

---

# Avaliação da Diluição de Citações entre *Preprints* e Artigos da Área da Saúde

*Assessment of the Citation Dilution Between Health-Related Preprints and Journal Articles*

---

Pedro Henrique da Silva Rodrigues (1), Lucas George Wendt (2)

Sônia Elisa Caregnato (3)

(1) Universidade Federal do Rio Grande do Sul, Brasil, 790.rodrigues@gmail.com

(2) lucas.george.wendt@gmail.com

(3) sonia.caregnato@ufrgs.br



## Resumo

**Introdução:** fenômeno da diluição de citações ocorre quando *preprints* continuam a ser citados mesmo após a publicação de seus artigos correspondentes. Embora os *preprints* recebam um DOI, tornando-se parte do registro acadêmico, sua citação representa um desafio para editores, pois há o risco de as citações se concentrarem na versão preliminar, impactando as métricas do artigo e da revista. **Metodologia:** esta pesquisa de caráter descritivo-exploratório, com abordagem quantitativa, teve como objetivo analisar a diluição em *preprints* do medRxiv depositados entre 2019 e 2021, cujos artigos foram publicados até 2022. O corpus analisado incluiu 18.596 *preprints*. A coleta de dados utilizou o OpenAlex e a validação via Crossref permitiu classificar as citações como “diluídas” (citação exclusiva ao *preprint* após a publicação do artigo). Foi proposto um indicador original, denominado Aumento Potencial de Citações (APCit), por meio do qual o impacto da diluição foi quantificado. O APCit estima o aumento percentual de citações que o artigo ou revista poderia ter alcançado se a diluição não tivesse ocorrido. **Resultados:** dos *preprints* analisados, 9.412 (51%) receberam citações diluídas, totalizando 58.737, o que representa 7% das citações totais recebidas pelos artigos correspondentes. O pico de citações diluídas ocorre no primeiro mês após a publicação do artigo, e a diluição cessa em média após 8 meses. A duração da diluição de citações variou de 0 a 51 meses após a publicação dos artigos revisados por pares, concentrando-se majoritariamente entre 1 e 12 meses (70%), seguida pelo intervalo de 13 a 24 meses (21%), com menor ocorrência entre 25 e 51 meses (9%). A taxa média de APCit para os 9.330 artigos analisados foi de 15%. A mediana do APCit-R foi de 10%, indicando que 50% das revistas poderiam ter um aumento de até 10% nas citações de artigos originados de *preprints*. **Conclusão:** a diluição de citações tem o potencial de prejudicar as métricas dos periódicos científicos. Embora soluções como a unificação de citações sejam debatidas, elas enfrentam obstáculos técnicos e éticos, e bases como WoS e Scopus mantêm a contagem separada. A adoção de medidas para garantir o vínculo e a atualização das referências é fundamental para mitigar o problema.

**Palavras-chave:** preprints; diluição; dispersão; perda de citações

## Abstract

**Introduction:** citation dilution occurs when preprints continue to be cited even after the publication of their corresponding peer-reviewed articles. Although preprints are assigned a DOI and thus become part of the scholarly record, their citation poses challenges for editors, as citations may concentrate on the preliminary version, potentially affecting the metrics of both the article and the journal. **Methodology:** this descriptive-exploratory study, with a quantitative approach, aimed to analyze citation dilution in medRxiv preprints deposited between 2019 and 2021, whose corresponding articles were published by 2022. The analyzed corpus comprised 18,596 preprints. Data collection was conducted using OpenAlex, and validation via Crossref enabled the classification of citations as “diluted” (i.e., citations exclusively to the preprint after the article’s publication). An original indicator, termed Potential Citation Increase (PCI), was proposed to quantify the impact of dilution. PCI estimates the percentage increase in citations that an article or journal could have achieved had dilution not occurred. **Results:** of the analyzed preprints, 9,412 (51%) received diluted citations, totaling 58,737, which represents 7% of the total citations received by the corresponding articles. The peak of diluted citations occurs in the first month after article publication, and dilution ceases on average after 8 months. The duration of citation dilution ranged from 0 to 51 months following the publication of peer-reviewed articles, with most cases concentrated between 1 and 12 months (70%), followed by 13 to 24 months (21%), and a smaller proportion between 25 and 51 months (9%). The average PCI rate for the 9,330 analyzed articles was 15%. The median PCI at the journal level was 10%, indicating that 50% of journals could have experienced up to a 10% increase in citations from articles originating from preprints. **Conclusion:** citation dilution has the potential to negatively affect journal metrics. Although solutions such as citation unification have been discussed, they face technical and ethical challenges, and databases such as WoS and Scopus continue to maintain separate citation counts. The adoption of measures to ensure proper linking and updating of references is essential to mitigate this issue.

**Keywords:** preprints; dilution; dispersion; loss of citations

## 1 Introdução

---

*Preprints* são manuscritos divulgados ao público antes de passarem pela tradicional revisão por pares. De acordo com Cobb (2017), essa prática teve início na década de 1960 com a criação dos Grupos de Troca de Informações e tem, entre seus objetivos, acelerar a disseminação de resultados de pesquisa e obter considerações de outros especialistas antes da publicação formal do manuscrito em um periódico.

Atualmente, os *preprints* recebem um identificador persistente, geralmente um DOI (Digital Object Identifier), o que os torna parte permanente do registro acadêmico, facilitando sua descoberta e citação (Ettinger *et al.* 2022). Nesse sentido, tem-se observado um crescimento tanto na publicação quanto na citação de *preprints* nos últimos anos. De acordo com a base de dados Dimensions®<sup>(1)</sup>, o número de citações a *preprints* passou de 213 em 1990 para mais de 690 mil em 2024, representando um crescimento superior a 3.000%. O total acumulado de citações a esses

---

RODRIGUES, Pedro Henrique da Silva; WENDT, Lucas George; CAREGNATO, Sônia Elisa. Avaliação da Diluição de Citações entre *Preprints* e Artigos da Área da Saúde. *Brazilian Journal of Information Science: research trends*, vol.20, publicação contínua, 2026, e026014. DOI: <https://doi.org/10.36311/1981-1640.2026.v20.e026014>.

documentos já ultrapassa 4 milhões conforme contabilizado por essa base. Nessa direção, o crescimento no uso de *preprints* como fonte de informação tem despertado a necessidade de estudos que avaliem o impacto desse tipo de publicação na comunicação científica.

Pesquisas como a de Moshontz *et al.* (2021) destacam que *preprints* e seus artigos correspondentes podem acumular citações de forma independente. Os autores observam ainda que, em algumas bases de dados, como o Google Scholar, ambas as versões podem ser associadas ao perfil dos autores, permitindo a unificação das citações. No entanto, não se sabem as consequências dessa diluição para as revistas científicas que publicam os trabalhos anteriormente veiculados como *preprints*.

