papers never cited?

Certainly, some papers remain uncited because their content is of little or no impact, but that is not the only reason. It might be because they have been published in a journal not read by researchers to whom the paper might be interesting. It might be that they represent important but 'negative' work reporting a blind alley to be avoided by others. The publication may be a commentary in an editorial, rather than a normal journal article and thus of general rather than research interest, or it might be that the work is a 'sleeping beauty' that has yet to be recognized for its significance.

Other papers can be very highly cited: hundreds, even thousands of times. Again, there are multiple reasons for this. Most frequently cited work is being recognized for its innovative significance and impact on the research field of which it speaks. Impact here is a good reflection of quality: it is an indicator of excellence. But there are other papers which are frequently cited because their significance is slightly different: they describe key methodology; they are a thoughtful and wide-ranging review of a field; or they represent contentious views which others seek to refute.

Citation analysis cannot make value judgments about why an article is uncited nor about why it is highly cited. The analysis can only report the citation impact that the publication has achieved. We normally assume, based on many other studies linking bibliometric and peer judgments, that high citation counts correlate on average with the quality of the research.



The figure shows the skewed distribution of more or less frequently cited papers from a sample of UK authored publications in cell biology. The skew in the distribution varies from field to field. It is to compensate for such factors that actual citation counts must be normalized, or rebased, against a world baseline.

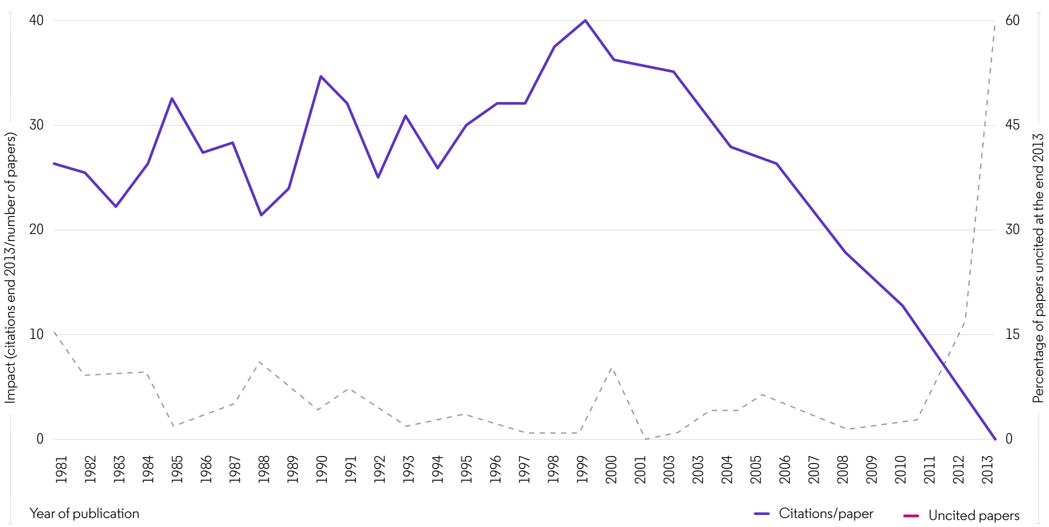
We do not seek to account separately for the effect of self-citation. If the citation count is significantly affected by self-citation then the paper is likely to have been infrequently cited. This is therefore only of consequence for low impact activity. Studies show that for large samples at national and institutional level, the effect of self-citation has little or no effect on the analytical outcomes and would not alter interpretation of the results.

Time factors

Citations accumulate over time. Older papers therefore have, on average, more citations than more recent work. The graph below shows the pattern of citation accumulation for a set of 33 journals in the journal category *Materials Science, Biomaterials*. Papers less than eight years old are, on average, still accumulating additional citations. The citation count goes on to reach a plateau for older sources.

The graph shows that the percentage of papers that have never been cited drops over about five years. Beyond five years, between 5% and 10% or more of papers remain uncited.

