Authorship concentration in health sciences journals from Latin America and the Caribbean

Leonardo Ferreira Fontenelle

https://doi.org/10.1590/SciELOPreprints.3647

Authorship concentration in health sciences journals from Latin America and the Caribbean

Concentração de autoria em revistas em ciências da saúde da América Latina e do Caribe

Authors

Leonardo Ferreira Fontenelle | leonardof@leonardof.med.br | https://orcid.org/0000-0003-4064-433X

Acknowledgment

Author contributions

L. F. Fontenelle is the sole author.

Research dataset

The original data were obtained from BIREME. The derived dataset was deposited at Zenodo: https://doi.org/10.5281/zenodo.7120989.

Ethical aspects

Approval by a research ethics committee was not sought because this study relied solely on openly available data.

Competing interests

The author is an editor of a journal included in the dataset, *Revista Brasileira de Medicina de Família e Comunidade* (ISSN 2179-7994). While he is not the journal's most prolific authors in the study period, he does collaborate with the most prolific author, besides being himself one of the journal's most prolific authors, even discounting editorials.

Abstract

Objective: To describe authorship concentration indexes (proportion of articles by the most prolific author [PPMP]; Gini coefficient) among health sciences journals indexed in LILACS, and to compare them to what a previous study found in MEDLINE.

Methods: Bibliographic data were obtained from LILACS for systematically indexed journals with at least 50 signed articles (that is, with at least one individual author) from 2015 to 2019. Authors were identified by their name or, when possible, ORCID id. The PPMP was the number articles by the journal's most prolific author, divided by the number of signed articles. The Gini coefficient counted more than once articles with multiple authors author. For comparison purposes, MEDLINE data were reweighted to have the same distribution of journal size (number of signed articles) as LILACS journals.

Results: The study included 568 journals, with a median size of 166 signed articles. The median PPMP was 4.5% (95th percentile 12.9%), and the median Gini coefficient was 0.149 (95th percentile 0.310). The PPMP increased with journal size, while the Gini coefficient decreased. MEDLINE journals had lower PPMP and higher Gini coefficients, but this difference disappeared after the data were reweighted.

Conclusions: LILACS inclusion criteria are effectively countering any pro-endogeneity effect ownership by universities might have on regional journals. Journal evaluation should mind journal size when examining authorship concentration indexes. Formal derivation of their relationship with journal size would allow more precise interpretation of such indexes.

Keywords

Resumo

Objetivo: Descrever índices de concentração de autoria (proporção de artigos do autor mais prolífico [PPMP]; coeficiente de Gini) entre as revistas de ciências da saúde indexadas na LILACS, e compará-las com o que um estudo anterior encontrou na MEDLINE

Métodos: Foram obtidos dados bibliográficos da LILACS para periódicos indexados sistematicamente com pelo menos 50 artigos assinados (ou seja, com pelo menos um autor individual) de 2015 a 2019. Os autores foram identificados pelo nome ou, quando possível, pela identificação ORCID. A PPMP foi o número de artigos do autor mais prolífico da revista, dividido pelo número de artigos assinados. O coeficiente de Gini contava mais de uma vez os artigos com vários autores. Para fins de comparação, os dados MEDLINE foram reponderados para ter a mesma distribuição de tamanho de revista (número de artigos assinados) que as revistas da LILACS.

Resultados: O estudo incluiu 568 revistas, com um tamanho mediano de 166 artigos assinados. A mediana da PPMP foi de 4,5% (percentil 95 12,9%), e a mediana do coeficiente de Gini foi de 0,149 (percentil 95 0,310). A PPMP aumentou com o tamanho do periódico, enquanto o coeficiente de Gini diminuiu. As revistas da MEDLINE tinham menores PPMP e maiores coeficientes de Gini, mas esta diferença desapareceu depois que os dados foram reponderados.

Conclusões: Os critérios de inclusão da LILACS estão efetivamente neutralizando qualquer efeito pró-endogeneidade atribuível às revistas regionais serem publicadas por universidades. A avaliação das revistas deve levar em conta o tamanho do periódico ao examinar os índices de concentração de autoria. A derivação formal de sua relação com o tamanho da revista permitiria uma interpretação mais precisa de tais índices.

Palavras-chave

Publicações Periódicas como Assunto; Bibliometria; Autoria; Universidades; América Latina; Região do Caribe; Países em Desenvolvimento; LILACS.

