



10° EBBC

Encontro Brasileiro de
Bibliometria e Cientometria



DIFERENÇAS DE GÊNERO DE AUTORES E DOCUMENTOS DA AMÉRICA LATINA COMPARANDO AS BASES DE DADOS:

OpenAlex, Ranking Aberto de Leiden, Scopus e SciELO

Thamyres T. Choji

<https://orcid.org/0000-0002-8158-6124>
choji@uca.es
Universidad de Cádiz, España (INIBICA)

Alysson Mazoni

<https://orcid.org/0000-0001-5265-6894>
afmazoni@usp.br
Universidade de São Paulo(USP)

Jose A. Moral-Munoz

<https://orcid.org/0000-0002-6465-982X>
joseantonio.moral@uca.es
Universidad de Cádiz, España (INIBICA)

Manuel J. Cobo Martin

<https://orcid.org/0000-0001-6575-803X>
mjcobo@decsai.ugr.es
Universidad de Granada, España

Rodrigo Costas

<https://orcid.org/0000-0002-7465-6462>
rcostas@cwts.leidenuniv.nl
Leiden University, The Netherlands

RESUMO:

Este estudo compara a presença de autores latino-americanos nas bases OpenAlex, Ranking Aberto de Leiden, Scopus e SciELO, analisando diferenças de gênero e tempo de carreira. Os resultados indicam que OpenAlex apresenta perfil mais inclusivo e maior representação de autores em início de carreira. LRCCOE e Scopus mostram maior seletividade e maior presença masculina, enquanto SciELO apresenta maior proporção de mulheres e menor diferença de gênero em estágios seniores. As desigualdades variam de acordo com a base utilizada.

Palavras-chave: Diferenças de gênero, bases de dados, OpenAlex, SciELO.

EIXO TEMÁTICO:

Ciência, Tecnologia e Inovação

MODALIDADE:

Bases e Fontes de Dados

DOI: 10.22477/x.ebbc.970

COMO CITAR ESTE ARTIGO:

CHOJI, Thamyres T.; MAZONI, Alysson; MORAL-MUNOZ, Jose A.; COBO MARTIN, Manuel J.; COSTAS, Rodrigo. Diferenças de gênero de autores e documentos da América Latina comparando as bases de dados: OpenAlex, Ranking Aberto de Leiden, Scopus e SciELO. In: EBBC - Encontro Brasileiro de Bibliometria e Cientometria. 10. **Anais [...]**. Curitiba, 2026. DOI: 10.22477/x.ebbc.970. Disponível em: <https://doi.org/10.22477/x.ebbc.970>.



1 INTRODUÇÃO

Durante os últimos anos, o estudo das desigualdades de gênero na produção científica vem ganhando espaço. Esses estudos mostram que as mulheres representam a minoria dos autores e estão distribuídas de maneira desigual entre as áreas do conhecimento, linhas de pesquisa e países. Essas diferenças também são observadas a níveis de produtividade, criação de patentes e impacto acadêmico alcançado (SUGIMOTO; LARIVIÈRE, 2023). Estas características podem ser consideradas resultado de uma série de características estruturais e sistemáticas.

Na bibliometria, essas diferenças são medidas por meio de bases de dados bibliográficas. Essas fontes podem ser privadas ou de acesso aberto, como Scopus e OpenAlex, respectivamente. Elas também podem apresentar diferentes escopos, indexando documentos científicos produzidos em escala global ou em nível mais local, como é o caso de Dimensions, SciELO ou RedaLyc. Cada uma delas possui suas especificidades e pode variar em termos de cobertura de tipos de documentos, revistas científicas, idiomas, entre outros aspectos.

Desde uma perspectiva crítica, essas bases de dados não são representações neutras da produção científica. Ao contrário, constituem infraestruturas epistêmicas que moldam o que é visível e medível na ciência (KHANNA et al., 2022). A América Latina é um cenário particularmente interessante para observar essas diferenças, considerando a presença de um circuito de divulgação científica regional consolidado (BEIGEL, 2014). Um passo fundamental para enfrentar as desigualdades epistêmicas é compreender a produção de conhecimento no Sul Global e sua acessibilidade a outras regiões. Dessa forma, analisar o panorama de gênero em diferentes bases de dados na América Latina permite verificar se as diferenças previamente identificadas são consistentes ou variam de acordo com a base bi-

bliográfica utilizada.

Até o momento, não há análises sistemáticas sobre possíveis diferenças de visibilidade em nível de autores entre bases de dados regionais e internacionais. Assim, este trabalho contribui para a literatura ao comparar a presença de autores latino-americanos nas bases OpenAlex, Scopus, *Ranking* Aberto de *Leiden* (LRCCOE) e SciELO.

2 METODOLOGÍA

Com o objetivo de analisar a presença de autores latino-americanos em bases de dados internacionais e regionais, este estudo comparou quatro fontes: OpenAlex, LRCCOE, Scopus e SciELO. Cada uma foi selecionada para representar diferentes contextos e critérios de indexação.

OpenAlex é uma base de dados de acesso aberto com ampla cobertura e critérios inclusivos de indexação, além de disponibilizar identificadores únicos de autor, o que facilita a desambiguação. LRCCOE é um subconjunto aberto derivado do OpenAlex, porém com critérios mais restritivos, incluindo exigências relacionadas ao idioma e ao perfil internacional das revistas indexadas (NEES JAN VAN; VISSER; WALTMAN, 2024). Scopus é uma base de dados privada, amplamente utilizada em estudos bibliométricos e reconhecida por seu alcance internacional, embora seus critérios de indexação não sejam totalmente públicos. Por fim, SciELO é uma biblioteca digital de acesso aberto com foco regional, reunindo periódicos científicos da América Latina e promovendo a comunicação científica no contexto regional.

Neste estudo, OpenAlex foi utilizado como universo inicial dos documentos e autores latino-americanos, com base em DOI e afiliação institucional. As demais bases funcionam como filtros desse universo, a partir de critérios editoriais, lingüís-

ticos, de tipo de documento e de internacionalização, o que pode alterar o perfil dos autores observados e, consequentemente, as métricas de gênero.

Primeiramente, a partir do OpenAlex, identificaram-se todos os documentos com DOI em que pelo menos um autor possuía afiliação em um dos seguintes países: Argentina, Bolívia, Brasil, Chile, Colômbia, Costa Rica, Cuba, República Dominicana, Equador, El Salvador, Guatemala, Honduras, México, Nicarágua, Panamá, Paraguai, Peru, Porto Rico, Uruguai e Venezuela. Foram identificados 4.853.054 documentos, totalizando 5.156.185 autores. Dentre estes, 3.594.162 (69,7%) possuíam afiliação em países da América Latina.

Considerando que apenas o OpenAlex dispõe de identificadores únicos de autor, utilizou-se o DOI para verificar se os documentos identificados também estavam indexados na LRCCOE, Scopus e SciELO. A partir dessa correspondência, foi possível identificar os autores presentes em cada uma das bases. A etapa seguinte consistiu na identificação do gênero dos autores. Para isso, utilizou-se o primeiro nome e o país de afiliação, seguindo metodologia consolidada na literatura por Kozlowski et al., 2022, além da biblioteca Python GenderGuesser (PÉREZ, 2016). Foi possível inferir o gênero de 79,2% dos autores latino-americanos. Uma limitação intrínseca à identificação de gênero em larga escala é a adoção de uma classificação binária, uma vez que outras identidades de gênero requerem autoidentificação individual. Posteriormente, calculou-se a idade de carreira com base na diferença entre o primeiro e o último ano de publicação de cada autor, sendo estabelecidas quatro categorias: menos de 5 anos, entre 5 e 10 anos, entre 10 e 15 anos e mais de 15 anos.

Por fim, considerando o objetivo de analisar a presença de autores por gênero nas quatro bases de dados, foram excluídos do conjunto final os autores cujo gênero

não pôde ser identificado. O conjunto de dados final contém informações sobre autores, gênero e documentos publicados, permitindo a análise comparativa entre as diferentes bases.

3 RESULTADOS E DISCUSSÃO

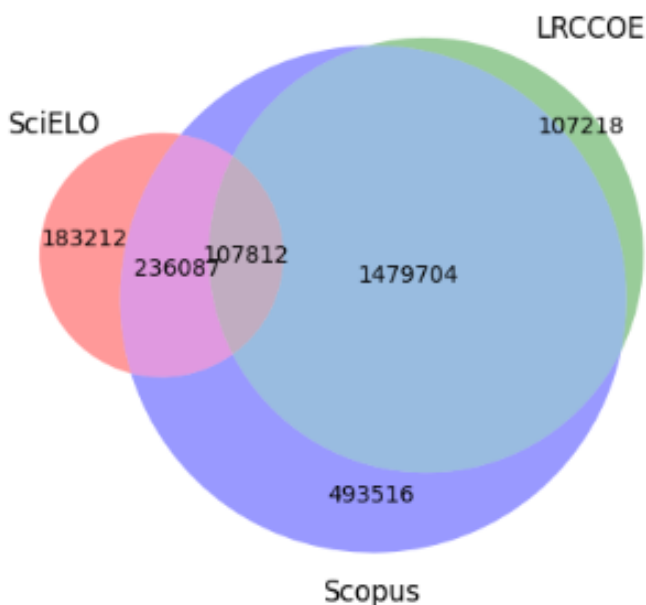
O corpus analisado incluiu 4.853.054 documentos (Tabela 1) com 2.845.611 autores com gênero identificados (Tabela 2), dos quais 50,4% são mulheres. Com relação a cobertura de documentos, Scopus foi a base de dados com a maior cobertura entre as três comparadas, correspondendo a 47,7% do universo de documentos identificados no OpenAlex. LRCCOE apresentou cobertura intermediária (35,1%), enquanto a SciELO registrou a menor proporção (11,1%). Ainda que represente apenas 11% do total, esse percentual não deve ser interpretado como residual, considerando o caráter regional e seletivo da SciELO no contexto latino-americano.

Tabela 1: Número de documentos produzidos por autores latino-americanos indexados nas diferentes bases de dados analisadas.

Base de Dados	N documentos	%referência
OpenAlex (referência)	4.853.054	
LRCCOE	1.704.423	35,1%
Scopus	2.317.119	47,7%
SciELO	536.800	11,1%



Overlap of documents across data sources

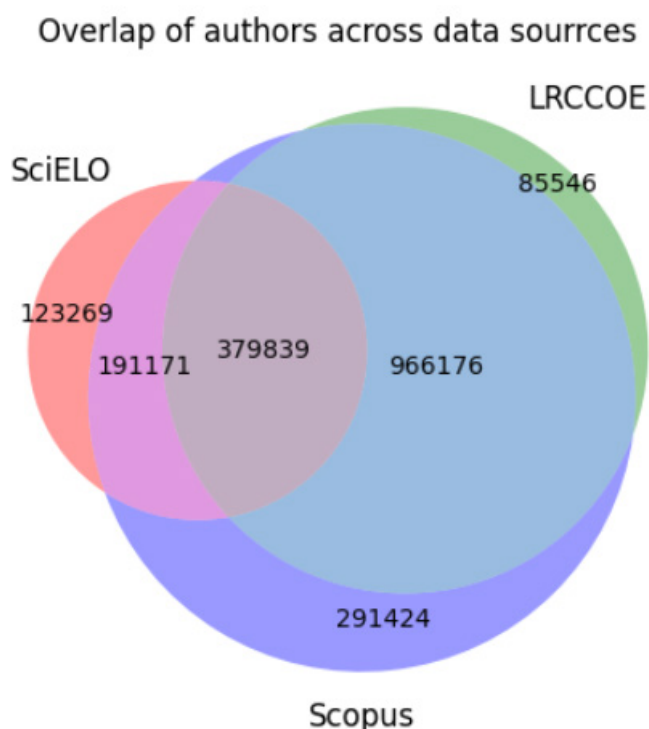


Enquanto à distribuição por gênero, OpenAlex apresentou maior proximidade à paridade (50,4%). Em contraste, LRCCOE e Scopus apresentaram maior presença masculina, em torno de 51,7%. Para SciELO o cenário foi o inverso, cerca de 52% dos autores foram mulheres (Tabela 2). Estes resultados sugerem que a diferença de gênero detectada entre os autores latino-americanos não é um valor fixo, ao invés disso, esse valor pode estar relacionado com a escolha da base de dados utilizada para medir tais diferenças.

Tabela 2: Quantidade de autores identificados por gênero em cada uma das bases de dados analisadas, produtividade e produtividade média.

	Autores identificados	%M	%H	Autoria		Média autoria	
				Mulheres	Homens	W	M
OpenAlex (referência)	2.845.611	50,4%	49,6%	2.812.951	3.573.579	2,0	2,5
LRCCOE	1.160.978	48,3%	51,7%	2.031.488	2.781.166	3,6	4,6
Scopus	1.483.578	48,4%	51,6%	2.279.098	3.031.965	3,2	4,0
SciELO	592.118	51,9%	48,1%	1.631.124	2.184.935	5,3	7,7

Figura 2: Sobreposição de autores identificados entre Scopus, SciELO e Ranking Aberto de Leiden.



Por autoria se considera a quantidade de vezes que cada um dos autores únicos é autor de um trabalho. Neste caso, (Tabela 2), as autoras apresentaram menor média de publicações em todas as bases de dados. A diferença entre a média de autorias por gênero foi menor no OpenAlex (0,5) e maior para os autores identificados na SciELO (2,4). LRCCOE e Scopus esta diferença foi de 1 e 0,8, respectivamente. Embora a menor produtividade seja conhecida pela literatura (PRADIER; CÉSPEDES; LARIVIÈRE, 2026; SÁNCHEZ-JIMÉNEZ *et al.*, 2024), os resultados sugerem que a magnitude dessa diferença também depende da base de dados utilizada. As diferenças observadas na produtividade média por gênero podem estar relacionadas a características das próprias bases de dados relacionadas à cobertura disciplinar. Determinadas áreas do conhecimento apresentam padrões distintos de coautoria, volume de publicações e participação feminina. Assim, variações na representação de áreas como Ciências da Saúde, Ciências Sociais ou Engenharias podem influenciar tanto a média de produtividade quanto a distribuição de gênero observada.

A sobreposição (Figuras 1 e 2) mostra que SciELO apresenta um subconjunto de documentos e autores distintos do circuito internacional representado por LRC-COE e Scopus, sugerindo que a escolha da base altera o que se torna visível.

Os resultados apresentados neste trabalho são preliminares. No entanto, em conjunto, eles reforçam a ideia de que bases de dados bibliográficas não são instrumentos neutros, mas infraestruturas que configuram diferentes padrões de visibilidade científica. As desigualdades de gênero observadas variam não apenas em magnitude de autores e documentos, mas também em sua produtividade. Assim, a escolha da base de dados não é apenas uma decisão técnica, mas também epistemológica.

4 CONCLUSÕES

Os resultados indicam que a escolha da base de dados influencia a medição das desigualdades de gênero na ciência latino-americana. Observou-se que bases com critérios mais inclusivos tendem a apresentar maior proximidade à paridade, enquanto bases mais seletivas e internacionalizadas mostram maior presença masculina. SciELO, por sua vez, apresentou maior proporção de mulheres. De modo geral, as variações identificadas indicam que as diferenças de gênero não se manifestam de forma homogênea, mas são fatores que dependem de critérios de cobertura de cada uma das bases de dados. Por isso, a seleção da base de dados não é apenas uma decisão metodológica, mas também epistemológica, pois pode condicionar a forma como as desigualdades são representadas e interpretadas. O presente trabalho está em curso. As seguintes etapas consistem em identificar as diferentes características dos autores identificados nas bases de dados apresentadas.

AGRADECIMENTOS

Os autores agradecem a Universidad de Cádiz por financiar a participação no EBBC, através da ajuda Plan Propio–UCA 2025-2027. Também gostaríamos de agradecer ao projeto FAPESP “Multi-observatório de dinâmica de ciência, tecnologia e inovação (MultiObs)”, 25/05419-7. Esta contribuição é parte da tese doutoral da primeira autora, Thamyres T. Choji.

REFERÊNCIAS

BEIGEL, Fernanda. Publishing from the periphery: Structural heterogeneity and segmented circuits. The evaluation of scientific publications for tenure in Argentina’s CONICET. **Current Sociology**, [S. l.], v. 62, n. 5, p. 743–765, 2014. DOI: 10.1177/0011392114533977. Disponível em: <https://doi.org/10.1177/0011392114533977>. Acesso em: 06 nov. 2025.

KHANNA, Saurabh; BALL, Jon; ALPERIN, Juan Pablo; WILLINSKY, John. Recalibrating the scope of scholarly publishing: A modest step in a vast decolonization process. **Quantitative Science Studies**, [S. l.], v. 3, n. 4, p. 912–930, 2022. DOI: 10.1162/qss_a_00228. Disponível em: https://doi.org/10.1162/qss_a_00228. Acesso em: 09 fev. 2026.

KOZLOWSKI, Diego; LARIVIÈRE, Vincent; SUGIMOTO, Cassidy R.; MONROE-WHITE, Thema. Intersectional inequalities in science. **Proceedings of the National Academy of Sciences**, [S. l.], v. 119, n. 2, p. e2113067119, 2022. DOI: 10.1073/pnas.2113067119. Disponível em: <https://pnas.org/doi/full/10.1073/pnas.2113067119>. Acesso em: 08 abr. 2022.

NEES JAN VAN, Eck; VISSER, Martijn; WALTMAN, Ludo. **Opening up the CWTS Leiden Ranking: Toward a decentralized and open model for data curation**. 2024. Disponível em: <https://www.leidenmadtrics.nl/articles/opening-up-the-cwts-leiden-ranking-toward-a-decentralized-and-open-model-for-data-curation>. Acesso em: 9 jan. 2026.

PÉREZ, Israel Saeta. **gender-guesser: Get the gender from first name**. , 2016. Disponível em: <https://github.com/lead-ratings/gender-guesser>. Acesso em: 29 maio. 2022.

PRADIER, Carolina; CÉSPEDES, Lucía; LARIVIÈRE, Vincent. How multilingual is scholarly communication? Mapping the global distribution of languages in publications and citations. **Journal of the Association for Information Science and**



Technology, [S. l.], v. n/a, n. n/a, p. 1–15, 2026. DOI: 10.1002/asi.70055. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.70055>. Acesso em: 06 fev. 2026.

SÁNCHEZ-JIMÉNEZ, Rodrigo; GUERRERO-CASTILLO, Pablo; GUERRERO-BOTE, Vicente P.; HALEVI, Gali; DE-MOYA-ANEGÓN, Félix. Analysis of the distribution of authorship by gender in scientific output: A global perspective. **Journal of Informetrics**, [S. l.], v. 18, n. 3, p. 101556, 2024. DOI: 10.1016/j.joi.2024.101556. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1751157724000695>. Acesso em: 12 jul. 2024.

SUGIMOTO, Cassidy R.; LARIVIÈRE, Vincent. **Equity for women in science: dismantling systemic barriers to advancement**. Cambridge, Massachusetts: Harvard university press, 2023. DOI: 10.2307/j.ctv37qqwrz. Disponível em: <https://www.jstor.org/stable/j.ctv37qqwrz>. Acesso em: 22 mar. 2024.