Account must be taken of these time factors in comparing current research with historical patterns. For these reasons, it is sometimes more appropriate to use a fixed five-year window of papers and citations to compare two periods than to look at the longer-term profile of citations and lack of citations for a recent year and an historical year.



Discipline factors

Citation rates vary between disciplines and fields. For the UK science base as a whole, ten years produces a general plateau beyond which few additional citations would be expected. On the whole, citations accumulate more rapidly and plateau at a higher level in biological sciences than physical sciences, and natural sciences generally cite at a higher rate than social sciences.

Papers are assigned to disciplines (journal categories or research fields) by the *Web of Science*, bringing cognate research areas together. Before 2007, journals were assigned to the older, well established Current Contents categories which were informed by extensive work by Thomson and with the research community since the early 1960s. This scheme has been superseded by the 254 *Web of Science* journal categories which allow for greater disaggregation for the growing volume of research which is published and abstracted.

Papers are allocated according to the journal in which the paper is published.

Some journals may be considered to be part of the publication record for more than one research field. As the example below illustrates, the journal Acta Biomaterialia is assigned to two journal categories: *Materials Science, Biomaterials and Engineering, Biomedical*.

Very few papers are not assigned to any research field and as such will not be included in specific analyses using normalized citation impact data. The journals included in the *Web of Science* databases and how they are selected are detailed here: clarivate.com/webofsciencegroup/solutions/webofscience-core-collection-editorial-selection-process/

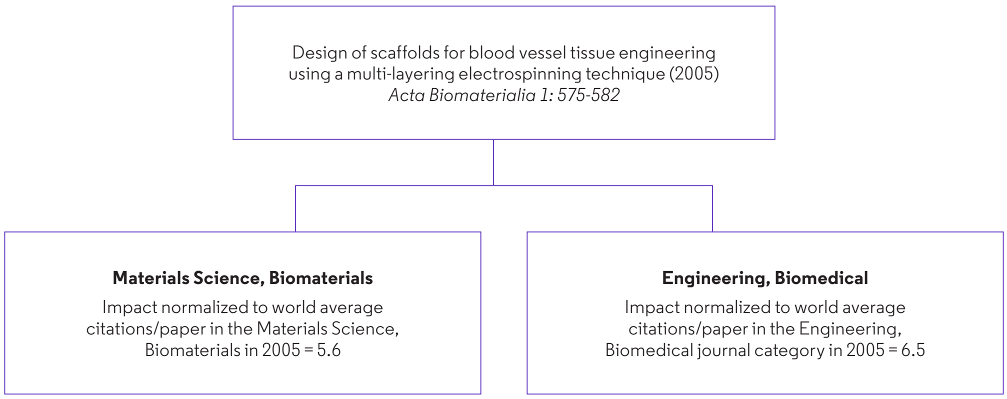
Some journals with a very diverse content, including the prestigious journals *Nature* and *Science* were classified as *Multidisciplinary* in databases created prior to 2007. The papers from these *Multidisciplinary* journals are now re-assigned to more specific research fields using an algorithm based on the research area(s) of the references cited by the article.

Normalized Citation Impact

Because citations accumulate over time at a rate that is dependent upon the field of research, all analyses must take both field and year into account. In other words, because the absolute citation count for a specific article is influenced by its field and by the year it was published, we can only make comparisons of indexed data after normalizing with reference to these two variables.

We only use citation counts for reviews and articles in calculations of impact, because document type influences the citation count. For example, a review will often be cited more frequently than an article in the same field, but editorials and meeting abstracts are rarely cited and citation rates for conference proceedings are extremely variable. The most common normalization factors are the average citations per paper for (1) the year and (2) either the field or the journal in which the paper was published. This normalization is also referred to as 'rebasing' the citation count.

Impact is therefore most commonly analysed in terms of 'normalized citation impact', or nci. The following schematic illustrates how the normalized citation impact is calculated at the paper level and journal category level.



