

Scientometric Analysis of Research Output from Brazil in Response to the Zika Crisis Using e-Lattes

Ricardo Barros Sampaio^{1†}, Antônio de Abreu Batista-Jr²,
Bruno Santos Ferreira³, Mauricio L. Barreto⁴, Jesús P. Mena-Chalco⁵

¹Regional Board of Brasília Fiocruz, Distrito Federal, Brazil

²Department of Informatics Federal University of Maranhão, São Luís, MA, Brazil

³Department of Teaching and Learning Studies, Laval University, Québec, Canada

⁴Center of Integrated Data and Knowledge about Health (CIDACS), Salvador, Brazil

⁵Center for Mathematics, Computation and Cognition, Federal University of ABC Santo André, SP, Brazil

Abstract

Purpose: This paper aims to test the use of e-Lattes to map the Brazilian scientific output in a recent research health subject: Zika Virus.

Design/methodology/approach: From a set of Lattes CVs of Zika researchers registered on the Lattes Platform, we used the e-Lattes to map the Brazilian scientific response to the Zika crisis.

Findings: Brazilian science articulated quickly during the public health emergency of international concern (PHEIC) due to the creation of mechanisms to streamline funding of scientific research.

Research limitations: We did not assess any dimension of research quality, including the scientific impact and societal value.

Practical implications: e-Lattes can provide useful guidelines for different stakeholders in research groups from Lattes CVs of members.

Originality/value: The information included in Lattes CVs permits us to assess science from a broader perspective taking into account not only scientific research production but also the training of human resources and scientific collaboration.

Keywords e-Lattes; Zika virus; Lattes Platform; Brazilian research output; Brazilian science

Citation: Sampaio, R.B., Batista-Jr, A.A.B., Ferreira, B.S., Barreto, M.L., & Mena-Chalco, J.P. "Scientometric analysis of research output from Brazil in response to the Zika crisis using e-Lattes." *Journal of Data and Information Science*, vol. 5, no. 4, 2020, pp. 137–146. <https://doi.org/10.2478/jdis-2020-0038>

Received: Feb. 1, 2020

Revised: Jun. 21, 2020;

Jul. 30, 2020

Accepted: Aug. 6, 2020



[†] Corresponding author: Ricardo Barros Sampaio (E-mail: rsampaio.br@gmail.com).

1 Introduction[Ⓣ]

E-lattes (Sampaio et al., 2019) is a data-mining tool that provides visualization of knowledge and de-duplication, extraction services from academic records (Lattes CVs) registered in the Brazilian scientific curricular database called Lattes Platform. The tool enables the study of multiple research topics ranging from research groups' performance to gender bias (Huang et al., 2020), science mapping (Chen, 2017), mentorship (Li nard et al., 2018), and academic genealogy (Damaceno et al., 2019).

In this article, we use e-Lattes to map the Brazilian scientific research response to the emergency of the Zika virus (Liu, Shi, & Qin, 2019). Additionally, we want to investigate whether the tool can provide supporting information for funders on the performance of different research groups, as well as validate the past research funding policies and support the development of new strategies.

In order to analyze this bibliometric event, the rapid growth in scientific publications on a specific topic, with concentrated cases in Brazil, we used the e-Lattes software (Sampaio et al., 2019), built using R language with support for statistical analysis and graphing, to analyze the data.

In 2016, Zika caused a public health crisis in Brazil (Oliveira et al., 2020). In that same year, the World Health Organization (WHO) declared it an international public health emergency. Immediately, various research groups directed their attention to research the Zika virus infection, transmission, and control, as well as the social impact of severe forms of the disease in the most affected societies.

We propose to apply the e-Lattes to validate the streamlining policies of research funding implemented by the Brazilian government between 2015 and 2018 inclusive. These policies aimed to speed up research to develop vaccines and other technologies to control the virus.

After 2016, Brazilian scientific production on Zika increased sharply, although just a few core journals by that time had published on the topic (Singh, 2016). The world wide scientific production increase after 2015 has been attributed mainly to the association of Zika with microcephaly (Mart n-Acebes & Saiz, 2019). However, we think it also can be associated with the adoption of the fast-tracking funding policies designed to facilitate a rapid response from the scientific community (Singh, 2016).

[Ⓣ] This research is supported by the project "Plataforma de Vigil ncia de longo prazo para zika v rus e microcefalia no  mbito do SUS" (Long-Term Surveillance Platform for Zika virus and microcephaly within SUS), linked to the Center for Data Integration and Knowledge for Health (Cidacs1), with resources from the Brazilian Ministry of Health. Research is also partially supported by grant #BD-08792/17, Maranh o Research Foundation (FAPEMA). Lastly, we thank our colleagues from the Zika Research group very much for their support and the use of the software.