As citações a *preprints* não se restringem ao período anterior à publicação dos artigos correspondentes. Gehanno *et al.* (2022) analisaram 205 artigos de periódicos médicos, dos quais 60 (29,3%) citaram ao menos um *preprint*. Os autores identificaram que, dos 182 *preprints* citados, mais de 40% já haviam sido publicados como artigos em periódicos revisados por pares. Esse processo pelo qual as citações se fragmentam entre distintas versões ou etapas de um mesmo trabalho científico tem sido denominado como “diluição de citações” por autores como Avissar-Whiting (2022) e Smart (2022). A diluição acontece quando um estudo possui múltiplas versões e as citações a ele acabam sendo distribuídas entre essas versões e os diferentes canais de publicação nos quais aparece.

Para compreender melhor como os pesquisadores lidam com os *preprints*, Ni e Waltman (2024) realizaram uma pesquisa global (survey) com o objetivo de analisar a adoção, a familiaridade e as atitudes dos pesquisadores em relação aos *preprints* em diferentes disciplinas e regiões do mundo e constataram que muitos pesquisadores demonstram confusão sobre como citar *preprints* e como distingui-los das versões revisadas por pares. Além disso, os respondentes expressaram preocupação com a existência de múltiplas versões do mesmo artigo, o que pode gerar incerteza para os leitores e diluir as citações.

Nesse contexto, Chalepliglou e Koulouris (2023) reforçam que a citação a *preprints* representa um desafio relevante para os editores de periódicos, uma vez que há o risco de que as citações se concentrem no *preprint* em vez de no artigo final publicado. Gao, Wu e Zhu (2020)

levantam a mesma preocupação ao apontar que as citações direcionadas a um *preprint* podem afetar as métricas de citação do artigo correspondente publicado em periódico. De acordo com os autores, os *preprints* podem absorver parte das citações que, de outra forma, poderiam ter sido atribuídas ao artigo.

Para reduzir a diluição de citações, Gao, Wu e Zhu (2020) propõem unificar as citações de *preprints* e artigos publicados em um único registro. Contudo, essa solução tem limitações, já que os autores podem variar e os conteúdos podem diferir significativamente. Além disso, há questões éticas, como decidir se citações feitas ao *preprint* devem permanecer associadas a ele e ao repositório responsável por sua publicação. A Crossref, por sua vez, não recomenda mesclar citações, defendendo que cada documento tenha suas próprias métricas.

Além disso, a citação não reconhece apenas o conteúdo científico, mas também o trabalho editorial e de divulgação do periódico. Dessa forma, transferir automaticamente para o artigo final as citações que o *preprint* recebeu de forma independente poderia distorcer o sistema de recompensa acadêmica <sup>(2)</sup>. Acredita-se, ainda, que a fusão das citações inviabilizaria uma avaliação justa e transparente da contribuição de cada versão do trabalho em seus diferentes estágios. Nesse sentido, a Web of Science (WoS) e a Scopus, duas das principais bases de dados científicas do mundo, não mesclam ou transferem citações, contabilizando separadamente as citações ao *preprint* e ao artigo correspondente.

Nesse contexto, o presente artigo tem como objetivo analisar quantitativamente a diluição de citações em *preprints* do medRxiv, que foram posteriormente publicados como artigos entre 2019 e 2022. Para isso, buscou-se quantificar e classificar as citações direcionadas a esses *preprints*, observando sua distribuição temporal antes e depois da publicação dos artigos correspondentes. Adicionalmente, o estudo investigou a duração da diluição, determinando por quanto tempo os *preprints* continuaram a ser citados após a publicação. Por fim, avaliou-se o impacto dessa prática para os periódicos científicos, utilizando a métrica de Aumento Potencial de Citações (APCit), procedimento apresentado neste texto como possibilidade para cálculo do aumento potencial de citações na relação *preprint*/artigo revisado por pares.

---

RODRIGUES, Pedro Henrique da Silva; WENDT, Lucas George; CAREGNATO, Sônia Elisa. Avaliação da Diluição de Citações entre *Preprints* e Artigos da Área da Saúde. *Brazilian Journal of Information Science: research trends*, vol.20, publicação contínua, 2026, e026014. DOI: <https://doi.org/10.36311/1981-1640.2026.v20.e026014>.

## 2 Conceito e Evolução dos *Preprints*

---

De acordo com Rodriguez (2019), os *preprints* surgiram da confluência de quatro fatores: a necessidade de validação por pares antes do envio a periódicos, o avanço nos padrões de publicação desde o início da pesquisa até sua divulgação, o movimento de Acesso Aberto, que defende o livre acesso à informação científica, e a demanda por rapidez na publicação, seja para superar atrasos ou para estabelecer prioridade autoral.

Embora a prática remonte aos anos 1960, o primeiro servidor de *preprints*, como conhecemos hoje, foi o arXiv.org, lançado em 1991. Segundo Alvarez e Caregnato (2017), em áreas como a Física, a prática de divulgar resultados de pesquisa por meio de *preprints* teve início na década de 1960, com o propósito de agilizar a comunicação entre pesquisadores e promover o acesso aberto à produção científica. Atualmente, o arXiv hospeda cerca de 2,4 milhões de trabalhos acadêmicos em áreas como física, matemática, ciência da computação, biologia quantitativa, finanças quantitativas, estatística, engenharia elétrica, ciência de sistemas e economia. Assim, a *internet* teve um papel importante no desenvolvimento dos *preprints*, bem como dos próprios periódicos científicos. Conforme destacam Crespo e Caregnato (2006), a publicação científica periódica migrou para a *internet* como seu principal suporte de conteúdo, um desenvolvimento que, inicialmente, enfrentou baixa aceitação e estava restrito a um público limitado. Com o tempo, a redução de custos e o aprimoramento das Tecnologias da Informação e Comunicação (TICs) popularizaram o acesso e ampliaram a aceitação desse novo meio em diversas áreas científicas.

Nesse sentido, Puebla, Polka e Rieger (2022) conceituam os *preprints* como manuscritos acadêmicos disponibilizados em plataformas ou repositórios *online* com o intuito de promover o compartilhamento amplo e gratuito de pesquisas, sem barreiras de acesso. De acordo com os autores, esse tipo de documento costuma ser semelhante à versão submetida a periódicos científicos e pode ser publicado antes, durante ou independentemente do processo de submissão a uma revista. Geralmente é disponibilizado em um servidor de *preprints* poucos dias após o envio, sem avaliação por pares, permanecendo acessível de forma gratuita na *internet*. Além disso, esses servidores, em regra, não exigem a transferência dos direitos autorais, permitindo que o autor mantenha a propriedade intelectual e facilitando a reutilização do material. Nesse contexto, os

*preprints* se contrapõem ao modelo tradicional de acesso fechado, que frequentemente exige que o autor transfira os direitos autorais à editora ou ao periódico. Conforme Silva e Silveira (2019), o conteúdo em acesso aberto pode ser protegido por licenças *Creative Commons* ou *Science Commons*, as quais permitem que o autor ou a instituição ajustem os termos de uso conforme seu interesse, possibilitando o direito de utilizar, distribuir e remixar as informações.

