Cuba Publications in the Science Citation Index Expanded: publication characteristics, institutions and journals

Julián Monge-Nájera¹; https://orcid.org/0000-0001-7764-2966
Dianelis Olivera Batista²; https://orcid.org/0000-0002-7081-9092
Yuh-Shan Ho³*; https://orcid.org/0000-0002-2557-8736

1. Laboratorio de Ecología Urbana, Vicerrectoría de Investigación, Universidad Estatal a Distancia, 2050 San José, Costa Rica; julianmonge@gmail.com
2. Department of Information Science, Central University “Marta Abreu” of Las Villas, Santa Clara, Cuba; dolivera@uclv.cu
3. Trend Research Centre, Asia University, No. 500 Lioufeng Road, Wufeng, Taichung 41354, Taiwan; ysho@asia.edu.tw (*Correspondence)

Received 14-V-2021. Corrected 06-IX-2021. Accepted 23-IX-2021.

ABSTRACT

Introduction: In contrast with other tropical countries, Cuba has been frequently studied from the point of view of scientometrics. It has been reported that Cuban researchers often failed to cite other Cuban researchers or to collaborate with them and that 78 % of the Cuban scientific output is published in Cuban journals and mostly missed by Scopus and the Web of Science. However, there are no recent comprehensive studies of science in Cuba.

Objective: To quantify the Cuban scientific output, in all disciplines, until the beginning of 2021.

Methods: We analyzed publications from Cuba, dated 1900 to 2021, that reached the index Science Citation Index Expanded.

Results: We retrieved a total of 23,576 publications, mostly articles. In this particular database, English is the dominant language, and, over time, articles have become longer and increased the number of authors and references. Numerically, the leading institution is Universidad de La Habana. Research is strongly concentrated around medical subjects. Collaboration teams led by foreign authors have more citations recorded by the database, where the number of Cuban articles has decreased strongly since 2008.

Conclusion: Our conclusions only apply to the fraction of Cuban science covered by the Science Citation Index Expanded (under 22 %). We recommend three main improvements: to increase collaboration among Cuban scientists; to expand research areas beyond medical subjects; and to improve the quality of Cuban journals.

Key words: tropical countries; impact factor; bibliometrics; science and development; scientific production.

In contrast with other tropical countries, Cuba, a Caribbean island of 11 million inhabitants, has been frequently studied from the point of view of scientometrics. A report that covered clinical essays, from 1991 to 2001, found that a mean of 17 articles were published every year (74 % in English), with a mean of seven authors each, that in 96 journals from 17 countries (Ruiz et al., 2002). The values for dengue studies were consistent with that report, with a mean of 6 authors per article, in 45 journals and 12 countries. The second study
additionally reported a mean of six citations per article, in total; and a historical increase in the number of articles in the field of biomedicine, locally led by the Instituto de Medicina Tropical “Pedro Kourí” (Arencibia-Jorge et al., 2008). All reports, however, have limited value because American databases do not reflect the reality of Cuban publications, which appear mostly in Spanish and in journals not properly covered by those databases or other indices (Araujo Ruiz et al., 2005).

A decade ago, other studies found that education, just like other areas of research, grew in output from 2003 to 2012 (Cruz Ramírez et al., 2014) and that Scopus and the Web of Science covered poorly the growing scientific output of Cuba, making them insufficient to study Cuban science (Arencibia-Jorge & de Moya-Anegón, 2010).

More recent authors have reported that Cuban researchers often fail to cite other Cuban researchers or to collaborate with them (Chinchilla-Rodríguez et al., 2015), a result also reported by Zacca-González et al. (2015). Along the same line, a study of the state-owned pharmaceutical company found a strong specialization of research on vaccines, little collaboration and low visibility of publications (Arencibia-Jorge et al., 2016).

For the period 2003-2012, it was reported that Cuban authors were research team leaders only in small studies, published in Spanish and with low citation values in foreign databases (Chinchilla-Rodríguez et al., 2016). Over time, Cuban researchers working in Europe—mostly, students—decrease their collaboration with researchers working in the island, but their collaboration is numerically significant, with 991 institutions and 58 countries participating (Palacios-Callender & Roberts, 2018).

