

XXV ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – XXV ENANCIB
GT 7 – Produção e Comunicação da Informação em Ciência, Tecnologia & Inovação

PERIÓDICOS PERDEM CITAÇÕES AO PUBLICAR ARTIGOS ORIGINADOS DE PREPRINTS?
ANÁLISE DA DILUIÇÃO DE CITAÇÕES POR PREPRINTS DO MEDRXIV

DO JOURNALS LOSE CITATIONS WHEN PUBLISHING ARTICLES FROM PREPRINTS? ANALYSIS
OF CITATION DILUTION BY MEDRXIV PREPRINTS

Pedro Henrique da Silva Rodrigues – Universidade Federal do Rio Grande do Sul (UFRGS)
Sonia Elisa Caregnato – Universidade Federal do Rio Grande do Sul (UFRGS)

Modalidade: Trabalho Completo

Resumo: A pesquisa analisou a diluição de citações na área da saúde, comparando as citações a preprints do medRxiv com as citações aos artigos correspondentes. Por meio de análise quantitativa, foi determinado o número de citações a preprints realizadas após a publicação dos respectivos artigos. Utilizou-se as bases de dados medRxiv, OpenAlex e Crossref, bem como suas APIs, para coleta dos dados. Foram identificadas 58.737 citações diluídas referentes a 9.412 preprints, equivalendo a 31% do total de citações a preprints (n=186.193) e 7% do total de citações a artigos (n=839.532). A média foi 6,24 (mediana=2; moda=1), com 76% dos preprints acumulando entre 1 e 5 citações diluídas. Mais de 70% dessas citações ocorreram nos 12 primeiros meses após a publicação do artigo correspondente. Conclui-se que citar preprints em vez dos artigos publicados resulta em perda de citações para os periódicos. Essa prática também gera preocupações quanto à confiabilidade da informação, pois o conteúdo do preprint pode não corresponder ao do artigo final. Por fim, são necessárias mais pesquisas para compreender a extensão do problema.

Palavras-chave: preprints; citações; diluição de citações.

Abstract: The research analyzed citation dilution in the health sciences by comparing citations to medRxiv preprints with citations to their corresponding articles. Through quantitative analysis, the number of citations to preprints made after the publication of their respective articles was determined. Data was collected using the medRxiv, OpenAlex, and Crossref databases and their APIs. A total of 58,737 diluted citations were identified for 9,412 preprints, accounting for 31% of all preprint citations (n=186,193) and 7% of all article citations (n=839,532). The average was 6.24 (median=2; mode=1), with 76% of preprints accumulating between 1 and 5 diluted citations. Over 70% of these citations occurred within the first 12 months following the publication of the corresponding article. It is concluded that citing preprints instead of the published articles results in a loss of citations for journals. This practice also raises concerns about the reliability of the information, as the content of the preprint may not match that of the final article. Finally, further research is needed to understand the extent of the problem.

Keywords: preprints; citations; dilution of citations.

1 INTRODUÇÃO

Os *preprints* são trabalhos científicos disponibilizados *on-line* antes da revisão por pares. Geralmente, são disponibilizados em repositórios que lhes atribuem um Identificador de Objeto Digital (DOI®), tornando-os um registro acadêmico permanente e citável (Ettinger *et al.*, 2022).

A citação a *preprints* vem sendo problematizada por parte da comunidade científica, que argumenta que citar esse tipo de publicação exige uma interpretação responsável do seu conteúdo, e que introduzir uma cultura de citação responsável é difícil, especialmente diante do crescimento das publicações predatórias (Silva, 2018).

Nesse contexto, a citação a *preprints* em áreas que exigem precisão, como a da saúde, torna-se uma questão delicada, pois a ausência de revisão significa que não há como atestar a confiabilidade das informações, abrindo margem para o uso de dados equivocados ou desatualizados. Diante disso, organizações relacionadas a publicações científicas da área médica, como a American Medical Writers Association, European Medical Writers Association e International Society for Medical Publication Professionals (2021), aconselham que *preprints* não sejam citados e, caso seja indispensável, que a referência seja feita na forma de comunicação pessoal, com o *link* em nota de rodapé.

Apesar disso, a presença de *preprints* em lista de referências tem sido frequente e tem provocado a diluição de citações, que se caracteriza como o fenômeno que ocorre quando um *preprint* já foi publicado como artigo revisado por pares, mas continua sendo citado no lugar da versão final publicada. Um exemplo dessa diluição foi identificado por Gehanno *et al.* (2022), que observaram que 40% de uma amostra de 182 *preprints* sobre *Covid-19*, depositados entre 2020 e 2021 no repositório medRxiv®, foram citados após a publicação dos respectivos artigos revisados por pares.

A diluição de citações pode ter várias causas, como a falta de acesso ao artigo publicado (quando este está em acesso fechado), desconhecimento de que o *preprint* já foi publicado como artigo, falhas na atualização de gerenciadores de referências (Fraser *et al.* 2020), ou ainda a ausência de vinculação entre o *preprint* e a versão final que dificulta o acesso a informações sobre o *status* da publicação (Davis, 2018). Outros fatores incluem a escolha deliberada em citar o *preprint* para contextualização, preferência pessoal do citante devido a

informações que foram suprimidas na versão final ou por discordância em relação a alterações feitas durante a revisão por pares.