Segundo Príncipe (2020), os servidores de *preprints* realizam uma avaliação rápida do texto submetido para verificar se o conteúdo possui caráter científico e se não há problemas éticos ou de plágio. Após essa checagem, o *preprint* é disponibilizado publicamente e pode receber comentários. Os autores, por sua vez, podem considerar essas contribuições e elaborar novas versões do documento original.

Quanto à função dos *preprints*, Rodríguez (2019) argumenta que sua principal função é tornar a informação publicamente disponível. Ao garantir o acesso público, gratuito e irrestrito aos resultados de pesquisa, os *preprints* contribuem para a redução de barreiras financeiras e para a democratização do acesso ao conhecimento (Sarabipour *et al.*, 2019; Puebla; Polka; Rieger, 2022). Além disso, os *preprints* desempenham um papel relevante na comunicação científica ao atuarem como documentos temporários, cuja função é suprir a lacuna temporal decorrente dos longos processos de publicação em periódicos. Assim, podem favorecer o acesso mais rápido a novos resultados e possibilitam a obtenção de *feedback* de diferentes pesquisadores (Larivière *et al.*, 2014; Flanagin *et al.*, 2020).

Nesse contexto, a disponibilização de *preprints* em acesso aberto, aliada à manutenção dos direitos autorais pelos próprios autores (e não pelas plataformas), cumpre uma função social associada à democratização da informação, à redução de barreiras de acesso e à maior rapidez na circulação do conhecimento científico. Por sua vez, a função científica dos *preprints* está relacionada à aceleração da comunicação, ao *feedback* aberto e ao aumento da transparência no processo de produção do conhecimento.

### 3 Preprints, DOI e Citações

---

Damasio (2023) ressalta que o uso de DOI em *preprints* é necessário, pois garante a vinculação da publicação ao sistema tradicional após sua divulgação. Dessa forma, ainda que os DOI do *preprint* e respectivo artigo sejam distintos, eles mantêm uma relação entre si.

A partir de 2016, os *preprints* passaram a ser oficialmente reconhecidos pelo Crossref como uma categoria de registro acadêmico, podendo receber um DOI e ter suas citações registradas. Conforme observa Farley (2024), a permanência dos registros é um princípio essencial, motivo pelo qual todo *preprint* cadastrado com um DOI deve continuar acessível. Dessa forma, os *preprints* permanecem disponíveis online mesmo após sua publicação em revistas científicas.

Em casos excepcionais que exijam a retirada de um *preprint*, como situações de plágio ou má conduta, recomenda-se que o DOI direcione para uma página contendo informações básicas, como o título, os autores e uma justificativa sucinta para a remoção. Ressalta-se ainda que os metadados associados aos *preprints* não podem ser apagados dos registros vinculados ao DOI (Farley, 2024).

A atribuição de DOI aos *preprints* facilita o registro de citações. Fraser *et al.* (2020) observaram que *preprints* estão sendo citados, independentemente de sua publicação em revistas revisadas por pares. Pagliaro (2021) destaca que *preprints* aparecem com frequência em citações de artigos de revistas, livros, conferências e apresentações revisadas por pares. Bertin e Atanassova (2022) examinaram o volume de citações a *preprints* em diferentes bases de dados no período de 2004 a 2019, constatando um aumento a partir de 2011, em paralelo ao surgimento e expansão de repositórios de *preprints*.

Analisando a estrutura IMRaD (Introdução, Métodos, Resultados e Discussão) em artigos da revista PLoS ONE, Bertin e Atanassova (2022) verificaram a distribuição das citações a *preprints*. Inicialmente, poderia se esperar que *preprints*, por serem estudos não validados por pares, fossem mencionados principalmente nas seções de introdução ou discussão, onde se contextualiza o trabalho ou se comparam resultados. Contudo, os autores observaram que a maior parte das citações ocorre na seção de metodologia, sugerindo que pesquisadores utilizam *preprints* para embasar a construção de novos conhecimentos, mesmo antes de sua validação formal por

pares. É importante notar que esses resultados se referem a periódicos da PLoS, focados em saúde, e podem não se aplicar a outras áreas do conhecimento.

A prática de citar *preprints* é influenciada pela cultura científica de cada disciplina (Kim, 2010). Áreas como física, matemática e ciência da computação apresentam maior tradição e aceitação de *preprints*, especialmente após o lançamento do repositório arXiv em 1991. Pickler (2019) aponta que a rápida disseminação de resultados nessas áreas é facilitada pela agilidade de publicação dos *preprints*.

Noruzi (2016) relata que, entre 1991 e 2016, *preprints* do arXiv receberam 135.782 citações, com 23.288 delas apenas em 2016. No entanto, Wang, Glänzel e Chen (2018) indicam que a frequência de citações a *preprints* diminuiu significativamente após sua publicação em revistas. Larivière *et al.* (2014) compararam as taxas de citação de *preprints* do arXiv, publicados ou não em revistas, com artigos revisados por pares indexados na Web of Science (WoS). Os resultados mostraram que *preprints*, independentemente de sua publicação em revistas, apresentam taxas de citação inferiores. Surpreendentemente, os *preprints* publicados em revistas não apresentaram taxas de citação significativamente superiores aos não publicados, contrariando a expectativa de maior impacto para estudos revisados por pares (Larivière *et al.*, 2014).

Sobre a confiabilidade dos *preprints*, Gehanno *et al.* (2022) compararam 124 *preprints* sobre Covid-19 do medRxiv com suas versões publicadas em revistas, identificando diferenças preponderantes em títulos, dados ou conclusões em quase metade dos casos. De forma semelhante, Bero *et al.* (2021) analisaram 67 *preprints* sobre Covid-19 e constataram discrepâncias nos resultados em 34% dos casos. Nesse contexto, o medRxiv alerta que *preprints* não devem ser usados como base para práticas clínicas ou divulgados como informações consolidadas (About [...], 2024). O uso de *disclaimers* acerca do tipo de publicação ao que correspondem os *preprints*, inclusive, é comum em repositórios.

Por outro lado, Klein *et al.* (2019) investigaram a similaridade entre *preprints* e suas versões publicadas, analisando 9.848 pares do arXiv e 1.443 do bioRxiv. Utilizando algoritmos de comparação textual, os autores verificaram que 95% dos textos do arXiv e 85% dos do bioRxiv apresentavam similaridade de 0,7 (em uma escala de 0 a 1), considerando o corpo do texto,

excluindo título, autores, resumo e referências. Já Akbaritabar *et al.* (2022) analisaram as mudanças ocorridas nos manuscritos durante o processo de revisão por pares, utilizando as alterações nas listas de referências entre a versão *preprint* e a versão final publicada como principal indicador. Foram mapeados 6.024 pares de *preprints* e suas respectivas publicações em revistas científicas. Eles classificaram as referências em três categorias: “adicionadas”, “removidas” ou “inalteradas”. Os resultados mostram que as listas de referências são estáveis. Cerca de 90% das referências presentes no *preprint* foram mantidas na versão publicada, enquanto aproximadamente 8% foram adicionadas e apenas 4% foram removidas.