Introduction

As dictated by the Council of Science Editors (2018), the World Association of Medical Editors (2009 Jul 25) and the International Committee of Medical Journal Editors (2021), editorial decisions should depend solely on the work's validity and interest to readers, not on any competing interests the editors might have, or the journal owner's commercial interests. Journals owned by universities and other research institutions can be particularly challenging in this regard, because editors may even have the same competing interests as the journal owners. Such journals' scope can be expected to substantially overlap the parent organization's research activities, and some of the journal editors, reviewers and authors can be expected to be affiliated with said organization. As reviewed by Barradas and Pinheiro (2016), such endogeneity may bias the peer review process to be more lenient and to incorporate less diverse points of view, undermining the journal's contribution to the scientific record; also see Sarigöl and colleagues (2017).

Recently, a high-profile scandal involving editors affiliated to the same institution as the authors motivated Locher and colleagues (2021 Mar 30) to propose an authorship concentration index (proportion of articles by the most prolific author) as an indicator of endogeneity. A second index (the Gini coefficient of inequality) was soon proposed by Scanff and colleagues (2021), who validated both indexes in a representative sample of journals indexed in MEDLINE. When either index was above its 95th percentile, the journal's most prolific author was more likely than not to participate in the journal's editorial board. Furthermore, the most prolific authors were more likely to have their articles accepted in less than three weeks, and this publication lag was even shorter in journals with a larger number of articles by their most prolific author. As discussed by Scanff and colleagues (2021), while there are legitimate reasons for editors to publish in their own journals, and for articles from prolific authors to have shorter evaluation times, both authorship concentration indexes can be used for flagging journals as potentially endogenous, "self-promotion journals" (as in Locher and colleagues (2021 Mar 30)), or "nepotistic journals" (as in Scanff and colleagues (2021)).

As noted by Scanff and colleagues (2021), about one third of the journals suspected of endogeneity were published in at least one language other than English, even if MEDLINE is skewed towards journals in English. This skewness also means the proposed 95th cutoffs may not generalize to health sciences journals in other languages, which are often not indexed in MEDLINE. Such journals are more likely to be included in the World Health Organization's Global Index Medicus, which comprises five regional bibliographic indexes targeting low- and middle-income countries. The first and foremost of such regional indexes is the Latin American and the Caribbean Literature on Health Sciences (LILACS), which is maintained by the Pan American Health Organization's Latin American and Caribbean Center on Health Sciences Information (BIREME, from its original name in Portuguese) (Clark and Castro 2002).

Most scientific journals in the region are owned by universities (Fischman et al. 2010; Beigel et al. 2021), raising the possibility of journals in LILACS being even more endogenous than those in MEDLINE. On the other hand, the region has a longstanding tradition of including endogeneity in journal evaluations (Amorim et al. 2015; Rozemblum

et al. 2015; Paz Enrique et al. 2016; Boas et al. 2021), and the lack of significant endogeneity is one of criteria for journal selection in LILACS (Suga 2019). The overall endogeneity of LILACS-indexed journals should, thus, result from the tension between these two factors. The objective of this study was, then, to describe both authorship concentration indexes among health sciences journals indexed in LILACS, and to compare them to what Scanff and colleagues (2021) found in MEDLINE.

Methods

This bibliographic study covered LILACS from years 2015 to 2019, and included journals with at least 50 articles with at least one identified individual author ("signed articles"). In January 2022, the iAH (Interface for Access on Health Information) search interface in the BIREME's Virtual Health Library was used to search for any documents indexed as journal articles in that study period. In May 2022, BIREME provided a database listing journal data such as (abbreviated) journal title, LILACS indexing, and subjects (described as Health Sciences Descriptors – DeCS). All data were downloaded in the ISO 2709 format, imported into the R statistical environment 4.2.1 (R Core Team 2022) with its package stringi 1.7.8 (Gagolewski 2022) and code written for this purpose (Fontenelle 2022), and then filtered to include only signed articles from systematically indexed journals with at least 50 of those.