Chaleplioglou e Koulouris (2023), assim como Moshontz *et al.* (2021), destacam que o uso de DOIs diferentes para o *preprint* e o artigo correspondente – prática recomendada pelo Comitê de Ética em Publicações¹ e Crossref[®] – cria dois conjuntos separados de citações. A Crossref reconhece essa separação e recomenda que as citações de cada versão sejam contadas de forma independente, já que, embora sejam fruto do mesmo trabalho, são documentos distintos².

Por outro lado, Gao, Wu e Zhu (2020) sugerem unir as citações ao *preprint* e ao artigo final em um único registro para evitar a diluição de citações. No entanto, essa proposta esbarra em questões técnicas e éticas: em alguns casos, os autores do *preprint* e do artigo não são exatamente os mesmos, o que poderia fazer com que autores recebessem ou deixassem de créditos indevidamente. Além disso, a citação também serve como reconhecimento pelo trabalho editorial e divulgação do artigo. O sistema de recompensas acadêmicas seria distorcido ao atribuir à revista citações que o *preprint* recebeu de forma independente do processo editorial – mesmo que tenham sido feitas quando o artigo já estava disponível. Impediria também uma avaliação justa, transparente e precisa da contribuição científica de cada documento (*preprint* ou artigo) em seus diferentes estágios de desenvolvimento.

O problema de citar o *preprint* em vez do artigo correspondente vai além da possível perda de citações para as revistas. Uma preocupação recorrente é de que podem ocorrer mudanças significativas entre as duas versões: Gehanno *et al.* (2022) encontraram divergências em títulos, dados ou conclusões em quase metade dos 124 *preprints* sobre *Covid-19* que analisaram. Já em outro estudo, Bero *et al.* (2021) observaram que 34% dos 67 *preprints* sobre *Covid-19* mostravam resultados diferentes em comparação aos artigos. Isso significa que o conteúdo citado de um *preprint* pode não estar presente ou ter sido alterado no artigo final, o que pode causar confusão e tornar a citação imprecisa.

Neste trabalho, apresenta-se parte dos resultados da dissertação de mestrado intitulada "Análise da diluição de citações entre preprints do medRxiv e artigos publicados em periódicos da área da saúde". O objetivo geral foi analisar o fenômeno da diluição de citações em artigos da área da saúde originados de preprints disponibilizados no medRxiv entre 2019

¹ <https://publicationethics.org/guidance/case/can-two-be-assigned-same-manuscript>

² <https://archive.org/details/gmail-crossref-re-citations/page/n7/mode/2up>

e 2021. Além disso, procurou-se estimar o Aumento Potencial de Citações (APCit), uma taxa que calcula, em porcentagem, o aumento potencial no número de citações que um artigo ou revista poderia ter alcançado caso não houvesse a diluição. A importância da pesquisa está em compreender melhor os efeitos do uso crescente de *preprints* na comunicação científica e os possíveis impactos nas citações de artigos e periódicos.