This article in the journal Acta Biomaterialia is assigned to two journal categories: *Materials Science, Biomaterials and Engineering, Biomedical.* The world average baselines for, as an example, *Materials science, Biomaterials* are calculated by summing the citations to all the articles and reviews published worldwide in the journal Acta Biomaterialia and the other 32 journals assigned to this category for each year and dividing this by the total number of articles and reviews published in the journal category. This gives the category-specific normalized citation impact (in the above example the category-specific nciF for *Materials Science, Biomaterials* is 5.6 and the category-specific nciF for Engineering, Biomedical is higher at 6.5).

Most papers (nearly two-thirds) are assigned to a single journal category while a minority of them are assigned to more than five.

The average (normalized) citation impact can be calculated at an individual paper level where it can be associated with more than one journal category. It can also be calculated for a set of papers at any level from a single country to an individual researcher's output. In the example above, the average citation impact of the Acta Biomaterialia paper can be expressed as ((5.6 + 6.5)/2) = 6.1.

World average impact data are sourced from the *Web of Science* National Science Indicators baseline data for 2015.

Mean Normalized Citation Impact

Research performance has historically been indexed by using average citation impact, usually compared to a world average that accounts for time and discipline. As noted, however, the distribution of citations amongst papers is highly skewed because many papers are never cited while a few papers accumulate very large citation counts. That means that an average may be misleading if assumptions are made about the distribution of the underlying data.

In fact, almost all research activity metrics are skewed: for research income, PhD numbers and publications there are many low activity values and a few exceptionally high values. In reality, therefore, the skewed distribution means that average impact tends to be greater than and often significantly different from either the median or mode in the distribution. This should be borne in mind when reviewing analytical outcomes.

Impact Profiles

We have developed a bibliometric methodology that shows the proportion of papers that are uncited and the proportion that lie in each of eight categories of relative citation rates, normalized (rebased) to world average.[23] An Impact Profile enables an examination and analysis of the strengths and weaknesses of published outputs relative to world average and relative to a reference profile. This provides much more information about the basis and structure of research performance than conventionally reported averages in citation indices.

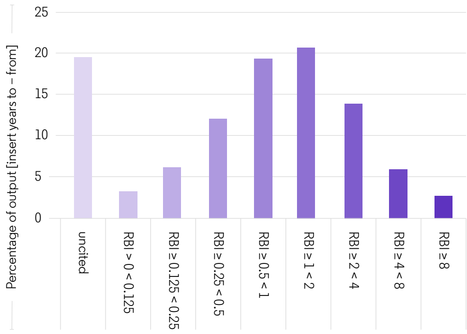
Papers which are "highly cited" are often defined in our reports as those with an average citation impact greater than or equal to 4.0, i.e. those papers which have received greater than or equal to four times the world average number of citations for papers in that subject published in that year. This differs from the database of global highly cited papers, produced by the *Web of Science* which are the top one percent most frequently cited for their field and year. The top percentile is a powerful indicator of leading performance but is too stringent a threshold for most management analyses.

The proportion of uncited papers in a dataset can be compared to the benchmark for the UK, the USA or any other country. Overall, in a typical ten-year sample, around one-quarter of papers have not been cited within the 10-year period; the majority of these are, of course, those that are most recently published.

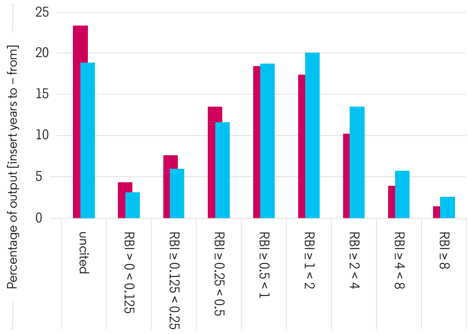
[23] Adams J, Gurney K & Marshall S (2007) Profiling citation impact: A new methodology. *Scientometrics* 72: 325-344

The Impact Profile histogram can be presented in several ways which are illustrated below.