2 Methodology

2.1 About e-Lattes

The e-Lattes tool is an R package for data prospecting and information mining from the Lattes Platform. The tool provides a Graphical User Interface (GUI) for non-technical users². Its differential is that it deals with output and input data in JavaScript Object Notation (JSON)³ a text-based, language-independent data interchange format.

More precisely, e-Lattes collects and manages data from a set of Lattes CVs in Extensible Markup Language (XML), creating and manipulating several data structures. Internally, it visualizes them as a Venn diagram, see Figure 1. The tool presents the output in the form of many indicators and graphics.

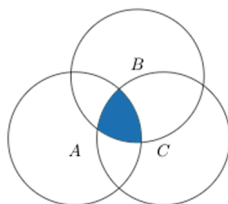


Figure 1. e-Lattes Vision, internally, after handling Lattes CVs of researchers C, A, and B. Each set includes the publications of a researcher. The shaded area indicates the publications co-authored by the three researchers.

However, before this, e-Lattes carries out data de-duplication eliminating duplicate information from the set of Lattes CVs. It eliminates duplicate copies of the same publication, from different Lattes CVs comparing the edit distance between the titles. The edit distance between two strings is the minimum number of edit operations⁴ required to transform one string into the other. The higher the edit distance, the more the titles are likely to belong to different articles.

It is also necessary to point out that e-Lattes identifies, from a Lattes CV set considered, two researchers who co-authored at least one article, creating an endogenous co-authorship network that enables the computation of diverse network metrics (Gaur et al., 2019).

2.2 Lattes Platform

The Lattes Platform is a curricular information system made available by CNPq (Brazilian National Council of Research) which includes the recording of curricular,



² <http://elattes.com.br/>

³ <https://www.rfc-editor.org/info/rfc4627>

⁴ Specifically, the e-Lattes tool uses the Levenshtein distance that allows string operations of deletion, insertion, and substitution.

academic data of researchers acting in scientific research in Brazil. Due to its abundance of information and increasing reliability, Lattes CVs have become an indispensable element for the analysis of merit of funded research with national resources.

All researchers and institutions are encouraged to keep their records up to date through Lattes, recording all their main academic activities and publications. The Lattes Curriculum allows the user to add four levels of information (Major Area, Area, Subarea, and Specialty) as defined by CNPq. Researchers might choose one or more of these subareas and Specialties to represent their scientific capabilities.

2.3 Case study

The chosen universe on the theme of Zika for this paper has to do with the fact that Brazil is the second largest producer of scientific knowledge in the area based on the number of publications. Furthermore, Brazil has been the epicentre of the Zika Virus Infection (ZIKAV), and its recent epidemic in Americas. This has had a direct effect on science production due to the fast track publication policy adopted by several scientific journals in the world.

For this reason, we believe that the analysis of researchers in the area, taking into consideration that the Lattes platform is a Brazilian database, might still have an impact on the field, even though mostly international researchers were not considered for the analysis. For future work we should be able to map areas of convergence between the Lattes Platform and other scientific publication databases.

The ZIKAV, and the neurological manifestations related to it in particular, constitutes a very complex scenario with little accumulated knowledge. Several aspects are completely new, such as the presence of the virus in the Americas and its rapid expansion, as well as the appearance of severe neurological manifestations associated with ZIKAV such as microcephaly, Guillain-Barré syndrome and possibly other neurodegenerative manifestations as yet unknown. A thorough understanding of the various factors required to cope with the epidemic requires a broad, supportive, collective and on-going scientific research effort to clarify various aspects of the virus infection, transmission, entry, dispersal, and interaction with humans, as well as the impacts generated on the health system.

On the Zika topic we test e-Lattes potential to provide support information to assist researchers/funders in their everyday activities. From a set of Lattes CVs of Zika researchers registered in the Lattes Platform, we use e-Lattes to map the group scientific research output in response to the PHEIC of Zika virus infection.

To find the Zika researchers, we perform a search from the Lattes Platform, using the search terms zika, zikv, and zkv. The query returned 9,178 CVs. Nevertheless,



we only selected 1,594 among them. The ZIKAV group, as we call it, includes researchers with a Ph.D. with at least one scientific publication related to Zika reported in their CVs.

Once the universe of our research defined, we decided to use all the data available on those Lattes CVs, including general data, research areas, publications, academic supervising, and co-authorship networks. For this specific group, we did not find any patents or other technological developments registered on their CVs.