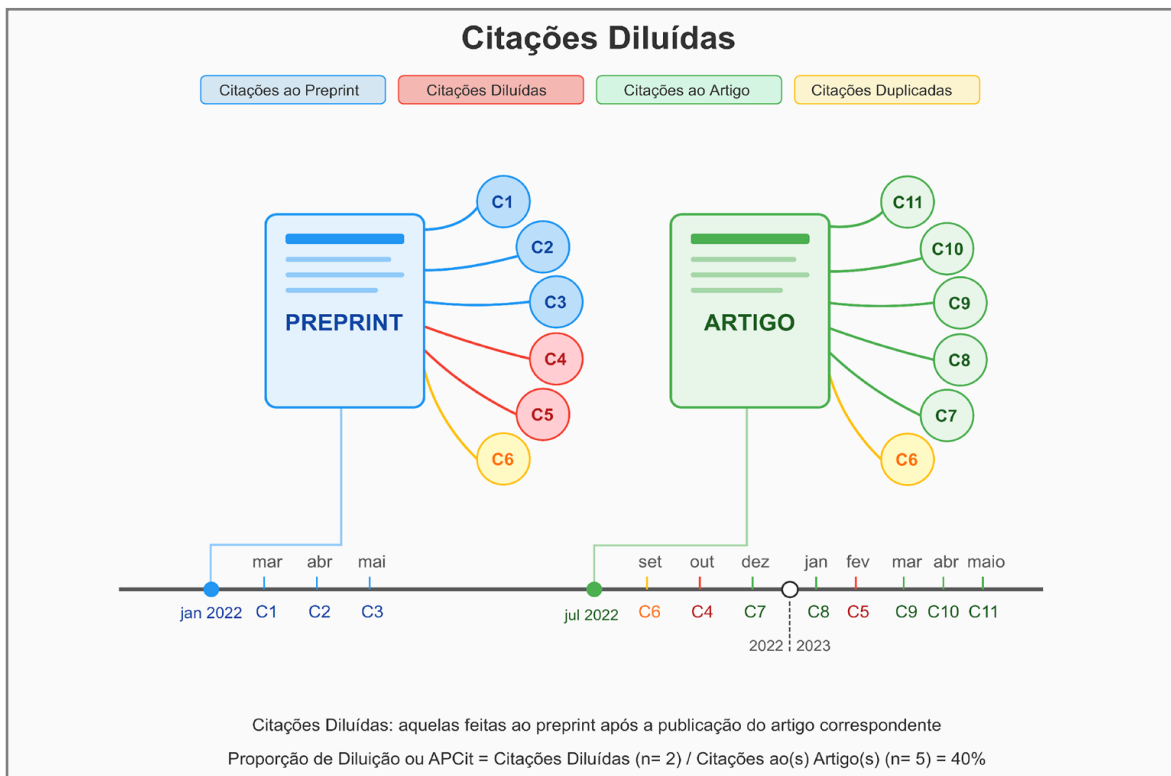
2 PROCEDIMENTOS METODOLÓGICOS

Este estudo adota uma abordagem quantitativa, descritivo-exploratória, para investigar a diluição de citações causada por *preprints* na área médica. Primeiramente, foram identificados *preprints* publicados no medRxiv entre 2019 e 2021, coletando via API os metadados (DOI, título, autoria, data de publicação e, quando disponível, DOI do artigo correspondente, se publicado). Em seguida, filtrou-se esse conjunto para incluir apenas *preprints* que viraram artigos revisados por pares entre 2019 e 2022. Esses intervalos de publicação foram definidos a fim de garantir uma janela adequada para publicação e citação – ou seja, tempo suficiente para que os *preprints* e artigos pudessem ser publicados e citados. Já as citações, coletadas na etapa seguinte, abrangeram o período da publicação do trabalho citado até 31 de dezembro de 2023, uma vez que as coletas ocorreram em 2024.

Na etapa seguinte, extraiu-se do OpenAlex™ metadados sobre os artigos – título, periódico, índice h do periódico, data de publicação e DOI – e as listas de citantes dos *preprints* e dos artigos correspondentes – com DOI e data de publicação –, permitindo classificar cada citação como anterior ou posterior à publicação do artigo correspondente. Verificou-se também, via API Crossref, as listas de referências dos citantes, buscando classificar o *corpus* em citações diluídas, citações duplicadas (quando o *preprint* e o artigo são citados juntos) e citações não correspondentes. As citações não correspondentes ocorreram quando nem o *preprint* nem o artigo constavam na lista de referências do citante, desse modo elas não foram incluídas nas análises deste estudo. Essa ausência pode ocorrer devido ao não depósito ou depósito parcial de referências nos metadados da Crossref pelos editores.

A Figura 1 exemplifica a ocorrência de citações duplicadas e diluídas, objeto de estudo deste trabalho.

Figura 1 – Identificação de citações diluídas e duplicadas por análise temporal entre *preprints* e artigos publicados



Fonte: elaborado pelos autores (2025).

Por fim, quantificou-se a diluição de citações e comparou-se o total de citações diluídas com o total de citações recebidas pelos artigos correspondentes. Em seguida, foi calculado o Aumento Potencial de Citações (APCit), uma taxa gerada através da fórmula $APCit (\%) = (\text{Número de Citações Diluídas} / \text{Número de Citações ao Artigo ou Revista}) \times 100$. O APCit expressa a porcentagem de crescimento que o número de citações recebidas pelo artigo (ou revista) poderia ter alcançado caso as citações feitas ao *preprint* tivessem sido feitas ao artigo publicado.

Determinou-se também a duração da diluição de citações ao longo do tempo, analisando por quanto tempo os *preprints* continuaram a ser citados após a publicação dos artigos. Adicionalmente, foi verificada a relação entre o Aumento Potencial Geral de Citações (APCit-G) – uma variação do APCit que considera o total de citações de uma revista no período de diluição – e o índice h das revistas que apresentaram citações diluídas.

As fontes de dados foram:

- a) medRxiv, para coletar pares de *preprint*–artigo;

- b) OpenAlex, para obter metadados das publicações e citações;
- c) Crossref, para verificação e classificação das citações.

Os dados foram extraídos por meio das APIs de cada plataforma, usando *scripts* em Python, organizados e analisados em planilhas do Excel®.

O medRxiv foi selecionado porque, desde seu lançamento em junho de 2019, tornou-se o principal repositório na área da saúde. Nele, todos os *preprints* recebem DOI, o que facilita sua identificação, citação e indexação em mecanismos de busca. Além disso, pesquisas como a de Gehanno *et al.* (2022) já mostraram diluição de citações em *preprints* do medRxiv, reforçando sua relevância para este estudo. O OpenAlex foi escolhido por ser uma das poucas fontes abertas e gratuitas que oferecem metadados completos de citantes, incluindo datas de publicação – um requisito essencial para a análise deste estudo. Por fim, a Crossref foi usada para verificar e classificar as referências, pois é a principal base de depósito de DOIs e reúne metadados detalhados dos trabalhos, inclusive listas de referência fornecidas pelos editores.

3 RESULTADOS

Foram analisados 18.596 pares de *preprints* depositados no medRxiv entre 2019 e 2021 e seus artigos correspondentes publicados em periódicos entre 2019 e 2022. A coleta de citações para esses trabalhos foi realizada até 31 de dezembro de 2023. Os *preprints* acumularam ao todo 186.193 citações (média de 10 por *preprint*), enquanto os artigos correspondentes receberam 839.532 citações (média de 54 por artigo).