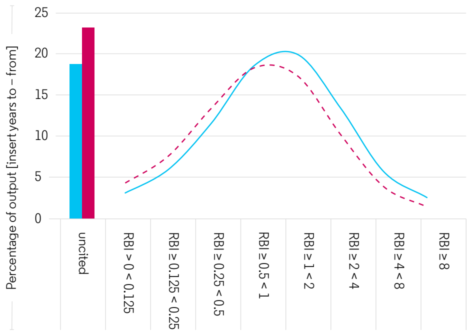
A



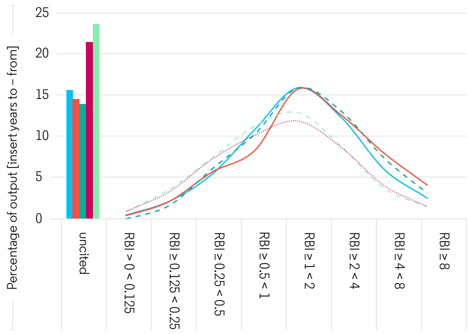
B



C



D



A: is used to represent the total output of an individual country, institution or researcher with no benchmark data. Visually it highlights the numbers of uncited papers (weaknesses) and highly cited papers (strengths).

B & C: are used to represent the total output of an individual country, institution or researcher (client) against an appropriate benchmark dataset (benchmark). The data are displayed as either histograms (B) or a combination of histogram and profile (C). Version C prevents the 'travel' which occurs in histograms where the eye is drawn to the data most offset to the right, but can be less easy to interpret as categorical data.

D: illustrates the complexity of data which can be displayed using an Impact Profile. These data show research output in defined journal categories against appropriate benchmarks:

— client, research field X;
— client, research field Y;
— client, research field Z;
— benchmark, research field X+Y;
— benchmark, research field, Z.

Impact Profiles enable an examination and analysis of the balance of published outputs relative to world average and relative to a reference profile. This provides much more information about the basis and structure of research performance than conventionally reported averages in citation indices.

An Impact Profile shows what proportion of papers are uncited and what proportion are in each of eight categories of relative citation rates, normalized to world average (which becomes 1.0 in this graph). Normalized citation rates above 1.0 indicate papers cited more often than world average for the field in which that journal is categorized and in their year of publication.

Attention should be paid to:

- The proportion of uncited papers on the left of the chart
- The proportion of cited papers either side of world average (1.0)
- The location of the most common (modal) group near the center
- The proportion of papers in the most highly cited categories to the right, (≥4 x world, ≥8 x world).

What are uncited papers?

It may be a surprise that some journal papers are never subsequently cited after publication, even by their authors. This accounts for about half the total global output for a typical, recent 10-year period.

We cannot tell why papers are not cited. It is likely that a significant proportion of papers remain uncited because they are reporting negative results which are an essential matter of record in their field but make the content less likely to be referenced in other papers. Inevitably, other papers are uncited because their content is trivial or marginal to the mainstream. However, it should not be assumed that this is the case for all such papers.

There is variation in non-citation between countries and between fields. For example, relatively more engineering papers tend to remain uncited than papers in other sciences, indicative of a disciplinary factor but not a quality factor. While there is also an obvious increase in the likelihood of citation over time, most papers that are going to be cited will be cited within a few years of publication.

What is the threshold for "highly cited"?

The Web of Science Group has traditionally used the term "Highly Cited Paper" to refer to the world's one percent of most frequently cited papers, considering year of publication and field. After reviewing the outcomes of a number of analyses, we have chosen a more relaxed definition for our descriptive and analytical work. We deem papers that are in the world's top 10% of most frequently cited papers, considering year of publication and field, to be relatively highly cited for national comparisons.

Appendix 2

University- industry collaboration methodology

The analysis presented here uses data from the *Web of Science,* obtained through searches performed at the normal WoS interface available to researchers.