The ORCID (Open Researcher and Contributor ID) identifier was used to disambiguate authors whenever possible; it was available for 7.2% of the "authorships" (authors per article times number of articles). After minimal processing of author names (such as removing extraneous elements), the ORCID id was used instead of author names when such names occurred with one and only one ORCID id. The procedure was repeated after author names had their diacritics removed (names were encoded in ASCII) and hyphens replaced with white space (because Spanish-language authors sometimes introduce them to ensure indexing by first family name). In the end, there were 148 (0.5%) unique author names associated with more than one ORCID id while also sometimes occurring with no ORCID id. Because it was not clear which ORCID id to impute, these authors (and those with no ORCID id author) were identified by the author name, as in Scanff and colleagues (2021).

Authorship concentration at the journal level, aggregating all five years, was indicated by the proportion (or rather percentage) of articles published by the most prolific author and the Gini coefficient. Each journal's most prolific author was whoever had published more articles in that journal, and the proportion of articles was simply the ratio between the number of these articles and the journal size (total amount of signed articles in that journal). The Gini coefficient (zero meaning complete equality and one meaning complete inequality) was calculated following Jasso (1979); see Davidson (2009) for a discussion of the relative merits of each way of calculating the Gini coefficient.

Both indexes were described by their median, interquartile range, total range and 95th percentile. The relationship between the indexes and journal size was described using scatter plots drawn with ggplot2, version 3.3.5 (Wickham 2016). The relationship of journal subjects and countries with these variables was described by overlaying the same scatter plots with one smoothing spline for each subject or country with 20 or more journals.

For better comparison, data from Scanff and colleagues (2021) were reweighted so that journal sizes would have the same distribution as among LILACS journals. The present study did not include a sensitivity analysis restricting data to the research articles, because the former study found the results to be largely the same, and because in the present study's data only 244 journals had at least 50 signed articles with an explicit publication type; there would be even fewer journals with at least 50 research articles.

Results and discussion

The search retrieved 148,667 articles, distributed across at least 933 journals (some rare entries didn't have an ISSN). Of these, 130,312 articles were published in 698 journals systematically indexed by LILACS in the period. Finally, 126,774 of these were signed articles published in 568 journals with at least 50 signed articles during years 2015 to 2019.

The median journal size was 166 signed articles (Table 1). Meanwhile, MEDLINE journals studied by Scanff and colleagues (2021) had a median size of 500 signed articles (IQR, 262 to 964). After reweighting, MEDLINE journals had the same size as those in LILACS (1).

Table 1: Authorship concentration indexes in LILACS journals, 2015 to 2019, compared to MEDLINE journals

Journal characteristics	LILACS (n = 591)	MEDLINE (n = 5468)
Total articles		
median, IQR	166 [102; 278]	173 [105; 291]
range	[50; 1975]	[50; 108990]
Signed articles		
median, IQR	166 [102; 274]	169 [105; 286]
range	[50; 1974]	[50; 107342]
Articles by the most prolific author		
median, IQR	8 [5; 12]	8 [5; 14]
range	[1; 96]	[1; 767]
More than one author tied as most prolific	138 (24.3%)	1520.5 (27.8%)
Percentage of articles by the most polific author		
median, IQR	4.48% [3.10%; 7.25%]	4.58% [2.94%; 7.44%]
range	[0.87%; 61.73%]	[0.13%; 39.91%]
95th percentile	12.92%	14.92%
Gini coefficient		
median, IQR	0.149 [0.099; 0.203]	0.148 [0.096–0.209]
range	[0.000; 0.484]	[0.000; 0.741]
95th percentile	0.310	0.338

IQR, interquartile range
Source: own data (LILACS) and Scanff and colleagues (2021)

The journals' most prolific author had a median of 8 articles, or 4.5% of the signed articles, with a 95th percentile of 12.9% (Table 1). One quarter of the journals had more than one