3.1 Número de Citações Diluídas

Foram identificadas 58.737 citações diluídas, distribuídas em 9.412 *preprints*. Essas citações representam 31% do total das citações aos *preprints* (n=186.193) e 7% em relação ao total das citações aos artigos (n=839.532). A média de citações diluídas foi de 6,24, com mediana de 2 e moda de 1, o que indica que, embora o volume total de citações diluídas seja alto, elas se concentram em um número relativamente pequeno de *preprints*. 76% dos 9.412 *preprints* registraram entre 1 e 5 citações diluídas.

3.2 Duração da Diluição

A duração da diluição refere-se ao período durante o qual os *preprints* continuaram recebendo citações após a publicação dos respectivos artigos. A duração máxima observada

foi de 51 meses. No entanto, mais de 70% das citações diluídas ocorreram nos primeiros 12 meses após a publicação do artigo. A média de duração da diluição foi de 9 meses, com mediana e moda de 8 meses. Esses dados indicam que as citações aos *preprints* tendem a diminuir de forma rápida e progressiva após a publicação do artigo correspondente. Ainda assim, os *preprints* continuaram a ser citados após um ano (20,73%), dois anos (7,46%) e até três anos (1,10%) da publicação dos artigos. A Tabela 1 mostra a distribuição de citações aos *preprints* ao longo dos meses seguintes à publicação dos artigos.

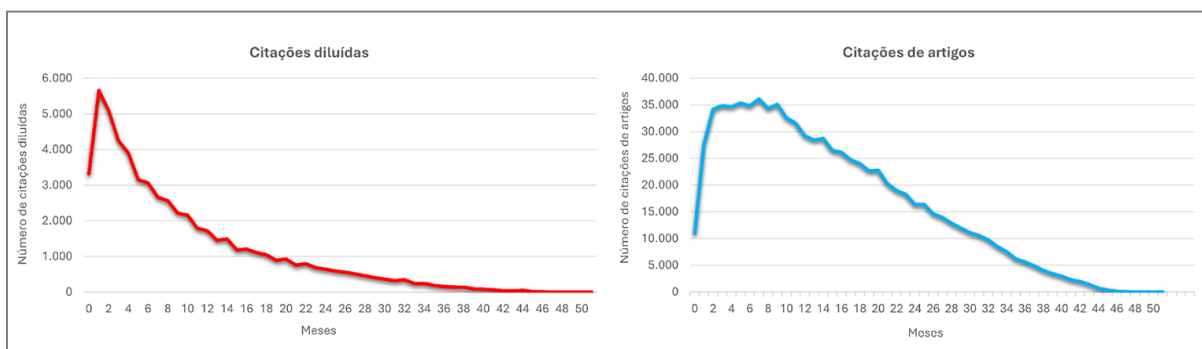
Tabela 1 – Distribuição das citações diluídas por intervalo de meses após a publicação do artigo

Intervalo de meses	Citações diluídas	Percentual
0-12	41.533	70,71%
13-24	12.179	20,73%
25-36	4.379	7,46%
37-48	645	1,10%
49-51	2	0,00%
Total	58.737	100%

Fonte: dados da pesquisa (2025).

Adicionalmente, comparou-se a distribuição mensal das citações diluídas com a das citações de artigos. O Gráfico 1 mostra os resultados dessa comparação:

Gráfico 1 – Distribuição mensal das citações diluídas e das citações de artigos



Fonte: dados da pesquisa (2025).

Observa-se que, em 24 meses, as citações diluídas caíram de mais de 5.000 para cerca de 600, ou seja, uma redução de 88%. No mesmo intervalo, as citações aos artigos passaram de 36.000 para cerca de 18.000, uma queda de aproximadamente 50%. Esses dados indicam

que as citações tanto aos *preprints* quanto aos artigos diminuem ao longo do tempo, mas os *preprints* sofrem um declínio proporcionalmente mais acentuado, cerca de 38% maior que dos artigos.

3.3 Impacto da Diluição de Citações nos Artigos

Neste estudo, o impacto da diluição de citações foi medido pelo Aumento Potencial de Citações (APCit). O APCit é uma estimativa que considera um cenário ideal, sem a diluição de citações, onde todas as citações diluídas são atribuídas ao artigo pelos autores citantes. Ele é calculado utilizando o número de citações ao artigo (ou à revista) como denominador, conforme a fórmula: $APCit = (\text{citações diluídas} / \text{citações ao artigo ou revista}) \times 100$.

A análise incluiu 9.412 pares de preprint-artigo, nos quais todos os *preprints* foram citados após a publicação do artigo correspondente. Deste total, 9.330 artigos receberam uma ou mais citações. Os 82 artigos restantes não foram citados, embora seus respectivos *preprints* tenham acumulado citações diluídas. Dessa forma, o APCit foi aplicado apenas aos 9.330 artigos. A Tabela 2 apresenta um resumo estatístico dos resultados obtidos.