While the database InCites carries data for the percentage of articles with industry co-authorship, their data is incomplete as the database is not yet able to classify correctly the nature of a large number of business organizations in Brazil (and elsewhere, for that matter). To obtain the data shown here we devised a search routine especially built to unveil the business sector in Brazil. The procedure involved obtaining the data for all scientific documents in the database with at least one author in Brazil (>300,000 records), then analyzing the organizations to which the authors were affiliated (>22,000), and then classifying among these the ones which were in the business sector. In the end we had more than four thousand organizations. At this point we ran a search looking for items in which the authors were in one of the 4,000+ business sector organizations and each and any university (obtained in a separate list).

Obtaining complete data is challenging. The *Web of Science* (and Scopus) has an incomplete classification of business sector organizations, so that their indicator on co-authorships between universities and industry undercounts the real size of the collaborations. The presently accessible database categorizes as "industry" mostly multinational companies and has weak (or no) categorization of medium and small Brazilian companies[24].

[24] This fact should not be construed as a criticism to the bibliometric bases. If anything, we might have criticism towards the way the databases are frequently used. These databases were initially developed to assist the Science Community in finding useful references to use in their work. They are heirs to the tradition initiated by the Science Citation Index, created by Eugene Garfield, as a database of information on publications, classified by sector, author, topic. Many of us scientists will remember the trips to the library to consult the Science Citation Index (and the Current Abstracts) to try and keep abreast of the new publications in our fields. Part of the challenge here is that the databases were not built to be ranking instruments. Still, they can assist ranking or science policy studies, if used with due care.

The *Web of Science* offers, however, search tools that can be used to compensate for the lack of categorization. We used the following methodology:

1. We started with a search for all items with addresses in "Brazil" (or "Brasil"), obtaining a list of approximately 700,000 items.

2. Then we used the "Analysis Tools" to obtain the list of institutions mentioned in the "Address" field of each publication. This list had 22,000 entities.

3. The list of entities was analyzed line by line to identify the ones belonging to the business sector that were then used to compose a search string applied to the field "Organization Name".

4. In addition to the list of search terms obtained in (3) we added to the search string items ending in terminations that characterize business sector entities such as LTDA, LLC, SpA, LTD, BV, INC and others, using wildcards for the initial portion of the organization name.

5. We added to the resulting list of search terms the names of organizations in Brazil known for having R&D activities (e.g. companies funded by FAPESP's Small Business Innovative Research Program, companies listed on the newspaper Valor Econômico's list of most innovative companies in Brazil). This allowed us to add more than 200 organization names to the search string.

6. Finally, we added the names of all entities considered by InCites as companies.

7. The list of entities had more than 4,000 items at this point.

8. Then it was necessary to eliminate redundancies to comply with the restriction of the search tool for the *Web of Science* that limits the number of search terms to about 6,000, apparently considering the boolian operators in the calculation. Many companies appear under similar names. In other cases, it was necessary to consolidate names: for example, works from Petrobras CENPES appear under several entity names such as 'Leopoldo Miguez', Cenpes Petrobras, 'Centro Leopoldo M', and other combinations.

9. Then the search string was built with the companies' names and the names of the universities in Brazil.

About Web of Science Group

Web of Science Group, organizes the world's research information to enable academia, corporations, publishers and governments to accelerate the pace of research. It is powered by the *Web of Science* – the world's largest publisher-neutral citation index and research intelligence platform. Its many well-known brands also include *Converis*, *EndNote*, *Kopernio*, *Publons*, *ScholarOne* and the *Institute for Scientific Information* (ISI). The 'university' of *Web of Science Group*, ISI maintains the knowledge corpus upon which the index and related information and analytical content and services are built; it disseminates that knowledge externally through events, conferences and publications and it carries out research to sustain, extend and improve the knowledge base. *Web of Science Group* is a *Clarivate Analytics* company.

Contact our experts today:

+1 215 386 0100 (U.S.)
+44 (0) 20 7433 4000 (Europe)

webofsciencegroup.com