Segundo os autores, as referências alteradas (adicionadas ou removidas) serviram principalmente para modificar a interpretação dos resultados ou o enquadramento teórico do trabalho. As ciências médicas e naturais foram as que apresentaram maior grau de transformação nesse aspecto. Portanto, quando um pesquisador cita um *preprint*, corre o risco de embasar sua pesquisa em um estudo com interpretações precipitadas, metodologias desatualizadas ou premissas teóricas que foram refutadas e corrigidas durante o processo de revisão por pares (correspondentes a cerca de 10% de referências modificadas).

Para compreender o que confere credibilidade aos *preprints*, Soderberg, Errington e Nosek (2020) entrevistaram 3.759 pesquisadores de diversas áreas, incluindo ciências sociais, biológicas, físicas, matemáticas e da saúde. Os resultados indicaram que transparência, acesso aberto a dados e verificação independente das alegações são mais valorizados do que a identidade dos autores, a revisão por pares ou a frequência de uso do *preprint*. A maioria dos entrevistados (69,73%) demonstrou apoio aos *preprints*, enquanto 15,16% foram contrários e 14,95% neutros. Quanto ao uso, 72,92% dos pesquisadores visualizaram ou baixaram *preprints* pelo menos ocasionalmente, sendo a visualização (70,63%) mais comum que o envio (29,85%) (Soderberg; Errington; Nosek, 2020).

George *et al.* (2021) destacam que, com o aumento do número de *preprints*, é necessário implementar mecanismos para garantir a qualidade e confiabilidade das pesquisas. Além disso, a citação de *preprints* exige cuidado para assegurar a validade das descobertas referenciadas. Assim,

embora a citação de *preprints* seja uma prática consolidada em algumas áreas, sua aceitação e confiabilidade variam, exigindo uso crítico e contextualizado.

Embora este estudo não se aprofunde nas motivações específicas para citar *preprints*, é importante observar que as razões para referenciá-los mesmo após a publicação de artigos revisados podem incluir: referências desatualizadas, quando pesquisadores não revisam suas listas de citações apesar da versão final estar disponível (Fraser *et al.*, 2020); a reutilização de citações, em que autores copiam referências de trabalhos prévios sem verificação (Davis, 2018); a falta de acesso, quando o autor não consegue obter o artigo correspondente, especialmente em periódicos pagos (Fraser *et al.*, 2020); a ausência de vínculo claro entre o *preprint* e o artigo final (Avisar-Whiting, 2022); e a escolha pessoal, na qual os autores preferem o *preprint* por discordarem das alterações da revisão por pares ou por informações omitidas na versão publicada (Avisar-Whiting, 2022).

## 4 Procedimentos Metodológicos

---

Trata-se de uma pesquisa de caráter descritivo-exploratório, com abordagem quantitativa, que, segundo Richardson (1999), vale-se de técnicas estatísticas que vão desde as mais básicas, como cálculo de percentuais, médias e desvios-padrão, até procedimentos mais sofisticados, como coeficientes de correlação e análises de regressão. A pesquisa exploratória tem como objetivo reunir informações sobre um objeto, delimitando o campo de estudo e identificando as condições em que ele se manifesta (Severino, 2013). Já o caráter descritivo se dá pela apresentação de características de uma população ou fenômeno determinado (Gil, 2002).

Nesta pesquisa, buscou-se quantificar as citações direcionadas aos *preprints* após seus artigos correspondentes serem publicados em periódicos revisados por pares. Para isso, foi construído um *corpus* composto por pares de *preprints* e seus respectivos artigos, e coletadas as citações recebidas pelos *preprints* e artigos, considerando as datas de publicação de ambos os documentos e de cada citante. Escolheu-se como *corpus* todos os *preprints* publicados no repositório medRxiv entre 2019 e 2021, cujos artigos correspondentes foram publicados em periódicos científicos entre 2019 e 2022.

O medRxiv foi escolhido porque seus *preprints* são facilmente identificáveis, citáveis e indexados, o que aumenta a probabilidade de serem encontrados e citados. Além disso, estudos anteriores como o de Gehanno *et al.* (2022) já haviam indicado a ocorrência de diluição de citações no medRxiv. Já a delimitação temporal da publicação dos trabalhos, 2019 e 2021 para *preprints* e 2019 e 2022 para artigos, se deu por duas razões: primeiro para garantir que os *preprints* tivessem tempo adequado para serem publicados como artigos em periódicos revisados por pares, ou seja, *preprints* depositados entre 2019 e 2021 tiveram a oportunidade de serem publicados como artigos até 2022. A segunda razão foi garantir um período de tempo suficiente para que os trabalhos pudessem ser lidos e citados. Artigos publicados até 2022 tiveram um período de citação que se estendeu até 31 de dezembro de 2023. Essa escolha é justificada por estudos na área da bibliometria, como o de Glänzel, Schlemmer e Thijs (2003), que constataram que a maioria das citações (76%) ocorre nos três primeiros anos após a publicação de um artigo, e aproximadamente 80% na área biomédica dentro desse mesmo período de três anos. Artigos não citados nesses primeiros anos raramente são citados posteriormente, segundo os autores.

Além dos *preprints* e seus respectivos artigos, foram coletados os seguintes metadados: os identificadores DOI e ID OpenAlex dos artigos publicados e dos *preprints*; o nome, o ISSN-L e o índice h dos periódicos onde os artigos foram publicados; a data de publicação de cada *preprint* e artigo; e o DOI e a data de publicação de cada citante do *preprint* e do artigo.

Para a coleta dos metadados foi utilizado o OpenAlex, uma base de dados abertos lançada em 2022 que reúne mais de 240 milhões de trabalhos acadêmicos de diversas fontes, como Crossref e repositórios de *preprints*. A escolha pelo OpenAlex deveu-se à sua capacidade de disponibilizar, por meio de API, o número total de citações e a lista completa dos citantes com seus respectivos metadados (DOI, título, data, autoria, entre outros), um recurso que não é oferecido gratuitamente por outras bases. Além disso, utilizou-se a Crossref para uma etapa de validação e classificação das citações. As etapas da pesquisa são descritas a seguir, apresentadas como I, II, III, IV, V e VI:

### **I. Coleta Inicial de *preprints* do medRxiv**

Esta etapa foi realizada de forma automatizada, com objetivo de coletar os metadados de todos os *preprints* disponíveis na plataforma medRxiv desde o seu lançamento em 2019, utilizando

um script em Python (Script 1) <sup>(3)</sup>. O script estabeleceu conexão com a API (Application Programming Interface) do medRxiv e coletou os seguintes metadados: DOI do *preprint*, título, autores, data de publicação, e o DOI do artigo correspondente (quando já publicado). A coleta foi realizada em março de 2024 e todas as informações foram armazenadas em uma planilha eletrônica.