More recently, a comparative study in several databases estimated that 78 % of the Cuban scientific output is published in Cuban journals, and that, of those published in foreign journals, Scopus missed 33 % and the Web of Science missed 38 % (Galbán-Rodríguez et al., 2019).

The most recent study considered only evaluations of medical syllabi in Cuban journals and basically found that they have a mean of 0.6 citations per year (Fernández et al., 2020).

Considering that there is a need for a study that (1) considers all fields of Cuban science and (2) is updated, we analyzed article characteristics (subject, language, authorship), institutions and journals that appear in the Web of Science, as well as citations covered by that particular database, for all fields and updated to January 2021. We present the results as a solid basis for better decisions in the administration of science in Cuba.

MATERIALS AND METHODS

We used the Science Citation Index Expanded (“SCI-EXPANDED”), Web of Science Core Collection (updated January 21, 2021), and did an advanced search using the word “Cuba” in field country (CU) limited to the period 1900 to 2019. The journal impact factors ($IF_{2019}$) were extracted from the 2019 Journal Citation Reports ($JCR$). This study is part of a series on the scientific output of tropical countries and all methodological details have repeatedly been published and can be consulted in Calahorrano et al. (2020). We chose journal articles for further analysis because they represented the majority of document types, as well as whole research ideas and results. We also searched for relationships between article subject and number of journals in Web of Science categories. The Web of Science can classify a document in more than one type, for example, 120 proceedings papers were also classified as articles, and thus the sum of percentages can be higher than 100 %.

The second author, who is familiar with Cuban science, corrected database misspellings, errors and variability in institutional names. The “reprint author” field is the corresponding author, thus this study used “corresponding author”. If authorship is not defined as first or corresponding author, the first author was defined as both, similar to single institutional articles. The countries, institutions, and
collaboration were obtained from the author’s affiliation. “Country independent articles” and “single institute articles” were defined as “author’s affiliation is from Cuba” and “only one institute”, respectively. “Internationally collaborative articles” means that the coauthors are from different countries and “inter-institutionally collaborative articles” that the coauthors are from different institutions inside Cuba.

RESULTS

Document type, language, year of publication and citation impact: We retrieved 23,576 publications. In this database, most of the publications are articles, distantly followed by meeting abstracts and letters (Appendix 1). Articles have a mean of 21 authors and 15 citations; meeting abstracts 7 authors and 0.2 citations; and letters have on average 2 authors and 2 citations. The most cited document types are book chapters, with a mean of 48 citations; reviews with 32 citations; and articles with a mean of 15 citations (Appendix 1).

English is, by far, the dominant language. Spanish follows distantly, and the rest of the languages, led by Russian, are only marginal (Appendix 2). English articles have a mean of 23 authors, Spanish articles 4 authors and Russian articles also 4 authors. Citations do not fully correspond with the leading languages, because the highest mean citations, 25, is for publications in Chinese, followed by English with 16 citations and French with 4 citations per paper on average (Appendix 2).

The historical trend shows an increase in the number of authors, references and pages per paper over time, from 2 or 3 authors writing 7-page long papers with 6 to 30 references in the 1970s, to 20-60 authors writing 10-13 page articles with 30-50 references in the last decade (Appendix 3).

Collaboration pattern. Countries and institutions: Articles from teams led by foreign authors received twice as many citations in journals covered by this particular database (Fig. 1).

Most foreign collaboration was done with Spain, Mexico and Brazil; but the most cited papers were those published with teams from Switzerland, Sweden and the Netherlands (Appendix 4).

The leading collaborating institutions were the National Autonomous University (Mexico), the University of Sao Paulo (Brazil) and the Spanish National Research Council; but the highest citation rates were for publications with the Universities of Tokyo (Japan), Oslo (Norway) and Lund (Sweden) (Appendix 5).

The citation lifespan of Cuban articles is long, reaching four decades; but the pattern is different from those from the nearby Central American countries, because Cuban articles reached peak citation in the second year, while Central American articles were most cited in the 2-6 years following publication (Fig. 2).