author tied as the most prolific one. The largest percentage (61.7%, more than twice the next largest percentage) was observed in a journal (*Universitas Scientiarum*, with 81 signed articles) which identifies the editor in the articles' first page; often its articles in SciELO (and records in LILACS) make it seem like the editor is one of the authors. (This journal was not included in panels A and C of Figure 1 for better visualization.) As in Scanff and colleagues (2021), the proportion of articles by the most prolific author was lower and less varying in journals with more articles (panel A in Figure 1).
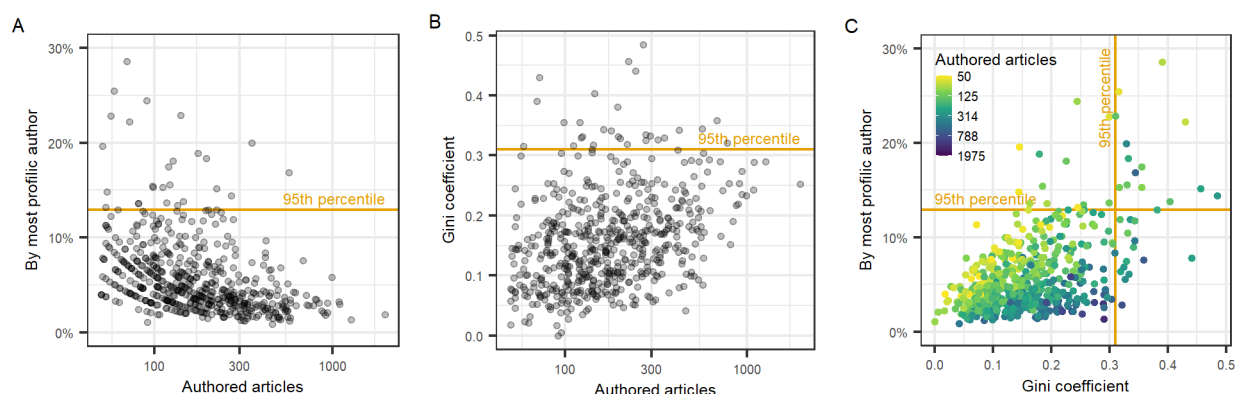
Figure 1: Pairwise scatter plot for journal size, percentage of articles by most prolific authors and Gini coefficient for LILACS journals, 2015 to 2019. One outlier journal was excluded from panels A and C for better visualization

The median Gini coefficient was 0.149, with a 95th percentile of 0.310 (1). No journal had an outstandingly high Gini coefficeint (panel B from Figure 1), but one journal (*Ágora*, with 91 signed articles) had a coefficient of zero. The journal had an average of two authors per article, and no author published more than one article during the study period. As in Scanff and colleagues (2021), the coefficient was higher on average but somewhat less varying for journals with more articles (panel B in Figure 1).

The covariance between the Gini coefficient and the proportion of articles by the most prolific author depended on the number of articles (panel C in Figure 1), as expected from the relationship between each index and the number of articles. Journals with more articles had a low proportion, independently of the Gini coefficient, while journals with fewer articles had larger proportions if they had larger Gini coefficients.

In comparison to journals indexed in LILACS, those indexed in MEDLINE had a lower proportion of articles by the most prolific author (median 2.9%, 95th percentile 10.6%) and a higher Gini coefficient (median 0.183, 95th percentile 0.355), as described by Scanff and colleagues (2021). However, after reweighting MEDLINE journals, both authorship concentration indexes in MEDLINE became virtually identical to those in LILACS (Table 1).

The importance of journal size appeared again when the LILACS journals were grouped by subject or country of publication. Figure 2 shows the average of each concentration index, by journal size, for the six countries with at least 20 journals (out of the 18 countries in the database). Likewise, Figure 3 shows the average of each concentration index, by journal size, for the six subjects with at least 20 journals (of the 183 subjects in the database). For each country and each subject, the proportion of articles by the most prolific author was lower and the Gini coefficient was higher for larger journals.
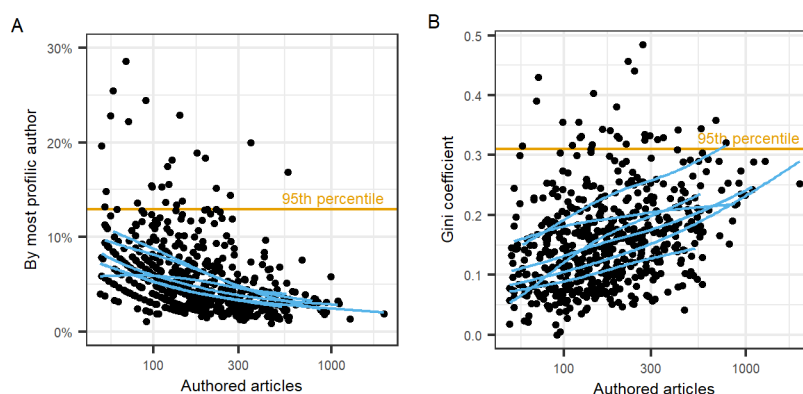
Figure 2: Scatter plot of the percentage of articles by the most prolific author and Gini coefficient, both by journal size. Each overlaid smoothing spline represents one country with 20 or more journals
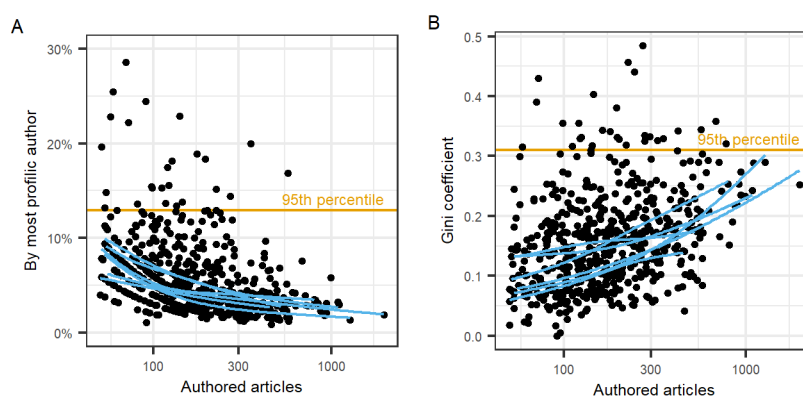