Tabela 2 – Estatísticas descritivas do Aumento Potencial de Citações (APCit para 9.330 artigos)

Medida	Valor	Observações
Contagem (N)	9.330	número de artigos
Média	15,37%	APCit médio
Moda	20,00%	APCit mais frequente
Desvio Padrão	29,22%	dispersão dos dados
Mínimo	0,17%	menor APCit observado
1º Quartil (Q1)	4,55%	25% dos artigos ≤ APCit 4,55%
Mediana (Q2)	8,70%	50% dos artigos ≤ APCit 8,70%
3º Quartil (Q3)	16,67%	75% dos artigos ≤ APCit 16,67%
Máximo	1.200%	maior APCit observado

Fonte: dados da pesquisa (2025).

O APCit médio foi de 15%, contudo, esse aumento potencial não é distribuído de forma uniforme entre os artigos, pois há grande variação nos dados devido à presença de *outliers*. No entanto, o cálculo dos quartis não é afetado por esses valores extremos, o que assegura uma representação mais fiel da distribuição dos dados. Observa-se que a maioria dos artigos (75%) teria um aumento potencial de até 16,67% no número de citações, enquanto metade dos artigos apresentaria um aumento de até 8,70%.

Os *outliers* surgiram em situações específicas em que a diluição das citações foi desproporcionalmente alta em relação ao número de citações recebidas pelos artigos correspondentes. Por exemplo, o APCit de 1.200% ocorreu em um caso em que o título do artigo era significativamente diferente do *preprint*³, o conteúdo era mais completo do que do artigo⁴ que, nesse caso, foi publicado como um artigo curto, em um periódico focado apenas no anúncio de resultados. Além disso, a alta relevância e atualidade do tema abordado no *preprint* – o sequenciamento genômico dos primeiros casos de *Covid-19* no Equador – certamente atraiu grande atenção inicial e resultou em um número elevado de citações para o *preprint*.

3.4 Impacto da Diluição de Citações nos Periódicos

Para avaliar o impacto da diluição de citações em periódicos que publicaram artigos originados de *preprints*, utilizou-se o Aumento Potencial Geral de Citações (APCit-G). O APCit-G é calculado dividindo o total de citações diluídas pelo total de citações recebidas por todos os artigos do periódico durante o período de diluição, em seguida, multiplicando-se o resultado por 100. O período de diluição corresponde aos anos em que os *preprints* continuaram a ser citados após a publicação dos artigos. Por exemplo, se um periódico publicou artigos originados de *preprints* em 2021 e 2022, e esses *preprints* diluíram citações em 2022 e 2023, o intervalo de diluição para esse periódico será de 2022 a 2023. Essa medida ajuda a entender se a prática de citar *preprints* em vez dos artigos publicados afeta significativamente a contagem geral de citações aos periódicos.

A análise foi realizada nas 2.003 revistas que publicaram os artigos originados dos 9.412 *preprints* que diluíram citações. Na Tabela 3 é apresentada a distribuição das revistas com base em intervalos de APCit-G.

³ Preprint: 10.1101/2020.06.11.20128330

⁴Artigo: 10.1128/mra.00996-20

Tabela 3 – Distribuição das revistas segundo os intervalos de APCit-G

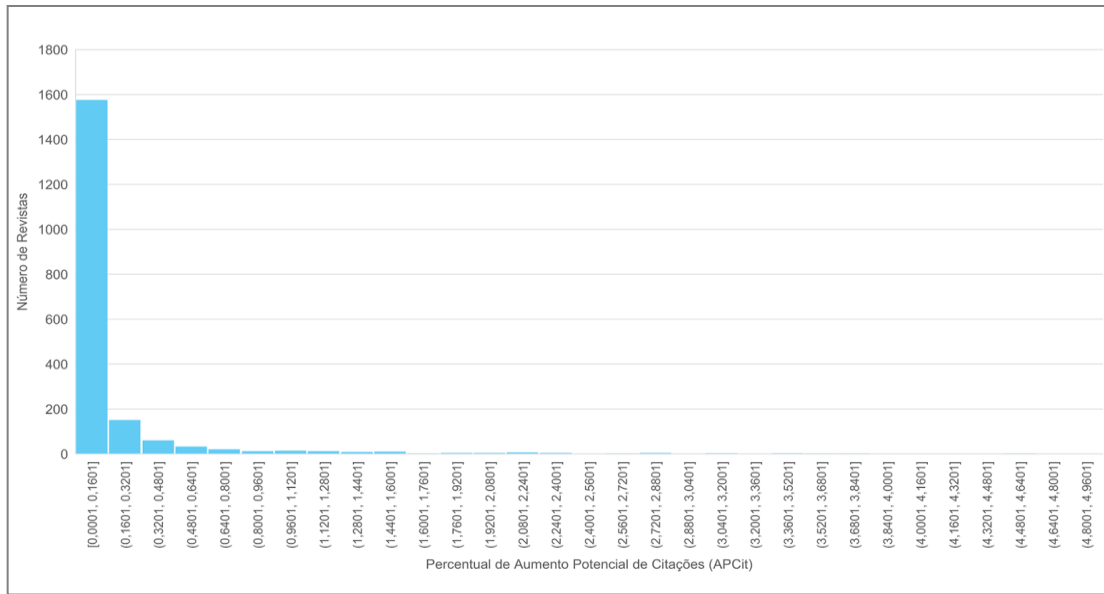
Intervalo APCit-G (%)	Número de revistas	Percentual
0 † 5	1.967	98,20%
5 † 10	16	0,80%
10 † 15	8	0,40%
15 † 20	3	0,15%
20 † 25	2	0,10%
25 † 30	3	0,15%
35 † 40	2	0,10%
40 † 45	1	0,05%
130 † 135	1	0,05%
Total	2.003	100%

Fonte: dados da pesquisa (2025).