By prospecting CVs from the Lattes Platform with e-Lattes, we assess whether the streamlining policies of research funding implemented by Brazilian research funding agencies have speeded up research output in the subject in Brazil in response to the Zika crisis.

Also, we classify Zika researchers by areas of knowledge, based on the knowledge area hierarchical classification scheme proposed by CNPq. It provides a simple mechanism to systematize and characterize information about researchers and research groups permitting four levels: (1) major area (e.g. Health Sciences), (2) area (e.g. Medicine), (3) sub-area (e.g. Surgery), and (4) specialty (e.g. Pediatric surgery). In this work, we use just levels (1) and (2) because they force the researcher to select a single option.

3 Results

Below we show a scientometrics analysis of research output from the ZIKAV group during the public health emergency. 853 articles were written, of which 661 can be found on the Web of Science.

The next analysis focused on supervision of masters dissertations, doctoral theses, and post-doctoral work, performed by the researchers. Given that the time to carry out Ph.D. research and writing up is long, the number of theses found was not very high. However, we expect an increase in the following years. Masters dissertations, on the other hand, are done in a shorter time and they show how intensely researchers adopted the subject. We also analyzed post-doc researchers supervised by a ZIKAV research group. However, for this type of academic supervision, the format of the work produced might have different outcomes, such as articles or books, and book chapters.

The number of masters dissertations was 2,668 for the period from 2015 to 2018. However, we only found 70 research works with titles including the term Zika, of which 46 were masters level, and 24 Ph.D. level. On the other hand, we did not find any post-doc work on Zika, this term was not cited in the titles of the 213 post-doc research pieces analyzed. The total number of post-doc researchers of the analyzed set from 2015 to 2018 was 344.



Research Paper

Table 1 presents the distribution of all the scientific articles that include the word Zika in their titles. The number of short and complete conference articles (131 + 18) is notable. Furthermore, Figure 2 illustrates the considerable increase in the number of researchers interested in the Zika topic between 2015 and 2018.

Table 1. Quantity of publications on Zika between 2015 and 2018 included in the sample.

Publication Type	Quantity
Conferences	149
Journal	65
Magazines	47
Books	5

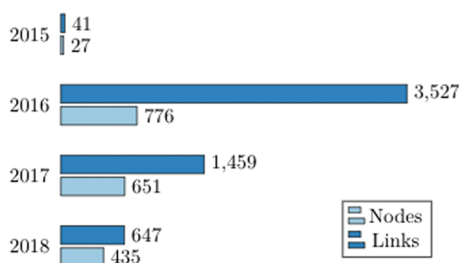


Figure 2. Researchers' interest in the topic Zika.

Additionally, Figure 3 and Table 2 show, respectively, the top 19 Brazilian researchers who work on Zika and their affiliations.

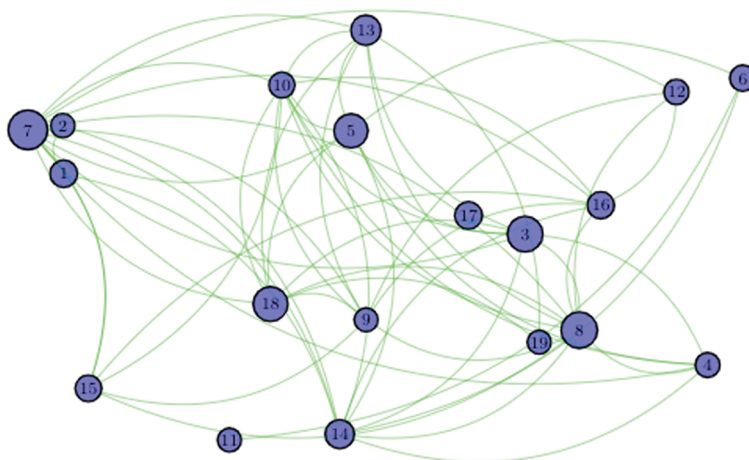


Figure 3. Scientific co-authorship between the top 19 most productive Zika researchers between 2015 and 2018. Each dot corresponds to a researcher, and each line the collaboration between them. The larger the dot size, the more the researcher published papers in the period.

Table 2. The affiliation of the top 19 most productive Zika researchers.