Inclusion of Cuban publications in the Science Citation Index Expanded was random before 1972, with only a minimal fraction of work entering the database; it only started to have any visible trend in the 1970s, when less than 200 documents were added every year, and increased to the current volume of around 800 documents and 20 citations per document per year (Fig. 3). The same applies to international collaborative publications, which were about 50 per year in the 1990s and reached 200 per year in recent times (Appendix 6).

Leading institutions and authors: In terms of publication output that reaches this database, the leading institutions were Universidad de La Habana, Universidad Central Marta Abreu de Las Villas, and the Centro de Ingeniería Genética y Biotecnología. However, the highest citation rates correspond to Centro de Neurociencias de Cuba; Centro de Aplicaciones Tecnológicas y Desarrollo Nuclear, and Centro de Química Farmacéutica (Appendix 7).

After a clear increase in citations during the 1980s, the top productive institutes in Cuba have remained stable in recent years, with
under 60 citations per year each, but Universidad de La Habana had a different pattern: after 2000 it greatly increased to 150 to 200 citations per year (Appendix 8).

**Web of Science categories and other data:** The most productive fields were Dairy and animal science and agriculture, Biochemistry and molecular biology, and Multidisciplinary materials science, but the field appearing in more journals was the latter (Appendix 9, Appendix 10).

The journals that published most of the articles were the Cuban Journal of Agricultural Science, the Revista de Neurología and the Medice Review, but the journals with the highest citation rates were Vaccine; Physical Review B; and PLoS One (Appendix 11).

Historically, all the leading fields grew in output from 1975 to 2010, but decreased afterwards; however, Multidisciplinary materials science seems to be recovering (Fig. 4). The leading journals have evolved differently over time, with the Cuban Journal of Agriculture having a slow growth but then disappearing from the database around 2010; and other journals mostly having a smaller presence in recent years (Appendix 12).

The most frequently used words in Cuban article titles were Cuba, effect, Cuban, analysis, and effects (Appendix 13); and the most frequently used author keywords were Cuba, taxonomy, QSAR (Quantitative structure-activity relationships), systematics and vaccine (Appendix 14).

The most cited articles where those resulting from Cuban participation in international medical megaprojects about human papillomavirus; body-mass index, and proteomics (Appendix 15). Historically, all top-cited papers...
Fig. 2. Citation life span for Cuban articles, compared with Central American articles.

Fig. 3. Historical trend of exclusively Cuban articles: output and mean yearly citations.
followed a pattern of slow increase, plateau and slow decrease in citations, never reaching more than 200 citations per year, but three papers published after 2015 grew exponentially from 400 to 600 citations in recent years: they are all about medical subjects and resulted from international megaprojects in which Cuba was not the leader (Appendix 16).

**DISCUSSION**

The higher citation rates of book chapters and reviews that we found for Cuba have also been reported previously by other authors for other countries and subjects (e.g. Leydesdorff & Felt, 2012; Torres-Salinas et al., 2013) and simply reflect the fact that books and reviews summarize much work and, thus, are useful for a larger number of scientific writers, standard articles are more specialized and addressed to a smaller number of researchers (Calver & Bradley, 2010; Leydesdorff & Felt, 2012).

When comparing our results with those of older studies done by Ruiz et al. (2002) and Arencibia-Jorge et al. (2008) it is clear that Cuban articles have become longer, now have more authors, and cite more literature than those from one or two decades ago, and this fully matches a world trend for science, with research becoming more complex and requiring larger teams and longer reports that, in turn, cite more previous publications (Palacios-Callender et al., 2016; Palacios-Callender & Roberts, 2018). The largest studies are often part of international megaprojects in the field of human health, and thus, are normally written in English and highly cited. It does not mean that they are better than studies with less citations, it only reflects the larger number of research done in health-related subjects (Zacca-González et al., 2015; Calahorrano et al., 2020).