## II. Coleta de metadados adicionais via API OpenAlex

Os metadados adicionais foram obtidos em março de 2024 via API do OpenAlex (utilizando o Script 2), incluindo:

- IDs OpenAlex: para artigos e *preprints*, essenciais para rastrear citações com o filtro `cites:ID` da API.
- Nome do periódico, ISSN-L e Índice h: para caracterizar as revistas e analisar o impacto da diluição de citações.
- Data de Publicação: de cada artigo e *preprint*, para estabelecer o marco temporal na identificação de citações diluídas.

## III. Seleção do corpus de pesquisa

Para a constituição do corpus da pesquisa, aplicando critérios específicos aos dados coletados anteriormente, a seleção seguiu dois critérios: a). Status de publicação: foram considerados apenas os *preprints* que já haviam sido publicados como artigos em periódicos; e b). Período de Análise: foram selecionados *preprints* disponíveis no medRxiv entre 2019 e 2021, cujos artigos correspondentes foram publicados entre 2019 e 2022.

## IV: Coleta das citações a *preprints*

Com o objetivo de coletar informações sobre as citações recebidas pelos *preprints* após a publicação dos artigos correspondentes, bem como o número total de citações recebidas pelos artigos correspondentes, *corpus* foi organizado em uma planilha Excel, na qual cada linha representava um par *preprint*-artigo para vincular os metadados de forma lógica. Utilizou-se o Script 3 para conectar à API OpenAlex e coletar as citações a *preprints*. Foram consideradas

---

RODRIGUES, Pedro Henrique da Silva; WENDT, Lucas George; CAREGNATO, Sônia Elisa. Avaliação da Diluição de Citações entre *Preprints* e Artigos da Área da Saúde. *Brazilian Journal of Information Science: research trends*, vol.20, publicação contínua, 2026, e026014. DOI: <https://doi.org/10.36311/1981-1640.2026.v20.e026014>.

apenas as citações a *preprints* que ocorreram após a data de publicação do artigo correspondente, com as datas comparadas individualmente para cada par. As consultas empregaram o parâmetro *filter=cites* para obter a lista completa dos citantes do *preprint*, incluindo suas datas de publicação e DOI. Além disso, foi coletado o número total de citações recebidas por cada artigo, juntamente com as datas de publicação dos seus citantes. Essa coleta foi realizada em abril de 2024.

## V. Validação das citações utilizando a API da Crossref

A coleta de dados foi realizada consultando os metadados dos artigos citantes via API Crossref, em abril de 2024. O objetivo foi verificar as referências dos trabalhos citantes para confirmar se o *preprint*, o artigo correspondente ou ambos foram citados, garantindo a qualidade e consistência dos dados de citação, além de classificar as citações em três tipos. Utilizou-se o Script 4 para pesquisar o DOI do *preprint* e do artigo correspondente nas referências bibliográficas do citante, conforme disponibilizado nos metadados da Crossref. As citações foram então classificadas em três categorias distintas:

- Citações duplicadas: quando tanto o *preprint* quanto o artigo final são citados no mesmo trabalho.
- Citações diluídas: refere-se à situação em que apenas o *preprint* é citado após a publicação do artigo correspondente.
- Citações não correspondentes: ocorre quando o OpenAlex indica que o trabalho citou o *preprint*, mas na validação não foi possível localizar o *preprint* na lista de referências do citante. Essas citações foram desconsideradas nas análises deste estudo.

Nessa etapa de validação com dados da Crossref, foram excluídas 53.767 citações classificadas como não correspondentes, o que representa 45,42% das 118.385 citações a *preprints* inicialmente identificadas pelo OpenAlex. Como mencionado anteriormente, considerou-se como “não correspondente” o caso em que a relação de citação indicada pelo OpenAlex não pôde ser confirmada na lista de referências do documento citante nos metadados da Crossref, acessados via API. Por exemplo, quando o OpenAlex indicava que o documento A citava o documento B,

verificava-se se B aparecia na lista de referências de A na Crossref. Se B não estivesse presente, a citação era classificada como não correspondente.

A alta proporção de não correspondências se explica, principalmente, pela ausência ou pelo depósito incompleto das listas de referências na Crossref por parte dos periódicos ou pelo não depósito do DOI na base da Crossref. Por exemplo, mesmo que o OpenAlex indique que A citou B, se os metadados de A na Crossref não incluírem sua lista de referências, não foi possível confirmar essa citação, o que leva à exclusão do par A-B do conjunto analisado. Cabe destacar que esse estudo considerou apenas documentos com DOI. Assim, todos os itens analisados possuem DOI e, em princípio, deveriam estar registrados na Crossref. Essa base é a principal agência de registro de DOI no mundo e responde por cerca de 95% resoluções de DOI (cliques em um link de DOI) <sup>(3)</sup>.

Após esse filtro, permaneceram 58.737 citações válidas para a análise. Embora essa exclusão represente um impacto percentual expressivo de perda de dados e sugira que o fenômeno da diluição possa ser ainda maior, ela garante a precisão e a confiabilidade do *corpus* final analisado, garantindo que todas as citações classificadas como diluídas foram de fato direcionadas exclusivamente ao *preprint*. Nesse sentido, com o processo de validação não restam dúvidas de que as citações aqui classificadas como diluídas referem-se apenas ao *preprint* após a publicação do respectivo artigo, com ausência de menção conjunta do artigo.

A Crossref foi escolhida porque, além de ser a principal agência de DOI do mundo, seus metadados frequentemente incluem as referências bibliográficas, permitindo identificar automaticamente (por meio do Script 4) se o *preprint* e/ou o artigo constam nas referências dos citantes.

## VI Análise dos Dados

Foi feita a compilação e análise estatística dos dados coletados e validados, para avaliar o fenômeno da diluição de citações e seu impacto. Esta etapa foi subdividida em diversas análises específicas:

---

RODRIGUES, Pedro Henrique da Silva; WENDT, Lucas George; CAREGNATO, Sônia Elisa. Avaliação da Diluição de Citações entre *Preprints* e Artigos da Área da Saúde. *Brazilian Journal of Information Science: research trends*, vol.20, publicação contínua, 2026, e026014. DOI: <https://doi.org/10.36311/1981-1640.2026.v20.e026014>.

a) Análise das citações diluídas: buscou detalhar a ocorrência de citações diluídas por *preprint* (média, mediana, moda, desvio padrão) e comparar as citações diluídas com as citações aos artigos por ano de publicação, analisando a diluição em função do intervalo de tempo entre *preprint* e artigo.

b) Análise do número de citações a *Preprints* antes e depois da publicação dos artigos correspondentes: identificar a dinâmica temporal das citações a *preprints*, quantificando e calculando a variação percentual na quantidade de citações antes e após a publicação dos artigos.