ID	Researcher	Affiliation
1	Marcio Roberto Teixeira Nunes	Instituto Evandro Chagas
2	Sandro Patroca da Silva	Instituto Evandro Chagas
3	Maurício Lacerda Nogueira	Escola de Medicina Sao Jose do Rio Preto
4	Thiago Moreno Lopes e Souza	Fundação Oswaldo Cruz
5	Mauro Martins Teixeira	Universidade Federal de Minas Gerais
6	Marli Tenório Cordeiro	Centro de Pesquisa Aggeu Magalhães
7	Ana Maria Bispo de Filippis	Fundação Oswaldo Cruz
8	Amilcar Tanuri	Universidade Federal do Rio de Janeiro
9	Renato Santana de Aguiar	Universidade Federal do Rio de Janeiro
10	Patricia Carvalho de Sequeira	Fundação Oswaldo Cruz
11	Edward Araujo Júnior	Universidade Federal de São Paulo
12	Melania Maria Ramos de Amorim	Inst de Medicina Integ Prof. Fernando Figueira
13	Giliane de Souza Trindade	Universidade Federal de Minas Gerais
14	Ernesto Torres de Azevedo Marques Jr	Centro de Pesquisa Aggeu Magalhães
15	Flavia Barreto dos Santos	Fundação Oswaldo Cruz
16	Rita Maria Ribeiro Nogueira	Fundação Oswaldo Cruz
17	Marco Antonio Orsini Neves	Fluminense Federal University
18	Erna Geessien Kroon	Universidade Federal de Minas Gerais
19	Jônatas Santos Abrahão	Universidade Federal de Minas Gerais

4 Discussion

To streamline the funding of research to respond to the Zika crisis, worldwide research funding agencies approved a new fast-track research funding policy for emergencies and the acceleration of research and development processes.

The bar plot in Figure 2 shows that the number of collaborations among Zika researchers during the public health emergency increased. This could demonstrate researchers' concern for the emergency as well as their response to increased research funding.

This funding policy also gave rise to diverse national scientific events producing a high rate of co-authorship between the key researchers investigating Zika, see Figure 3. Additionally we found that most articles on Zika from Brazil were published in conferences within Brazil, resulting in intense information exchange within the Brazilian community.

The funding and complexity around the Zika problem probably attracted researchers with diverse research skills represented by the CNPq knowledge area classification scheme, see Table 3. Despite this, more than 90% of the Zika researchers were from Health and Biological Sciences. They are principally from the areas of collective health and the areas of medicine.



Table 3. Number of Researchers per Area, Subarea and Specialty areas of research.

Great Area	#	Area	#	Subarea	#	Specialty	#
Health Sciences	822	Medicine	359	Molecular Biology	187	Virology	110
Biological Sciences	689	Public Health	159	Public Health	186	Infections Disease	83
Exact and Earth Sciences	108	Microbiology	126	Epidemiology	181	Medical Microbiology	33
Human Sciences	86	Biochemistry	101	Medical Clinic	175	Paediatrics	29
Applied Social Sciences	53	Immunology	87	Biology and Microorganisms	122	Neurology	24
Agricultural Sciences	40	Genetics	85	Maternal and child health	108	Proteins	21
Engineering	14	Nursing	51	Molecular Genetic	92	Molecular Biology	17
Linguistics, Literature and Arts	3	Parasitology	43	Cellular Immunology	76	Animals Infections Disease	15
		Physiotherapy	37	Applied Immunology	66	Human Protozoology	13
		Pharmacy	36	Applied Microbiology	64	Neurophysiology	13

The results suggest that the fast-track research funding model adopted by the Brazilian government did accelerate the development of the research on Zika. However, it is necessary to highlight that we did not assess any dimension of research quality (Aksnes, Langfeldt, & Wouters, 2019), including the scientific impact and societal value.

Additionally, the results also indicate that e-Lattes can provide supporting data on different types of research groups by mining the Lattes CVs and can help the funding agencies to validate their research funding policies and devise new projects.

5 Conclusions

Understanding how the Brazilian scientific community reacted to the recent public health Zika emergency is revealing. In this article, we demonstrated the use of the e-Lattes R package in a specific context. From the Lattes CVs of researchers acting in scientific research in Brazil, we validate the streamlining policies of research funding implemented by the Brazilian government to speed up the research on Zika.

Our results suggest that such policies boost research. We also conclude that from the data in the Lattes CVs, an assessment can be made from a broader perspective, one that takes into account not only scientific production but also the training of human resources through academic orientation and scientific collaboration.

Additionally, there are many challenges that still need to be overcome. The main one is combining data collected from Lattes CVs together with the publication and citation database from Microsoft Academic Graph (MAG) (Li et al., 2019), for instance.



Furthermore, we speculate that the international community could make use of e-Lattes in many ways. One possibility is to combine processed data with publication and citation data from a second database to consider the connection between researchers outside and inside the Lattes Platform.

Lane (2010) discussed how adequate incentives for researchers and academic institutions to use the Platform made Lattes a reliable and robust resource. In this sense, we believe that this model, which analyzes the profile of the researcher associated with his production, can be replicated by the international community.