Conforme os dados da Tabela 3, 98% das revistas apresentaram APCit-G entre 0 e 4,9%, enquanto apenas 0,8% apresentaram intervalos entre 5% e 9,9%. O Gráfico 2 apresenta uma análise mais detalhada da faixa de 0% a 4,9%, que engloba 98% das revistas analisadas.

Observa-se no Gráfico 2 que aproximadamente 79% das revistas teriam um aumento de até 0,16% no número total de citações. Embora este impacto possa ser considerado baixo, deve-se ressaltar que é necessário ponderar o contexto de cada revista (seu tamanho, número de publicações e citações recebidas). Por exemplo, 0,1% em uma revista com 1 milhão de citações equivale a 100 mil citações diluídas, enquanto em uma revista com mil citações equivale a apenas uma citação.

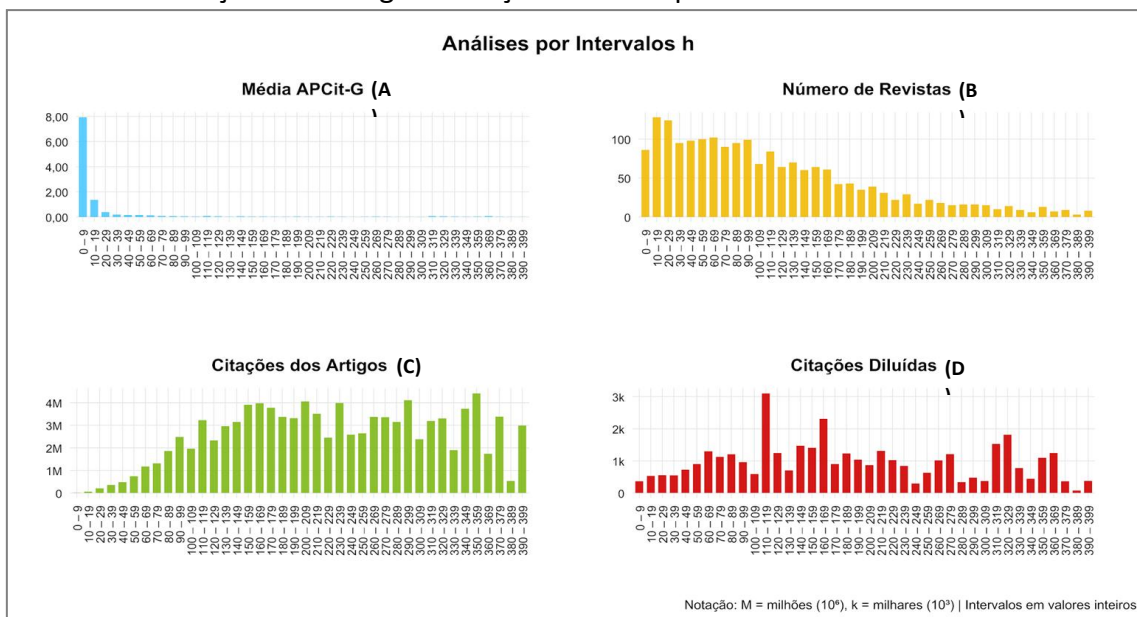
Gráfico 2 – Distribuição do APCit-G na faixa de 0% a 4,9%



Fonte: dados da pesquisa.

Para entender melhor como a diluição de citações afeta diferentes revistas, foi analisado se existe relação entre o APCit-G e o índice h das publicações. O índice h foi escolhido por levar em consideração tanto a quantidade de artigos quanto o número de citações. Nesta análise, as 2.003 revistas foram divididas em grupos com intervalos de 10 pontos de índice h e foi calculada a média de APCit-G para cada grupo, a quantidade de revistas, número de citações de artigos e citações diluídas por intervalo de índice h. O Gráfico 3 mostra os resultados.

Gráfico 3 – Distribuições das médias de APCit-G, quantidade de revistas, citações aos artigos e citações diluídas por intervalos de índice h



Fonte: dados da pesquisa (2025).

A relação entre os intervalos de índice h e o número absoluto de citações diluídas (D) mostra que revistas com índice h mais baixo (0-9) tem um número relativamente menor de citações diluídas (368 citações diluídas) do que aquelas com índice h mais elevado. No entanto, nota-se que as revistas com índice h menor registraram médias mais elevadas de APCit-G (A). Isso é visível na faixa de índice h entre 0 e 9, que apresentou a maior média de APCit-G (7,9%).