c) Duração máxima e média da diluição de citações: calcular por quanto tempo os *preprints* continuaram sendo citados no lugar dos artigos correspondentes (duração máxima e média, em meses), utilizando o Script 5

d) Análise da diluição de citações por revista: investigar a distribuição das citações diluídas entre as revistas, identificando as que tiveram maior número absoluto e maiores médias de citações diluídas.

e) Aumento potencial de citações (APC<sub>it</sub>): a taxa APC<sub>it</sub> é um indicador desenvolvido para avaliar o impacto da diluição de citações em artigos e periódicos. Ele busca responder à seguinte pergunta: qual seria o aumento percentual no número de citações que um artigo originado de *preprint*, ou uma revista, poderia ter alcançado caso não houvesse a diluição? É importante ressaltar que o APC<sub>it</sub> é uma estimativa e não uma previsão concreta. Não há garantia de que todos os autores que citaram o *preprint* após a publicação do artigo citariam a versão final publicada. Seu principal propósito é fornecer uma perspectiva quantitativa sobre a diluição de citações, funcionando como um indicador do impacto da diluição para artigos e periódicos.

A fórmula geral do APC<sub>it</sub> é a seguinte:  $APC_{it} (\%) = (xd / ya) \times 100$ , onde:

- $xd$  = número de citações diluídas (ou seja, citações feitas exclusivamente ao *preprint* após a publicação do artigo correspondente).
- $ya$  = número de citações ao artigo (referindo-se às citações que o artigo publicado de fato recebeu).

Uma limitação dessa fórmula é que ela não admite divisão por zero. Portanto, casos em que *preprints* receberam citações diluídas (numerador), mas os artigos ou a revista não receberam citações (denominador), são tratados separadamente.

A aplicação da fórmula APCit foi estendida às revistas, considerando duas abordagens distintas para avaliar o impacto da diluição:

a) APCit-R (Aumento Potencial Relativo de Citações): esta variação mede o impacto potencial das citações diluídas sobre o total de citações recebidas apenas pelos artigos que se originaram de *preprints* e foram publicados em uma determinada revista. Em outras palavras, ela verifica a proporção dessas citações diluídas dentro desse conjunto específico de artigos. É a razão entre a soma das citações diluídas dos *preprints* e a soma das citações aos artigos originados desses *preprints* publicados na revista analisada.

b) APCit-G (Aumento Potencial Geral de Citações): esta variação avalia o impacto das citações diluídas em relação ao total de citações recebidas por todos os artigos publicados pela revista durante o período de diluição. Ela ajuda a compreender como a diluição afetou o contexto geral de citações do periódico. A soma das citações diluídas dos *preprints* de uma revista é dividida pelo total de citações recebidas por todos os artigos (incluindo aqueles que não se originaram de *preprints*) publicados por essa revista durante o período em que as citações diluídas ocorreram (período de diluição). O "período de diluição" refere-se aos anos em que os *preprints* foram citados após a publicação do artigo, e pode variar entre as revistas.

c) Análise de citações duplicadas: analisar os casos em que um mesmo trabalho cita tanto o *preprint* quanto o artigo correspondente.

Por fim, cabe mencionar que no que se refere à identificação dos trabalhos, não foram encontradas duplicatas entre os *preprints* e artigos finais, visto que o DOI é um identificador único e exclusivo para cada documento. Foram detectados apenas vinte e sete (27) casos de DOIs inválidos, os quais foram excluídos do *corpus*. Em relação aos *preprints* com múltiplas versões, cabe ressaltar que o repositório medRxiv mantém o mesmo DOI para todas as atualizações de um manuscrito (ex.: versão 1, versão 2, etc.). Para os fins desta pesquisa, considerou-se sempre a data

da primeira versão, momento a partir do qual o *preprint* já se encontra público e apto a receber citações.

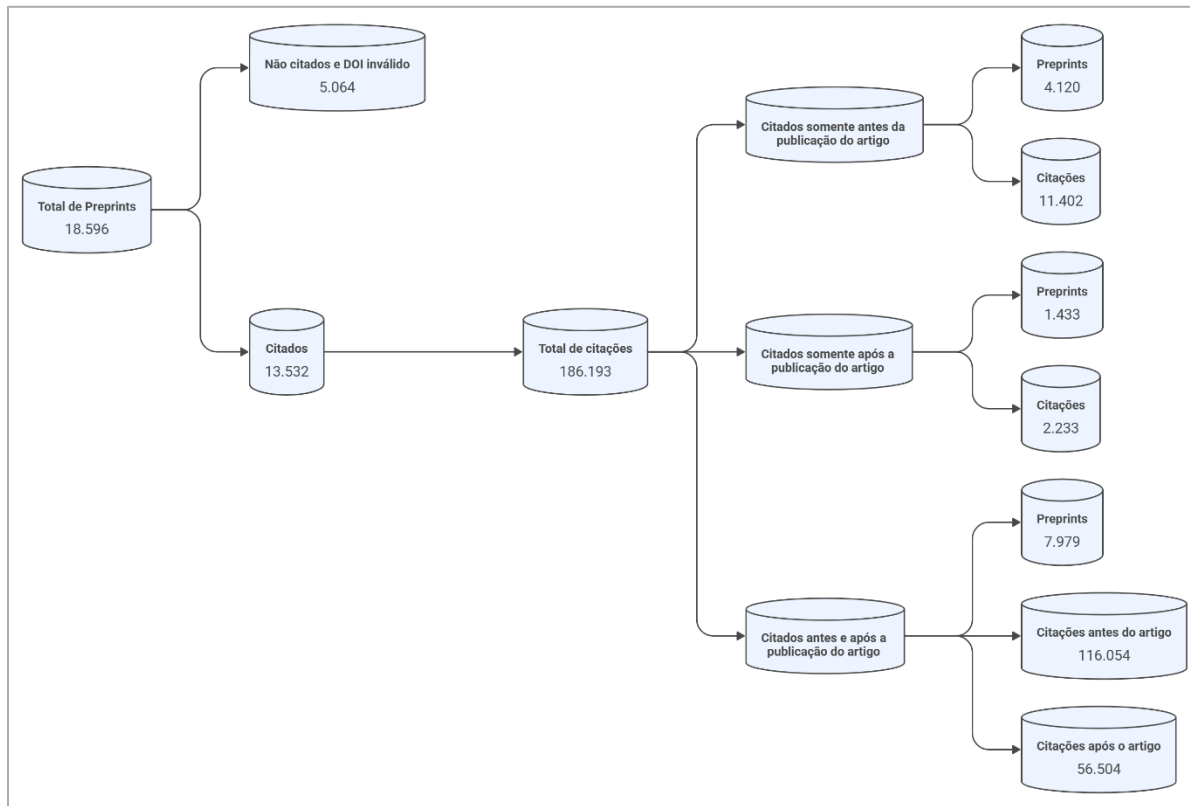
O índice *h* dos periódicos foi coletado via API do OpenAlex, a partir do identificador da revista localizado no campo "*source*": {"*id*": "[...]"} nos metadados de cada artigo. Identificaram-se 56 registros com ausência ou falha na coleta de metadados, onde 56 revistas não retornaram o ISSN-L ou o índice *h* de forma automatizada. Nesse cenário, a coleta e o preenchimento dessas informações foram realizados manualmente.