The increasing dominance of English in scientific reports, a dominance that is clear in the Cuban articles that we studied, can be misleading, because the Web of Science is known
to be heavily biased against articles published in other languages (Falagas et al., 2008; Monge-Nájera & Ho, 2017a). A very significant part of the Cuban technical and scientific literature is published in Spanish, the official language spoken on the island. All that literature, locally important and influential, is missing from the Web of Science (Arencibia-Jorge & de Moya-Anegón, 2010; Zacca-González et al., 2015). This poor coverage of Cuban science by the Web of Science and Scopus has been repeatedly reported but not solved but those databases, in detriment of Cuban science (Araujo Ruiz et al., 2005; Galbán-Rodríguez et al., 2019).

The publication, in Chinese, of articles with Cuban coauthors, is an anomaly that deserves further study, and their high citation rate could result from the large number of scientists active in China, the most populated country in the world.

The predominant collaboration of Cuban researchers with the ex-colonial power Spain, and with neighboring nations with powerful scientific institutions (Mexico and Brazil) also follows a general pattern that has been identified before in other countries throughout the tropics (Chinchilla-Rodríguez, et al., 2016; Palacios-Callender & Roberts, 2018; Sáenz et al., 2010). The higher citation rates in articles produced in collaboration with rich nations like Switzerland, Sweden and the Netherlands, can be explained by the access of researchers in those nations to larger budgets and thus to more influential journals (Palacios-Callender & Roberts, 2018).

The long citation lifespan of Cuban articles is typical of poorer countries, in which science advances slowly because of the low budgets assigned to research (Overbeck et al., 2018); a similar result has been reported for the rest of the Caribbean region, for example, in the case of Nicaragua (Monge-Nájera & Ho, 2017b). A curious result, though, is the peak citation of Cuban articles within two years of its publication: this is not normal in tropical countries (e.g. Arencibia-Jorge & de Moya-Anegón, 2010; Monge-Nájera & Ho, 2018) and may result from the Cuban concentration on health-related research, which is the most abundant and rapidly cited type of research (Arencibia-Jorge et al., 2016; Zacca-González et al., 2014). Nevertheless, the most common keywords in Cuban research reaching the Web of Science suggests a concentration on applied topics of local interest, which is normally a recipe for scientific stagnation (Toole, 2012).

The only other country in the region with a good number of scientometrics studies is El Salvador, and, like Cuba, El Salvador is a country that made an important investment in fields that are not limited to the natural sciences, particularly social sciences after the end of its civil war (Monge-Nájera & Ho, 2017c).

Overall, Cuban science has a good potential for participation in the Open Science movement that is prevalent in Latin America, but it needs to apply the following improvements: increase collaboration among Cuban scientists; expand its research areas beyond medical subjects; improve the quality of its own journals. It also needs to identify and correct the cause of its decline in the last decade. Cuba is yet to reach its full scientific potential and it should follow the example of other small nations that are more successful (Allik et al., 2020).

Ethical statement: authors declare that they all agree with this publication and made significant contributions; that there is no conflict of interest of any kind; and that we followed all pertinent ethical and legal procedures and requirements. All financial sources are fully and clearly stated in the acknowledgments section. A signed document has been filed in the journal archives.

ACKNOWLEDGMENTS

We thank Carolina Seas for her assistance with the literature and manuscript preparation.
RESUMEN

Publicaciones de Cuba en el Índice de Citación de Ciencia: características de la publicación, instituciones y revistas

Introducción: A diferencia de otros países tropicales, Cuba ha sido frecuentemente estudiada desde el punto de vista de la cienciometría. Se ha informado que los investigadores cubanos citan poco a otros investigadores cubanos, no colaboraron entre sí, y que el 78 % de la producción científica cubana se publica en revistas cubanas sin llegar a Scopus y la Web of Science. Sin embargo, no existen estudios recientes y generales de la ciencia en Cuba.

Objetivo: Cuantificar la producción científica cubana, en todas las disciplinas, hasta inicios de 2021.

Métodos: Analizamos publicaciones con dirección autoral de Cuba, fechadas de 1900 a 2021, únicamente en el Science Citation Index Expanded.