Figure 3: Scatter plot of the percentage of articles by the most prolific author and Gini coefficient, both by journal size. Each overlaid smoothing spline represents one subject with 20 or more journals

Concluding remarks

This study described the distribution, among health science journals indexed in LILACS, of two recently proposed indicators of endogeny. One indicator (the proportion of articles by the most prolific author) was higher and another (the Gini coefficient) was lower in LILACS journals than in MEDLINE ones. Both differences disappeared after adjustment for journal size.

In other words, if ownership by university does conduce to endogeneity in scientific journals, this effect is being countered by LILACS' inclusion criteria. Of note, the present study had the ORCID id for some authors, which ought to increase both concentration indexes in comparison to just using the authors' names. Furthermore, LILACS measures endogeny more directly, as the proportion of reviewers and editors from the same institution or geographic region, while the present study used more indirect measures. It is perhaps a testament to the effectiveness of the LILACS criteria that their effects are seen even when success is measured differently than LILACS itself measures it. Researchers

interested in measuring the effect on endogeneity of journal ownership by universities might have to do so in nascent journals not yet indexed in LILACS.

An implication for practice is that journal evaluations, both internal and external, should always mind the journal size when interpreting the authorship concentration indexes. While a journal publishing a thousand articles can be expected to have its *most* prolific author signing only 1% of them (ten articles), journals publishing fifty articles have a lower bound of at least 2% for even the *least* prolific authors. Likewise, larger journals can attend to larger and more heterogeneous scientific communities than smaller journals, and the Gini coefficient (or any other inequality measurement, really) for aggregate populations is larger than the Gini coefficient of their sub-populations, if there's any inequality among said sub-populations (Pyatt 1976).

This is likely to apply not only to health sciences journals from Latin America and the Caribbean, but also to those from other regions. If journals indexed in MEDLINE can have similar authorship concentration indexes as those indexed in LILACS, adjusting for journal size, then these indexes can be reasonably expected to have a similar distribution in other reputable bibliographic databases, at least in the health sciences. Perhaps even the relative abundance of non English-language journals in the outliers of Scanff and colleagues (2021) can also be explained by larger journals publishing in English, although verifying this is not in the scope of the present study. It remains an open question whether the distribution of authorship concentration indexes is the same in disciplines other than the health sciences.

Finally, research is needed to provide a more rigorous description of the relationship between journal size and authorship concentration indexes. If an ideally non-endogenous journal increased or decreased in number of signed articles, how would that affect the indexes? Such knowledge would enable more precise interpretation of the indexes, and allow one to confidently compare countries or subjects, as well as to investigate whether larger journals are generally less or more endogenous than smaller ones.

References

Amorim KM de O, Degani-Carneiro F, Ávila N da S, Marafon GJ. 2015. Evaluation Systems of Scientific Journals in Latin America. In: Alperín JP, Fischman G, editors. Made in Latin America: open access, scholarly journals, and regional innovations. Ciudad Autónoma de Buenos Aires: CLACSO. p. 61–74.

Barradas JS, Pinheiro LVR. 2016. Produtividade científica em defesa nacional. Tendências Pesqui Bras Em Ciênc Informação. 9(2). [accessed 2021 Dec 29]. https://revistas.ancib.org/index.php/tpbci/article/view/394.