## 5 Resultados e Discussões

---

O estudo analisou um *corpus* de 18.596 *preprints* depositados no medRxiv entre 2019 e 2021. Os resultados mostraram que 9.412 *preprints* (51%) foram citados ao menos uma vez após a publicação dos artigos correspondentes (citações diluídas), as quais foram distribuídas por 2.003 revistas. No total, foram contabilizadas 58.737 citações diluídas. Este número equivale a 7% do total de citações (n=839.532) recebidas pelos artigos correspondentes.

A Figura 1 apresenta um diagrama hierárquico da distribuição dos *preprints* e de suas citações, detalhando a proporção de *preprints* citados e não citados, bem como a temporalidade das citações (antes, após e antes e após a publicação dos artigos correspondentes).

Figura 1 – Diagrama hierárquico da distribuição dos *preprints* e de suas citações

Fonte: os autores (2026)

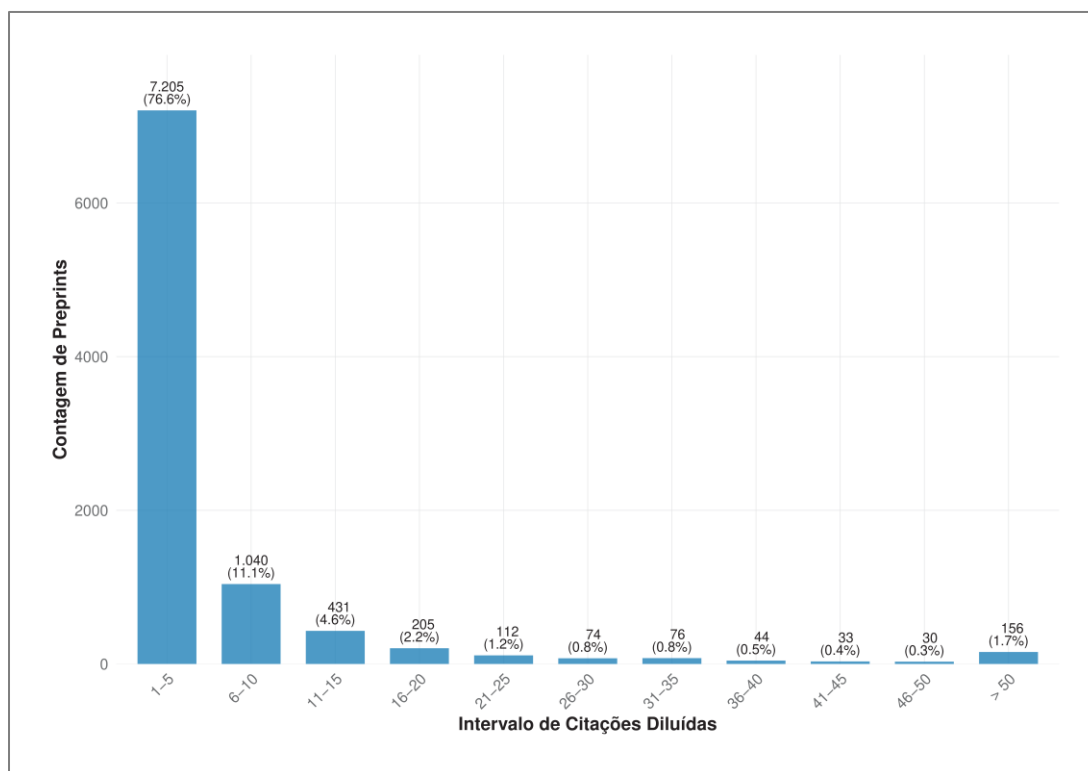
Os dados apresentados na figura 1 mostram que a maioria dos *preprints* foi citada ao menos uma vez (13.532 de 18.596), enquanto uma parte permaneceu sem citações ou apresentou DOI inválido (5.064). No conjunto dos *preprints* citados, observa-se uma distribuição assimétrica das citações ao longo do tempo. A maior parte dos *preprints* (7.979) foi citada tanto antes quanto após a publicação dos artigos correspondentes, concentrando também o maior volume de citações, especialmente no período anterior à publicação (116.054 citações, contra 56.504 posteriores).

### 5.1 Citações Diluídas: distribuição

A média de citações diluídas foi de 6, enquanto a mediana e a moda foram de 2 e 1, respectivamente. Isso mostra que a distribuição não é homogênea. A grande maioria dos *preprints* diluiu poucas citações, enquanto uma pequena parcela concentra um número mais elevado de citações diluídas. A distribuição heterogênea na quantidade de citações é esperada e compatível

com o que foi descrito por Price (1965), que afirma que a distribuição de citações costuma ser desigual. Segundo o autor, a maior parte dos artigos recebe poucas ou nenhuma citação, enquanto uma pequena parcela concentra um número elevado de citações. O comportamento se repete quando se trata das citações diluídas, pois, conforme observa-se no Gráfico 1, muitos *preprints* apresentaram poucas citações diluídas, enquanto poucos tiveram uma quantidade mais significativa.

Gráfico 1 – Distribuição de *Preprints* por Intervalo de Citações Diluídas



Fonte: dados da pesquisa (2025)

A Tabela 1 apresenta uma comparação entre a quantidade de citações diluídas e a quantidade de citações recebidas pelos artigos correspondentes, organizados de acordo com o ano de publicação de cada par *preprint*-artigo. Dessa forma, os resultados são apresentados segundo o intervalo de tempo entre a disponibilização do *preprint* e a publicação do artigo, o que permite examinar de que maneira a diluição de citações se manifestou em função do tempo decorrido entre ambas as versões.

Tabela 1 – Citações diluídas e citações a artigos segundo intervalo de tempo entre versões

Ano <i>preprint</i>	Ano artigo	Intervalo	Pares <i>preprints</i> -artigos	Citações diluídas	Média (diluição)	Mediana (diluição)	Moda (diluição)	Citações a artigos
2019	2019	<1 ano	27	58	2,15	1	1	891
2019	2020	1 ano	159	316	1,99	1	1	8.313
2019	2021	2 anos	26	47	1,74	1	1	1.099
2019	2022	3 anos	4	11	2,75	2,5	2	81
2020	2020	<1 ano	2.781	30.975	11,14	4	1	501.444
2020	2021	1 ano	2.375	10.919	4,60	2	1	140.067
2020	2022	2 anos	237	659	2,78	2	1	6.415
2021	2021	<1 ano	2.210	10.826	4,90	2	1	129.392
2021	2022	1 ano	1.592	4.925	3,09	1	1	51.825
<b>Total</b>			<b>9.412</b>	<b>58.737</b>	<b>6,24</b>	<b>2</b>	<b>1</b>	<b>839.532</b>

Fonte: dados da pesquisa (2025)

A análise da Tabela 1 mostra que a maior parte dos artigos originados de *preprints* foi publicada em até um ano após a versão preliminar. O ano de 2020 concentrou o maior número de pares *preprint*-artigo, bem como os maiores índices de citações diluídas e de citações direcionadas aos artigos, o que pode estar relacionado aos impactos da pandemia de Covid-19.