Resultados: Hallamos 23 576 publicaciones, en su mayoría artículos. En esta base de datos en particular, el inglés es el idioma dominante y, con el tiempo, los artículos se han vuelto más largos y han aumentado el número de autores y referencias. Numéricamente, la institución líder es la Universidad de La Habana. La investigación se concentra fuertemente en temas médicos. Los equipos de colaboración dirigidos por autores extranjeros tienen más citas registradas en esta base de datos estadounidense, donde el número de artículos cubanos ha disminuido fuertemente desde 2008.

Conclusión: Nuestras conclusiones solo se aplican a la fracción de la ciencia cubana cubierta por el Science Citation Index Expanded (menos del 22 %). Recomendamos tres mejoras principales: incrementar la colaboración entre los científicos cubanos; ampliar las áreas de investigación más allá de los temas médicos; y mejorar la calidad de las revistas cubanas.

Palabras clave: países tropicales; factor de impacto; bibliometría; ciencia y desarrollo; producción científica.

REFERENCES

Allik, J., Lauk, K., & Realo, A. (2020). Factors predicting the scientific wealth of nations. Cross-Cultural Research, 54(4), 364–397. https://doi.org/10.1177/1069397120910982

Araujo Ruiz, J. A., Van Hooydonk, G., Torricella Morales, R. G., & Arencibia-Jorge, R. (2005). Cuban scientific articles in ISI citation indexes and CubaCien-dencias databases (1988-2003). Scientometrics, 65(2), 161–171. https://doi.org/10.1007/s11192-005-0265-4

Arencibia-Jorge, R., Vega-Almeida, R. L., Sánchez-Taragó, N., & Araujo-Ruiz, J. A. (2008). Producción científica de Cuba sobre dengue 1981-2006: un análisis métrico en Scopus. Actamed, 18(3), 1–23.

Arenecibia-Jorge, R., & de Moya-Anegón, F. (2010). Challenges in the study of Cuban scientific output. Scientometrics, 83(3), 723–737. https://doi.org/10.1007/s11192-009-0150-7

Arenecibia-Jorge, R., Corera-Alvarez, E., Chinchilla-Rodriguez, Z., & de Moya-Anegón, F. (2016). Scientific output of the emerging Cuban biopharmaceutical industry: a scientometric approach. Scientometrics, 108(3), 1621–1636. https://doi.org/10.1007/s11192-016-2023-1

Calahorrano, L., Monge-Nájera, J., Wang, M. H., & Ho, Y. S. (2020). Ecuador publications in the Science Citation Index Expanded: institutions, subjects, citation and collaboration patterns. Revista de Biología Tropical, 68(1), 98–107. https://doi.org/10.15517/rbt.v68i1.37466

Calver, M. C., & Bradley, J. S. (2010). Patterns of citations of open access and non-open access conservation biology journal papers and book chapters. Conservation Biology, 24(3), 872–880. https://doi.org/10.1111/j.1523-1739.2010.01509.x

Chinchilla-Rodriguez, Z., Arencibia-Jorge, R., de Moya-Anegón, F., & Corera-Álvarez, E. (2015). Some patterns of Cuban scientific publication in Scopus: the current situation and challenges. Scientometrics, 103(3), 779–794. https://doi.org/10.1007/s11192-015-1568-8

Chinchilla-Rodriguez, Z., Zacca-González, G., Vargas-Quesada, B., & de Moya-Anegón, F. (2016). Benchmarking scientific performance by decomposing leadership of Cuban and Latin American institutions in Public Health. Scientometrics, 106(3), 1239–1264. https://doi.org/10.1007/s11192-015-1831-z

Cruz Ramírez, M., Escalona Reyes, M., Cabrera García, S., & Martínez Caridad, M. C. (2014). Análisis cienciometrico de las publicaciones educacionales cubanas en la WoS y Scopus (2003-2012). Revista Española de Documentación Científica, 37(3), 1–15. https://doi.org/10.3989/redc.2014.3.1119

Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, web of science, and Google scholar: strengths and weaknesses. The FASEB Journal, 22(2), 338–342. https://doi.org/10.1096/fj.07-9492LSF

Fernández, E., Quintana-Muñoz, L., Corrales-Reyes, I. E., & Chaple-Gil, A. M. (2020). Bibliometric indicators on the evaluation of medical curricula in Cuban biomedical journals. Revista Habanera de Ciencias Médicas, 19(1), 154–166.