In addition, the Platform's data could be extended or enriched with citation data and related metrics that, among other things, facilitate academic decision-making activities (i.e. hiring researchers, promoting jobs, financial support for research projects). In this context, e-Lattes acts as a model and infrastructure that facilitates the availability of data from the Lattes Platform and possible integration with other databases through data such as DOI and others.

Lastly, we believe that the wealth and increasing reliability of the information on the Lattes Platform justifies the attention of the international community as well as the Brazilian academic community. There is currently no other database with 6.5 million CVs such as Lattes.

Author Contribution

Proposing the research problems: Ricardo Barros Sampaio (rsampaio.br@gmail.com), Antônio de Abreu Batista-Jr (antonio.batista@ufma.br), Bruno Santos Ferreira (bruno.santos-ferreira.1@ulaval.ca), Mauricio L. Barreto (barreto.mauricio@gmail.com), Jesús P. Mena-Chalco (jesus.mena@ufabc.edu.br); Performing the research: RBS, AAB, BSF, JPM. Designing the research framework: RBS, AAB, BSF, JPM; Collecting and analyzing the data: RBS, AAB, BSF, JPM; Software development: RBS, AAB, BSF, JPM; Writing and revising the manuscript RBS, AAB, BSF, JPM.

References

- Aksnes, D.W., Langfeldt, L., & Wouters, P. (2019). Citations, citation indicators, and research quality: An overview of basic concepts and theories. *SAGE Open*, 9(1).
- Chen, C.M. (2017). Science mapping: A systematic review of the literature. *Journal of Data and Information Science*, 2(2), 1–40.
- Damaceno, R.J.P., Rossi, L., Mugnaini, R., & Mena-Chalco, J.P. (2019). The brazilian academic genealogy: Evidence of advisor-advisee relationships through quantitative analysis. *Scientometrics*, 119(1), 303–333.
- Gaur, V., Yadav, O.P., Soni, G., & S. Rathore, A.P. (2019). A review of metrics, algorithms and methodologies for network reliability. In *Proceedings of 2019 IEEE international conference on industrial engineering and engineering management (IEEM)* (pp. 1129–1133). DOI: 10.1109/IEEM44572.2019.8978688



Research Paper

- Gómez, E.J., Perez, F.A., & Ventura, D. (2018). What explains the lackluster response to zakat in Brazil? Exploring institutional, economic and health system context. *BMJ global health*, 3(5), 862. DOI: 10.1136/bmjgh-2018-000862
- Huang, J., Gates, A.J., Sinatra, R., & Barabási, A.-L. (2020). Historical comparison of gender inequality in scientific careers across countries and disciplines. *Proceedings of the National Academy of Sciences*, 117(9), 4609–4616. DOI: 10.1073/pnas.1914221117
- Lane, J. (2010). Let's make science metrics more scientific. *Nature*, 464(7288), 488–489.
- Li, J.C., Yin, Y., Fortunato, S., & Wang, D.S. (2019). A dataset of publication records for Nobel laureates. *Sci Data*, 6, 33. DOI: 10.1038/s41597-019-0033-6
- Liu, Z.Y., Shi, W.F., & Qin, C.F. (2019). The evolution of zika virus from Asia to the Americas. *Nature Reviews Microbiology*, 17(3), 131–139.
- Liénard, J.F., Achakulvisut, T., Acuna, D.E., & David, S.V. (2018). Intellectual synthesis in mentorship determines success in academic careers. *Nature Communications*, 9(4840). DOI: 10.1038/s41467-018-07034-y
- Martín-Acebes, M.A., & Saiz, J.C. (2019). The scientific response to zika virus. *Journal of clinical medicine*, 8(3). DOI: 10.3390/jcm8030369
- Oliveira, J.F., Pescarini, J.M., Rodrigues, M.S., Almeida, B.A., Henriques, C.M.P., Gouveia, F.C., ... Sampaio, R.B. (2020). The global scientific research response to the public health emergency of zika virus infection. *PLOS ONE*, 15(3), 1–16. DOI: 10.1371/journal.pone.0229790
- Sampaio, R.B., Ferreira, B.S., Batista-Jr, A.d., & Mena-Chalco, J.P. (2019). e-lattes: A new framework in R language for analysis of the lattes curriculum. In *Proceedings of the 17th international conference on scientometrics and informetrics, ISSI 2019* (pp. 2660–2661).
- Singh, N. (2016). Scientometric analysis of research on zika virus. *Virusdisease*, 27(3), 303–306.



This is an open access article licensed under the Creative Commons Attribution-NonCommercial-NoDerivs License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