A tendência de diminuição do APCit-G conforme o índice h aumenta sugere que periódicos com menor visibilidade, produtividade e citações podem ser mais impactados pela diluição de citações. Isso ocorre porque, para essas revistas, mesmo um número relativamente pequeno de citações diluídas pode representar uma parte mais significativa do total de citações recebidas. Nesse sentido, para fins de impacto (medido pelo APCit-G), as revistas com baixo índice h são mais afetadas pela diluição, pois as citações diluídas representam uma fatia maior de seu volume total de citações (C).

Quanto a distribuição das revistas (B), a maioria (54%, n= 1.085) possui índice h inferior a 109, com a maior concentração nos intervalos de 10-19 (n= 128) e 20-29 (n= 124). Isso significa que grande parte do *corpus* analisado se enquadra em faixas de índice h que, segundo a tendência geral observada, são mais suscetíveis a um maior impacto potencial da diluição.

3.5 Citações Duplicadas

A análise das citações diluídas não contabilizou as citações duplicadas, ou seja, aquelas em que o *preprint* e o artigo correspondente são citados simultaneamente. As citações duplicadas foram analisadas separadamente, onde constatou-se um total de 1.076 pares (5,79%), do *corpus* de 18.596 pares de *preprints* e artigos, que foram citados em conjunto. No total, foram identificadas 1.899 citações desse tipo. Essa quantidade é considerada pequena quando comparada às citações diluídas (n= 58.737), representando 3% do total de citações diluídas. Isso sugere que não é comum que os autores citem simultaneamente o *preprint* e o respectivo artigo publicado.

A maioria dos pares com citações duplicadas (71,84%) foi citada em conjunto apenas uma vez. 159 pares (14,78%) foram citados duas vezes e 62 (5,76%) foram citados três vezes. Combinando esses dados, cerca de 20% dos pares foram citados entre duas e três vezes. A partir de quatro citações duplicadas por par, a quantidade de pares diminui consideravelmente.

As causas das citações duplicadas não foram identificadas neste estudo, mas podem estar relacionadas à ausência de diretrizes claras para a citação de *preprints* e artigos, ao interesse dos autores em evidenciar ou comparar as versões, ou ainda à tentativa de aumentar artificialmente o número de citações aos próprios autores.

4 CONSIDERAÇÕES FINAIS

Verificou-se que o total de citações diluídas por *preprints* correspondeu aproximadamente 7% do total de citações aos artigos. A prática de citar *preprints* após a publicação do artigo resulta na perda de citações para os periódicos científicos. O impacto dessa perda pode variar conforme o periódico, mas parece ser proporcionalmente maior em periódicos com índice h mais baixo, o que pode dificultar seu crescimento, já que as métricas de citação são comumente usadas na avaliação das revistas. Entretanto, a relação entre o índice h e a diluição de citações requer mais estudos. Cabe mencionar que estudos como os de Colavizza *et al.* (2024) e Fu e Hughey (2019) apontam que artigos previamente disponibilizados como *preprints* tendem a acumular mais citações do que aqueles publicados diretamente, fenômeno conhecido como vantagem de citação. Contudo, permanece incerto se, para os periódicos, essa vantagem seria capaz de superar a eventual perda de citações decorrente da diluição. Para elucidar essa questão, são necessários estudos que analisem o mesmo conjunto de periódicos, controlando a heterogeneidade.

A correta vinculação entre *preprints* e seus artigos correspondentes é fundamental, mas não evita completamente a diluição de citações. No presente estudo, por exemplo, todos os pares *preprint*-artigo analisados estavam corretamente vinculados.

Embora os *preprints* acelerem a disseminação do conhecimento, citá-los em vez da versão revisada por pares levanta preocupações quanto à confiabilidade das informações. Isso ocorre porque os *preprints* podem sofrer alterações significativas – nos dados, títulos, conclusões – em relação à versão final publicada. Portanto, citar o artigo revisado por pares é a prática mais recomendada.

Uma limitação deste trabalho é ter se concentrado apenas em *preprints* do repositório medRxiv. Embora relevante, o medRxiv é apenas uma parte do universo de *preprints*, o que pode restringir a generalização dos resultados. Isso também pode subestimar o impacto da diluição nos periódicos, pois eles publicam artigos originados de *preprints* de diversas fontes, não apenas do medRxiv. Além disso, o estudo adotou uma abordagem quantitativa. Uma

análise qualitativa complementar seria necessária para investigar aspectos complementares como o porquê de os autores citarem *preprints* em vez dos artigos, se o conteúdo citado do *preprint* permanece o mesmo no artigo final e quais as características dos trabalhos que citaram a versão *preprint*.

Os *preprints* oferecem vários benefícios e estão alinhados com a ciência aberta. No entanto, isso não deve afastar o senso crítico sobre seus possíveis impactos negativos, especialmente quando são usados como referência bibliográfica no lugar dos artigos correspondentes. Por fim, ressalta-se a necessidade de pesquisas futuras para explorar questões como a variação dessa prática em outras áreas do conhecimento ou fora do contexto da pandemia, a relação entre a diluição e o tipo de acesso do artigo (aberto ou fechado) e possíveis diferenças na diluição entre pares vinculados e não vinculados.

AGRADECIMENTOS

Os autores agradecem o financiamento da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e do CNPq, concedido na forma de bolsas.

REFERÊNCIAS

AMERICAN MEDICAL WRITERS ASSOCIATION; EUROPEAN MEDICAL WRITERS ASSOCIATION; INTERNATIONAL SOCIETY FOR MEDICAL PUBLICATION PROFESSIONALS. AMWA-EMWA-ISMP joint position statement on medical publications, preprints, and peer review. **Current Medical Research and Opinion**, [S.l.], v. 37, n. 5, p. 861–866, 2021. Disponível em: <https://doi.org/10.1080/03007995.2021.1900365>. Acesso em: 8 maio 2025.

BERO, Lisa; LAWRENCE, Rosa; LESLIE, Louis; CHIU, Kellia; MCDONALD, Sally; PAGE, Matthew J.; GRUNDY, Quinn; PARKER, Lisa; BOUGHTON, Stephanie; KIRKHAM, Jamie J.; FEATHERSTONE, Robin. Cross-sectional study of preprints and final journal publications from COVID-19 studies: discrepancies in results reporting and spin in interpretation. **BMJ Open**, London, v. 11, e051821, 2021. Disponível em: <https://doi.org/10.1136/bmjopen-2021-051821>. Acesso em: 10 ago. 2025.

COLAVIZZA, Giovanni; CADWALLADER, Laura; LAFLAMME, Marc; DOZOT, Guillaume; LECORNEY, Simon; RAPPO, David; HRYNASZKIEWICZ, Iain. An analysis of the effects of sharing research data, code, and preprints on citations. **PLOS ONE**, [S. l.], v. 19, 2024. Disponível em: <https://doi.org/10.1371/journal.pone.0311493>. Acesso em: 10 ago. 2025.

CHALEPLIOGLOU, Artemis; KOULOURIS, Alexandros. Preprint paper platforms in the academic scholarly communication environment. **Journal of Librarianship and Information Science**, [S. l.], v. 55, n. 1, p. 43-56, 2023. Disponível em: <https://doi.org/10.1177/09610006211058908>. Acesso em: 10 ago. 2025.

XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB
Rio de Janeiro, RJ - 03 a 07 de novembro de 2025

DAVIS, Phil. Journals lose citations to preprint servers. *In: The Scholarly Kitchen*, [S.l.], 2018. Disponível em: <https://scholarlykitchen.sspnet.org/2018/05/21/journals-lose-citations-preprint-servers-repositories/>. Acesso em: 8 maio 2025.

ETTINGER, Cassandra L.; SADANANDAPPA, Madhumala K.; GÖRGÜLÜ, Kivanç; COGHLAN, Karen L.; HALLENBECK, Kenneth K.; PUEBLA, Iratxe. A guide to preprinting for early-career researchers. **Biology Open**, Cambridge, v. 11, n. 7, e059310, 2022. Disponível em: <https://doi.org/10.1242/bio.059310>. Acesso em: 10 ago. 2025.

FU, Darwin Y.; HUGHEY, Jacob J. Meta-research: Releasing a preprint is associated with more attention and citations for the peer-reviewed article. **eLife**, [S. l.], v. 8, e52646, 2019. Disponível em: <https://doi.org/10.7554/eLife.52646>. Acesso em: 10 ago. 2025.

FRASER, Nicholas; MOMENI, Fakhri; MAYR, Philipp; PETERS, Isabella. The relationship between bioRxiv preprints, citations and altmetrics. **Quantitative Science Studies**, Leuven, v. 1, n. 2, p. 618-638, 2020. Disponível em: https://doi.org/10.1162/qss_a_00043. Acesso em: 10 ago. 2025.

GAO, Yong; WU, Qiang; ZHU, Linna. Merging the citations received by arXiv-deposited e-prints and their corresponding published journal articles: problems and perspectives. **Information Processing & Management**, London, v. 57, n. 5, 102267, set. 2020. Disponível em: <https://doi.org/10.1016/j.ipm.2020.102267>. Acesso em: 10 ago. 2025.

GEHANNO, Jean-François; GROSJEAN, Julien; DARMONI, Stefan J.; ROLLIN, Laetitia. Reliability of citations of medRxiv preprints in articles published on COVID-19 in the world leading medical journals. **PLOS ONE**, San Francisco, v. 17, 2022. Disponível em: <https://doi.org/10.1371/journal.pone.0264661>. Acesso em: 10 ago. 2025.

MOSHONTZ, Hannah; BINION, Grace; WALTON, Haley; BROWN, Benjamin T.; SYED, Moin. A guide to posting and managing preprints. **Advances in Methods and Practices in Psychological Science**, [S. l.], v. 4, n. 2, 2021. Disponível em: <https://doi.org/10.1177/25152459211019948>. Acesso em: 10 ago. 2025.

SILVA, Jaime A. The preprint debate: What are the issues? **Medical Journal Armed Forces India**, [S. l.], v. 74, n. 2, p. 162-164, abr. 2018. Disponível em: <https://doi.org/10.1016/j.mjafi.2017.08.002>. Acesso em: 10 ago. 2025.

SMART, Pippa. The evolution, benefits, and challenges of preprints and their interaction with journals. **Science Editing**, [S. l.], v. 9, n. 1, p. 79-84, fev. 2022. Disponível em: <https://doi.org/10.6087/kcse.269>. Acesso em: 10 ago. 2025.