Galbán-Rodríguez, E., Torres-Ponjuán, D., Martí-Laheira, Y., & Arencibia-Jorge, R. (2019). Measuring the Cuban scientific output in scholarly journals through a comprehensive coverage approach. Scientometrics, 121(2), 1019–1043. https://doi.org/10.1007/s11192-019-03233-6
Leydesdorff, L., & Felt, U. (2012). “Books” and “book chapters” in the book citation index (BKCI) and science citation index (SCI, SoSCI, A&HCI). Proceedings of the American Society for Information Science and Technology, 49(1), 1–7. https://doi.org/10.1002/meet.2008.1450450209

Monge-Nájera, J., & Ho, Y. S. (2017a). Honduras publications in the Science Citation Index Expanded: institutions, fields and authors. Revista de Biología Tropical, 65(2), 657–668. https://doi.org/10.15517/rbt.v65i2.25893

Monge-Nájera, J., & Ho, Y. S. (2017b). Bibliometrics of Nicaraguan publications in the Science Citation Index Expanded. Revista de Biología Tropical, 65(2), 643–655. https://doi.org/10.15517/rbt.v65i2.23985

Monge-Nájera, J., & Ho, Y. S. (2017c). El Salvador publications in the Science Citation Index Expanded: subjects, authorship, collaboration and citation patterns. Revista de Biología Tropical, 65(4), 1428–1436. https://doi.org/10.15517/rbt.v65i4.28397

Monge-Nájera, J., & Ho, Y. S. (2018). Guatemala articles in the Science Citation Index Expanded: bibliometry of subjects, collaboration, institutions and authors. Revista de Biología Tropical, 66(1), 312–320. https://doi.org/10.15517/rbt.v66i1.29875

Overbeck, G. E., Bergallo, H. G., Grellie, C. E., Akama, A., Bravo, F., Colli, G. R., Magnusson, W. E., Moraes Tomas, W., & Fernandes, G. W. (2018). Global biodiversity threatened by science budget cuts in Brazil. BioScience, 68(1), 11–12. https://doi.org/10.1093/biosci/bix130

Palacios-Callender, M., Roberts, S. A., & Roth-Berghofer, T. (2016). Evaluating patterns of national and international collaboration in Cuban science using bibliometric tools. Journal of Documentation, 72(2), 362–390. https://doi.org/10.1108/JD-11-2014-0164

Ruiz, J. A. A., Jorge, R. A., & Calzado, C. G. (2002). Ensayos clínicos cubanos publicados en revistas de impacto internacional: estudio bibliométrico del periodo 1991-2001. Revista Española de Documentación Científica, 25(3), 254–266.

Sáenz, T. W., Thorsteinsdóttir, H., & de Souza, M. C. (2010). Cuba and Brazil: an important example of South-South collaboration in health biotechnology. MEDICC Review, 12(3), 32–35. https://doi.org/10.37757/MR2010.V12.N3.8

Toole, A. A. (2012). The impact of public basic research on industrial innovation: Evidence from the pharmaceutical industry. Research Policy, 41(1), 1–12. https://doi.org/10.1016/j.respol.2011.06.004

Torres-Salinas, D., Rodríguez-Sánchez, R., Robinson-García, N., Fdez-Valdúvia, J., & García, J. A. (2013). Mapping citation patterns of book chapters in the Book Citation Index. Journal of Informetrics, 7(2), 412–424. https://doi.org/10.1016/j.joi.2013.01.004

Zacca-González, G., Chinchilla-Rodríguez, Z., Vargas-Quesada, B., & de Moya-Anegón, F. (2014). Bibliometric analysis of regional Latin America’s scientific output in public health through SCImago journal & country rank. BMC Public Health, 14(1), 1–11. https://doi.org/10.1186/1471-2458-14-632

Zacca-González, G., Chinchilla-Rodríguez, Z., Vargas-Quesada, B., & de Moya-Anegón, F. (2015). Patterns of communication and impact of the Cuban scientific output in public health. Revista Cubana de Salud Publica, 41(2), 200–216.