Beigel F, Packer AL, Gallardo O, Salatino M. 2021. OLIVA: una mirada transversal a la producción científica indexada en América Latina. Diversidad disciplinar, colaboración institucional y multilingüismo en SciELO y Redalyc. [accessed 2021 Aug 17]. https://preprints.scielo.org/index.php/scielo/preprint/view/2653/version/2808.

Boas RFF, Campos F de F, Amaro B. 2021. Análise dos critérios formais de qualidade editorial: a política de classificação de periódicos científicos a partir do Qualis Periódicos. Informação Informação. 26(1):28–52. doi:10.5433/1981-8920.2021v26n1p28.

Clark OAC, Castro AA. 2002. Searching the Literatura Latino Americana e do Caribe em Ciências da Saúde (LILACS) database improves systematic reviews. Int J Epidemiol. 31(1):112–114. doi:10.1093/ije/31.1.112.

Council of Science Editors. 2018. CSE's White Paper on Promoting Integrity in Scientific Journal Publications. Wheat Ridge.

Davidson R. 2009. Reliable inference for the Gini index. J Econom. 150(1):30–40. doi:10.1016/j.jeconom.2008.11.004.

Fischman GE, Alperin JP, Willinsky J. 2010. Visibility and Quality in Spanish-Language Latin American Scholarly Publishing. Inf Technol Int Dev. 6(4):1–21.

Fontenelle LF. 2022. leofontenelle/authorship-concentration: Improved analytic code for "Authorship concentration in health sciences journals from Latin America and the Caribbean." doi:10.5281/zenodo.7121200. [accessed 2022 Sep 28]. https://zenodo.org/record/7121200.

Gagolewski M. 2022. **stringi** : Fast and Portable Character String Processing in *R*. J Stat Softw. 103(2). doi:10.18637/jss.v103.i02. [accessed 2022 Sep 23]. https://www.jstatsoft.org/v103/i02/.

International Committee of Medical Journal Editors. 2021. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. [accessed 2022 Feb 11]. http://www.icmje.org/recommendations/.

Jasso G. 1979. On Gini's Mean Difference and Gini's Index of Concentration. Am Sociol Rev. 44(5):867–870. doi:10.2307/2094535.

Locher C, Moher D, Cristea IA, Naudet F. 2021 Mar 30. Publication by association: how the COVID-19 pandemic has shown relationships between authors and editorial board members in the field of infectious diseases. BMJ Evid-Based Med.:bmjebm-2021-111670. doi:10.1136/bmjebm-2021-111670.

Paz Enrique LE, Peralta González MJ, Hernández Alfonso EA. 2016. Estudio bibliométrico de la Revista Centro Agrícola, Cuba. E-Cienc Inf. 6(2):1. doi:10.15517/eci.v6i2.25273.

Pyatt G. 1976. On the Interpretation and Disaggregation of Gini Coefficients. Econ J. 86(342):243–255. doi:10.2307/2230745.

R Core Team. 2022. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rozemblum C, Unzurrunzaga C, Banzato G, Pucacco C. 2015. Calidad editorial y calidad científica en los parámetros para inclusión de revistas científicas en bases de datos en Acceso Abierto y comerciales. Palabra Clave Plata. 4(2):64–80.

Sarigöl E, Garcia D, Scholtes I, Schweitzer F. 2017. Quantifying the effect of editor-author relations on manuscript handling times. Scientometrics. 113(1):609–631. doi:10.1007/s11192-017-2309-y.

Scanff A, Naudet F, Cristea IA, Moher D, Bishop DVM, Locher C. 2021. A survey of biomedical journals to detect editorial bias and nepotistic behavior. Siegerink B, editor. PLOS Biol. 19(11):e3001133. doi:10.1371/journal.pbio.3001133.

Suga SMY. 2019. Indización LILACS: acceso, visibilidad y calidad de la producción científica y técnica de América Latina y el Caribe. [accessed 2021 Apr 7]. https://pesquisa.bvsalud.org/portal/resource/es/oer-3868.

Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. https://ggplot2.tidyverse.org.

World Organization of Medical Editors. 2009 Jul 25. The Relationship Between Journal Editors-in-Chief and Owners. [accessed 2022 Feb 11]. https://wame.org/editorial-independence.

This preprint was submitted under the following conditions: