



ESTUDO SOBRE A DILUIÇÃO DE CITAÇÕES NOS ARTIGOS DERIVADOS DE *PREPRINTS* PUBLICADOS EM PERIÓDICOS DAS CIÊNCIAS DA SAÚDE

Pedro Henrique da Silva Rodrigues

 <https://orcid.org/0000-0001-5909-2067>.

 pedrohenrique@ufrgs.br.

 Universidade Federal do Rio Grande do Sul (UFRGS) |

 <https://ror.org/041yk2d64> | Porto Alegre, Brasil.

Sonia Elisa Caregnato

 <https://orcid.org/0000-0002-5676-2763>.

 sonia.caregnato@ufrgs.br.

 Universidade Federal do Rio Grande do Sul (UFRGS) |

 <https://ror.org/041yk2d64> | Porto Alegre, Brasil.

Eixo temático: Métodos Bibliométricos e de Citação

Modalidade: Resumo expandido

DOI: 10.22477/ix.ebbc.266

Resumo: Este estudo preliminar investigou a diluição de citações em periódicos que publicaram artigos derivados de *preprints* nas Ciências da Saúde. Trata-se de uma pesquisa quantitativa na qual, primeiramente, foram coletados os dez *preprints* mais citados, de acordo com o *Preprints Citation Index*; em seguida, foram obtidas as citações de cada *preprint*, diferenciando aquelas feitas antes e depois da publicação de cada artigo derivado dos *preprints*. Os resultados revelaram 607 citações diluídas entre 2022 e 2023. A maioria das citações diluídas foi observada no primeiro ano após a publicação do artigo, com uma notável redução a partir do segundo ano. O intervalo médio entre a publicação do *preprint* e do artigo foi inferior a um ano. Nas considerações finais, são oferecidas sugestões que podem ser seguidas pelos editores de periódicos, autores e repositórios de *preprints* para mitigar a diluição de citações.

Palavras-Chave: *Preprints*. Diluição de citações. Periódicos. Ciências da Saúde. *Preprints Citation Index*.

1 INTRODUÇÃO

Committee on Publication Ethics - Cope (2021), Higgins e Steiner (2021), Pagliaro (2021) e Rodríguez (2019) descrevem *preprints* como a versão inicial de um manuscrito acadêmico, disponibilizada em plataformas de acesso aberto, conhecidas como repositórios ou servidores de *preprints*, antes ou durante o processo de revisão por pares.

Entre as vantagens dos *preprints*, Aquino-Jarquin *et al.* (2018) e Ettinger *et al.* (2022) apontam a capacidade de garantir o reconhecimento adequado da autoria de descobertas científicas. Já Ciriminna *et al.* (2021) e Smart (2022) destacam a difusão mais rápida de novos conhecimentos como uma das vantagens dos *preprints*. Além disso, Sever, Eisen e Inglis (2019) enfatizam o baixo custo de operação de um repositório de *preprints*, possibilitando o compartilhamento dos documentos em acesso aberto, ou seja, sem custo para autores e leitores.

Por outro lado, Smart (2022) destaca desvantagens associadas aos *preprints*, tais como a falta de garantia de qualidade e confiabilidade, a possibilidade de disseminação de trabalhos fraudulentos ou enviesados, e o desafio de lidar com o volume excessivo de informações, o que dificultaria a identificação de trabalhos de qualidade. Além disso, Avissar-Whiting (2022) alerta para a dificuldade no controle de versões, mencionando que a vinculação entre o *preprint* e seu artigo correspondente nem sempre existe ou é feita de maneira eficiente.

Gehanno *et al.* (2022) conduziram um estudo para avaliar a confiabilidade dos *preprints*. Eles compararam 124 *preprints* sobre COVID-19, publicados no medRxiv, com suas respectivas publicações em periódicos. A pesquisa revelou diferenças significativas nos títulos, dados e conclusões em quase metade dos documentos analisados. De forma semelhante, Bero *et al.* (2021) compararam 67 *preprints* sobre COVID-19 com suas publicações subsequentes em periódicos. Os autores concluíram que 34% dos casos apresentavam alguma divergência nos resultados. Nesse sentido, o medRxiv recomenda que os *preprints* não sejam usados como guia para a prática clínica ou divulgados na mídia como informações consolidadas (About [...], 2024).

Além disso, Davis (2018) e Smart (2022) alertam para o risco de diluição de citações associado aos *preprints*. Isso ocorre quando os autores citam o *preprint* em vez do artigo correspondente publicado em uma revista. Avissar-Whiting (2022) esclarece que a diluição de citações é o fenômeno em que as citações se dispersam entre diferentes estágios e versões de um artigo, ou seja, quando existem múltiplas versões de um mesmo trabalho e as citações associadas a ele ficam dispersas entre essas versões. Para Davis (2018), citar *preprints* que se tornaram artigos pode prejudicar não somente os periódicos, que perdem reconhecimento público, mas também os autores, que ficam impossibilitados de agregar citações.

Importante observar que esse problema é particularmente relevante na atualidade, quando se observa, por exemplo, que 70,0% dos *preprints* na arXiv (2010-2015) foram publicados posteriormente em periódicos científicos (Alvarez; Caregnato, 2017). Nesse contexto, este estudo tem como objetivo identificar e quantificar a diluição de citações em revistas da área de ciências da saúde que publicaram artigos derivados de *preprints*. Coletamos dez *preprints*, mas um deles foi descartado. Portanto, apresentamos resultados preliminares baseados na análise de nove *preprints*.



2 PROCEDIMENTOS METODOLÓGICOS

Este estudo quantitativo foi composto pelas seguintes etapas: (I) coleta de *preprints* usando a base *Preprints Citation Index* da *Web of Science* (WoS); (II) associação dos *preprints* aos artigos derivados; (III) coleta de citações aos *preprints* com o *CitationChaser*; (IV) organização das citações distinguindo aquelas antes e após a publicação do artigo; (V) quantificação das citações diluídas ano a ano.

Haddaway, Grainger e Gray (2022) explicam que o *CitationChaser* é uma ferramenta de código aberto e gratuita para realizar uma busca rápida de citações. Essa ferramenta gera arquivos padronizados no formato RIS (Research Information Systems), facilitando a exportação e o manuseio dos dados em outros programas.

A escolha pelo *CitationChaser* foi motivada por fornecer metadados das citações como título, autoria, DOI (*Digital Object Identifier*) e data de publicação do trabalho citante, ou seja, a data da própria citação. Isso contribui significativamente para a análise temporal dos dados. Em relação ao desempenho, Bracchiglione *et al.* (2023) conduziram um estudo comparativo da sensibilidade e precisão na busca de citações usando cinco ferramentas: *WoS*, *Scopus*, *Google Scholar*, *Citationchaser* e *Paperfetcher*, e não encontraram diferenças significativas de desempenho entre elas.

Escolhemos restringir os resultados da busca no *Preprints Citation Index* aos documentos publicados no *medRxiv* devido ao fato de que nesse servidor os manuscritos são facilmente detectáveis, citáveis e indexados por vários mecanismos de busca e serviços de terceiros (About [...], 2024). Além disso, o *medRxiv* não aceita *preprint* que já tenha sido aceito para publicação, publicado ou postado em outro lugar. Isso é importante porque permite que a pesquisa identifique as citações feitas antes e depois da publicação do artigo.

A busca no *Preprint Citation Index* (Índice de Citação de *Preprints*), da *Web of Science*, identificou, sem a aplicação de filtros relacionados ao período de tempo, aos periódicos, à região ou a outros critérios, um total de 30.481 *preprints* publicados no *medRxiv*, os quais foram classificados com base no número de citações recebidas. Dentre esses, os 50 *preprints* mais citados foram selecionados. Desses, os 10 primeiros foram escolhidos para o estudo piloto, apresentado aqui para debate com a comunidade, visto que este se trata de um trabalho em andamento.

Neste trabalho, diferenciamos os seguintes tipos de citações:

- a. citações totais ao *preprint*: a quantidade total de citações atribuídas ao *preprint*, inclui tanto as feitas antes quanto após a publicação do artigo correspondente;
- b. citação ao *preprint*: citação atribuída corretamente ao *preprint*, feita até a publicação do artigo correspondente, se houver;
- c. citação diluída: citação atribuída ao *preprint* após a publicação do artigo correspondente;
- d. citação ao artigo: citação atribuída corretamente ao artigo.

A escolha pelas ciências da saúde se deu pelo fato dela ser uma área sensível quando se trata de



preprints, especialmente pelo seu potencial impacto na saúde pública e nas decisões clínicas. Nesse sentido, Flanagan, Fontanarosa e Bauchner (2020) afirmam que os servidores de *preprints* têm sido fonte de informação utilizada por políticos e médicos para apoiar tratamentos específicos, alguns sem comprovação científica, como o uso de hidroxicloroquina e plasma convalescente para o tratamento da COVID-19. Uma limitação deste estudo é a decisão de concentrar esforços em uma área específica do conhecimento, o que restringe a generalização dos resultados. Isso acontece devido às diferentes características e práticas de comunicação científica associadas às distintas comunidades científicas. Ou seja, as práticas de citação variam dependendo do campo de estudo.

3 RESULTADOS

Na Tabela 1, são exibidos os *preprints* juntamente com o ano de postagem do *preprint* e do artigo, o número de citações totais aos *preprints* até dezembro de 2023, além da quantidade de citações anuais atribuídas aos *preprints* após a publicação da versão revisada em periódico, ou seja, o número de citações diluídas.

O maior intervalo de tempo entre a publicação do *preprint* e a publicação do artigo foi de dois anos (IDP4). A maioria dos *preprints* teve sua versão em artigo publicada em menos de um ano, possivelmente explicado pelo contexto da COVID-19.

Tabela 1 - Resumo dos *preprints*: ano de postagem do *preprint* e do artigo, citações totais aos *preprints* até dezembro de 2023 e citações diluídas.

Preprints							
ID	Título do <i>preprint</i>	Preprint	Artigo	Citações totais ao <i>preprint</i>	Citações diluídas		
					2022	2023	Total
IDP1	SARS-CoV-2 Omicron has extensive but incomplete escape of Pfizer BNT162b2 elicited neutralization and requires ACE2 for infection	2021	2021	248	168	21	183
IDP2	Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data	2021	2021	513	60	8	68
IDP3	Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England	2021	2021	326	14	8	22
IDP4	Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia	2020	2022	230	-	17	17
IDP5	Reduced Neutralization of SARS-CoV-2 Omicron Variant by Vaccine Sera and Monoclonal Antibodies	2021	2022	242	-	11	11

Preprints							
ID	Título do <i>preprint</i>	<i>Preprint</i>	Artigo	Citações totais ao <i>preprint</i>	Citações diluídas		
					2022	2023	Total
IDP6	Emergence and spread of a SARS-CoV-2 variant through Europe in the summer of 2020	2021	2021	504	151	91	242
IDP7	Test sensitivity is secondary to frequency and turnaround time for COVID-19 surveillance	2020	2021	207	18	3	21
IDP8	Vaccine optimization for COVID-19: who to vaccinate first?	2020	2021	78	8	5	13
IDP9	The timing of COVID-19 transmission	2020	-	-	-	-	-
IDP10	Diverse Functional Autoantibodies in Patients with COVID-19	2021	2021	107	20	10	30

Fonte: Dados da pesquisa (2024).

Dos *preprints* aqui apresentados, apenas o IDP9 não foi publicado em uma revista e, portanto, não possui dados disponíveis. Embora cerca de 78% dos *preprints* da área médica sejam eventualmente aceitos para publicação em periódicos (Bai et al., 2023), alguns permanecem não publicados. Isso pode ocorrer devido à rejeição por parte de uma revista ou decisão dos autores de não submeter o *preprint* para publicação, por exemplo.

A diluição de citações foi calculada a partir do ano subsequente à publicação do artigo correspondente ao *preprint*, a fim de permitir que ele recebesse citações. Entre 2022 e 2023, o número total de citações diluídas dos nove *preprints* foi de 607. Observou-se uma redução significativa na diluição de citações a partir do segundo ano após a publicação do artigo.

Na tabela 2 é possível observar que a revista “Nature” publicou cinco artigos derivados de *preprints* entre 2022 e 2023 e acumulou um total de 540 citações diluídas, o que representa uma média de 108 citações por artigo com *preprint*. Esse valor destaca-se como o mais elevado dentre os dados coletados. Em contraste, as revistas “Science”, “Science Advances” e “eBioMedicine” registraram, respectivamente, 22, 34 e 11 citações diluídas no mesmo período.

Tabela 2 - Número de citações atribuídas aos *preprints* após o ano seguinte à publicação da versão revisada do artigo nos respectivos periódicos.

<i>Preprint</i>	Periódico	Ano da diluição		Total de citações diluídas	Média de diluição por artigo com <i>preprint</i>
		2022	2023		
IDP1; IDP2; IDP4; IDP6; IDP10	Nature	393	147	540	108
IDP3	Science	14	08	22	22



Preprint	Periódico	Ano da diluição		Total de citações diluídas	Média de diluição por artigo com preprint
		2022	2023		
IDP7; IDP8	Science Advances	26	08	34	17
IDP5	eBioMedicine	-	11	11	11

Fonte: Dados da pesquisa (2024).

Os resultados revelam uma grande disparidade na média de citações diluídas entre a revista “Nature” e as outras analisadas. Ainda não se sabe quais fatores podem influenciar na quantidade de citações diluídas. A princípio, pode-se pensar que os *preprints* com mais citações totais teriam mais citações diluídas. No entanto, o *preprint* com mais citações totais em nossos dados (IDP2) não foi o que teve a maior quantidade de citações diluídas. A revista “Nature” é atualmente uma das mais citadas do mundo, de acordo com o ranking do Google Acadêmico, que calcula o “Índice h5” com base em citações. No entanto, não se sabe exatamente o quanto a reputação ou citabilidade de um periódico pode influenciar a quantidade de citações diluídas. Acreditamos também que o tempo entre a postagem do *preprint* e a publicação do artigo correspondente pode aumentar as citações totais ao *preprint*, mas não necessariamente o número de citações diluídas. Isso ocorre porque a contagem das citações diluídas só começa após a publicação do artigo.

Foi possível observar que o intervalo médio de publicação entre o *preprint* e seu artigo correspondente foi curto, geralmente inferior a 1 ano. Nota-se que, mesmo após dois anos da publicação do artigo, os *preprints* continuam a ser citados. No entanto, acentua-se uma redução significativa no número de citações atribuídas aos *preprints* a partir do segundo ano após a publicação do respectivo artigo.

Existem diversas razões pelas quais os pesquisadores podem continuar a citar um *preprint*, mesmo após a publicação do artigo correspondente em uma revista científica. Isso pode ocorrer porque o autor teve acesso ao *preprint* antes da publicação do artigo, e ao sair a versão revisada, o autor não atualizou sua lista de referências ou, conforme observado por Davis (2018), há a possibilidade de que estejam copiando ou reutilizando referências antigas, ou que o *Google Scholar*, por preferência, direcione os leitores para o servidor de *preprints* em vez da revista, contribuindo assim para a continuidade das citações ao *preprint*.

4 CONSIDERAÇÕES FINAIS

Os dados preliminares apresentados neste estudo sugerem que a diluição de citações é um fenômeno real, mas que pode variar de acordo com o periódico e o tempo decorrido desde a publicação do artigo. Os dados revelam que o intervalo médio de publicação entre *preprint* e artigo é curto, geralmente menos de 1 ano e que mesmo após dois anos da publicação do artigo, os *preprints* ainda são citados, mas as citações diminuem significativamente a partir do segundo ano após a publicação. Assim, é possível afirmar que a diluição de citações é um fenômeno que ocorre, mas que parece mais comum nos primeiros anos após a publicação do artigo.



Com base nos resultados apresentados, algumas recomendações podem ser feitas para mitigar a diluição de citações. Primeiramente, os editores de periódicos devem orientar seus autores a atualizar suas listas de referências antes da publicação. Adicionalmente, pode-se estabelecer uma etapa editorial adicional, consistindo em verificar, se os *preprints* citados nos artigos já foram publicados em periódicos. Além disso, os editores e repositórios de *preprints* devem se comunicar para garantir a conexão entre o artigo publicado na revista e o *preprint* publicado no repositório de *preprints*. Por fim, os motores de busca acadêmicos, como o *Google Scholar*, devem considerar o artigo como a fonte primária de informações.

Com o avanço da pesquisa, a intenção é ampliar a análise ao incluir um volume maior de dados, visando obter uma visão mais abrangente da diluição de citações nas ciências da saúde. Além disso, considera-se importante explorar *preprints* de outras áreas para compreender como a diluição de citações ocorre em diferentes campos do conhecimento.

AGRADECIMENTO

Os autores agradecem o financiamento do CNPq e da CAPES, concedido na forma de bolsas.

REFERÊNCIAS

ABOUT medRxiv. In: **MedRxiv**: the preprint server for health sciences. 2024. [Site]. Disponível em: <https://www.medrxiv.org/content/about-medrxiv>. Acesso em: 19 jan. 2024.

ALVAREZ, G. R.; CAREGNATO, S. E. Preprints na comunicação científica da Física de Altas Energias: análise das submissões no repositório arXiv (2010-2015). **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 22, p. 104-117, 2017. DOI: <https://doi.org/10.1590/1981-5344/2830>. Disponível em: <https://www.scielo.br/j/pci/a/cVqRTjfxwpDvr7v49kCQwfh>. Acesso em: 18 mar. 2024.

AQUINO-JARQUIN, G. *et al.* Preimpresiones en biomedicina: ¿alternativa o complemento al modelo tradicional de publicación? **Gaceta Médica de México**, México, v. 154, p. 87-91, 2018. DOI: <http://dx.doi.org/10.24875/GMM.17002770>. Disponível em: https://www.gacetamedicademexico.com/frame_esp.php?id=114. Acesso em: 21 jan. 2024.

AVISSAR-WHITING, M. Downstream retraction of preprinted research in the life and medical sciences. **PLoS ONE**, San Francisco, v. 17, n. 5, p. 1-9, e0267971, May 2022. DOI: <https://doi.org/10.1371/journal.pone.0267971>. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0267971>. Acesso em: 21 jan. 2024.

BAI, A. D. *et al.* Comparison of preprint postings of randomized clinical trials on COVID-19 and corresponding published journal articles: a systematic review. **JAMA Network Open**, Chicago, v. 6, n. 1, p. e2253301, Jan. 2023. DOI: <https://doi.org/10.1001/jamanetworkopen.2022.53301>. Disponível em: <https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2800857>. Acesso em: 24 fev. 2024.

BERO, L. *et al.* Cross-sectional study of preprints and final journal publications from COVID-19 studies:



discrepancies in results reporting and spin in interpretation. **BMJ Open**, Londres, v. 11, n. 7, 2021. DOI: <https://doi.org/10.1136/bmjopen-2021-051821>. Disponível em: <https://bmjopen.bmj.com/content/11/7/e051821>. Acesso em: 27 mar. 2024.

BRACCHIGLIONE, J. A. P. *et al.* A comparison of different forward citation chasing tools for complementary searches for Cochrane systematic reviews. **Cochrane Database of Systematic Reviews**, [S. l.], 2023. DOI: <https://doi.org/10.1002/14651858.CD202301>. Disponível em: <https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD202301/full>. Acesso em: 2 mar. 2024.

CIRIMINNA, Rosaria *et al.* Reaping the benefits of open science in scholarly communication. **Heliyon**, [S. l.], v. 7, n. 12, p. e08638, Dec. 2021. DOI: <https://doi.org/10.1016/j.heliyon.2021.e08638>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2405844021027419>. Acesso em: 21 jan. 2024.

COPE - COMMITTEE ON PUBLICATION ETHICS. **Preprints**. Eastleigh, Inglaterra, [2021]. DOI: <https://doi.org/10.24318/R4WByao2>. Disponível em: <https://publicationethics.org/node/38176>. Acesso em: 20 fev. 2024.

DAVIS, Phil. Journals Lose Citations to Preprint Servers. *In*: THE SCHOLARLY KITCHEN. **Blog of Society for Scholarly Publishing**. [S. l.], May 21, 2018. Disponível em: <https://scholarlykitchen.sspnet.org/2018/05/21/journals-lose-citations-preprint-servers-repositories>. Acesso em: 21 jan. 2024.

ETTINGER, Cassandra L. *et al.* A guide to preprinting for early-career researchers. **Biology Open**, Cambridge, v. 11, n. 7, p. 1-8, July 2022. DOI: <https://doi.org/10.1242/bio.059310>. Disponível em: <https://journals.biologists.com/bio/article/11/7/bio059310/276073/A-guide-to-preprinting-for-early-career>. Acesso em: 21 jan. 2024.

FLANAGIN, Annette; FONTANAROSA, Phil B.; BAUCHNER, Howard. Preprints Involving Medical Research - Do the Benefits Outweigh the Challenges?. **JAMA Network**, Chicago, v. 324, n. 18, p. 1840-1843, Nov. 2020. DOI: <https://doi.org/10.1001/jama.2020.20674>. Disponível em: <https://jamanetwork.com/journals/jama/fullarticle/2772743>. Acesso em: 28 fev. 2024.

GEHANNO, Jean-François *et al.* Reliability of citations of medRxiv preprints in articles published on COVID-19 in the world leading medical journals. **PLoS ONE**, San Francisco, August 10, 2022. DOI: <https://doi.org/10.1371/journal.pone.0264661>. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0264661>. Acesso em: 27 mar. 2024.

HADDAWAY, Neal R.; GRAINGER, Matthew J.; GRAY, Charles T. Citationchaser: A tool for transparent and efficient forward and backward citation chasing in systematic searching. **Research Synthesis Methods**, [S. l.], v. 13, p. 533-545, 2022. DOI: <https://doi.org/10.1002/jrsm.1563>. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1002/jrsm.1563>. Acesso em: 25 mar. 2024.

HIGGINS, Jan; STEINER, Robert D. Author preprint behaviour and non-compliance with journal preprint policies: One biomedical journal's experience. **Learned Publishing**, [S. l.], v. 34, p. 389-395, Mar. 2021. DOI: <https://doi.org/10.1002/leap.1376>. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1002/leap.1376>. Acesso em: 20 fev. 2024.

PAGLIARO, Mario. Preprints in chemistry: An exploratory analysis of differences with journal articles.



Publications, Basel, v. 9, n. 1, p. 1-8, 2021. DOI: <https://doi.org/10.3390/publications9010005>. Disponível em: <https://www.mdpi.com/2304-6775/9/1/5>. Acesso em: 21 fev. 2024.

RODRÍGUEZ, Ernesto Galbán. Preprints and preprint servers as academic communication tools. **Revista Cubana de Información en Ciencias de la Salud (ACIMED)**, Havana, v. 30, n. 1, p. 1-27, 2019. Disponível em: <https://www.medigraphic.com/cgi-bin/new/resumenl.cgi?IDARTICULO=87074>. Acesso em: 20 fev. 2024.

SEVER, Richard; EISEN, Michael; INGLIS, John. Plan U: Universal access to scientific and medical research via funder preprint mandates. **PLoS Biology**, San Francisco, v. 17, n. 6, p. e3000273, June 2019. DOI: <https://doi.org/10.1371/journal.pbio.3000273>. Disponível em: <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3000273>. Acesso em: 21 fev. 2024.

SMART, Pippa. The evolution, benefits, and challenges of preprints and their interaction with journals. **ScienceEditing**, Gangnam-gu, v. 9, n. 1, p. 79-84, Feb. 2022. DOI: <https://doi.org/10.6087/kcse.269>. Disponível em: <https://www.escienceediting.org/journal/view.php?doi=10.6087/kcse.269>. Acesso em: 21 jan. 2024.