Observa-se que tanto o número absoluto quanto a média de citações diluídas diminuem à medida que aumenta o intervalo entre o *preprint* e a publicação do artigo correspondente. Para ilustrar, a média de diluição registrada para pares publicados em 2020 foi de 11,14. Já os *preprints* de 2020 cujos artigos foram disponibilizados em 2021 apresentaram média de 4,60, enquanto aqueles com artigos publicados em 2022 registraram apenas 2,78. Tendência semelhante ocorre com os *preprints* de 2021, cuja média de diluição foi de 4,9, caindo para 3,09 quando os artigos correspondentes foram publicados em 2022.

O comportamento nos permite inferir que quanto maior o intervalo de tempo entre o *preprint* e o artigo, menor tende a ser a média de citações diluídas. A dinâmica pode ser explicada pela observação de Glänzel, Schlemmer e Thijs (2003), segundo a qual a maioria dos trabalhos acadêmicos, inclusive na área da saúde, recebe maior número de citações nos primeiros anos após a publicação, com tendência de declínio ao longo do tempo. Assim, quando o artigo demora mais

tempo para ser publicado, sua visibilidade e capacidade de atrair novas citações já se encontram reduzidas em função da existência do *preprint* e conseqüentemente terá menos citações diluídas.

## 5.2 Duração da diluição de citações

---

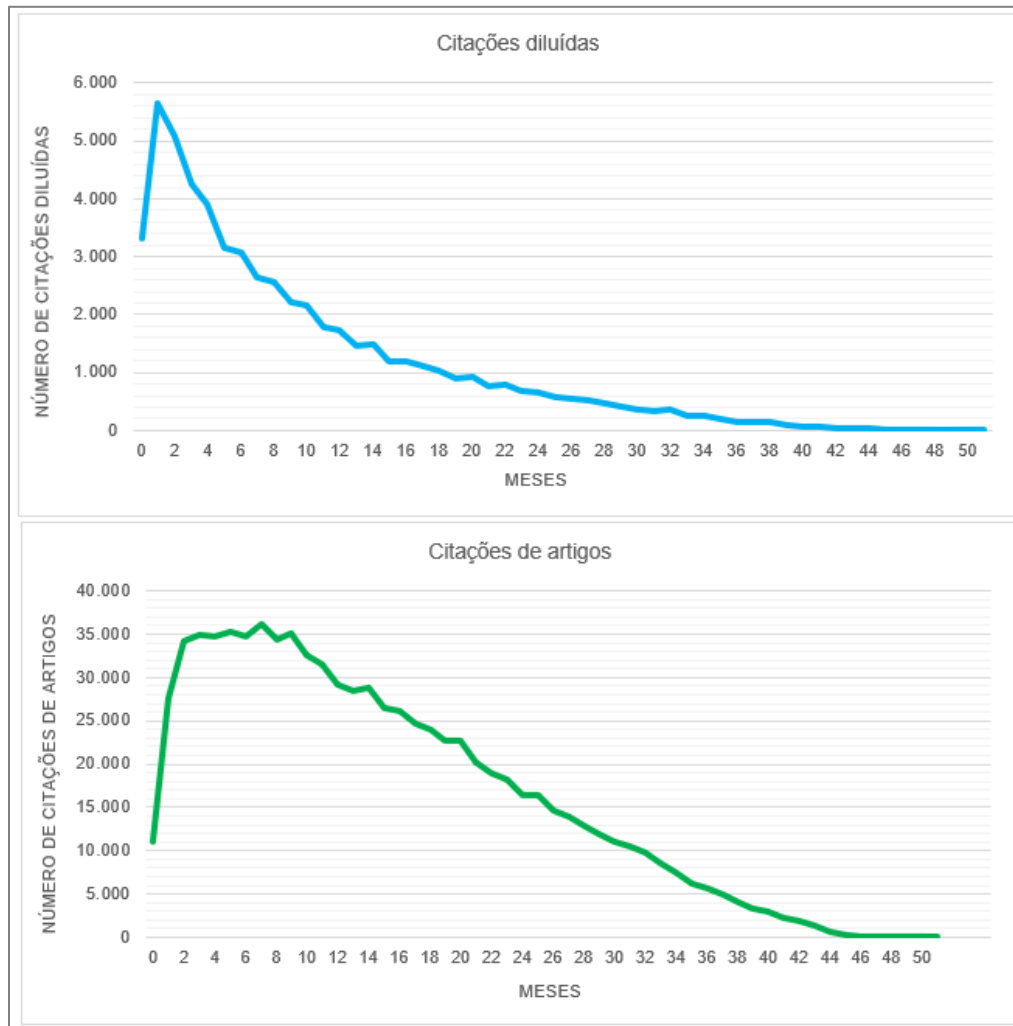
A Duração da Diluição de Citações é uma análise que buscou quantificar o tempo em que os *preprints* continuaram a ser citados após a publicação de seus respectivos artigos revisados por pares.

O tempo de diluição variou de 0 a 51 meses. A maioria das citações diluídas (70%) teve uma duração entre 1 e 12 meses. Cerca de 21% ficaram no intervalo de 13 a 24 meses. Após esse período, o número caiu, com os 9% restantes distribuídos entre 25 e 51 meses após a publicação da versão do texto revisada por pares.

Os dados corroboram o estudo de Fraser *et al.* (2020), que notou que poucos *preprints* são citados dois anos após a publicação do artigo final. As razões para que os *preprints* continuem sendo citados mesmo após um período considerável da publicação do artigo não são claras, mas podem incluir a dificuldade de acesso ao artigo pago, o desconhecimento da publicação da versão final ou a falta de atualização em programas de gerenciamento de referências, por exemplo.

Para avaliar se as citações aos artigos aumentam enquanto as citações aos *preprints* diminuem, também foi elaborado o histórico temporal das citações feitas aos artigos originados dos *preprints*. O cálculo seguiu o mesmo procedimento usado para as citações de *preprints*: primeiro, registrou-se a data de publicação de cada citante, depois subtraiu-se pela data de publicação do artigo citado. Assim, obteve-se a diferença, em meses, entre o artigo e os citantes. Por fim, analisou-se a frequência com que cada intervalo de meses apareceu nos resultados, conforme é mostrado no Gráfico 2.

Gráfico 2 – Distribuição temporal das citações diluídas e citações aos artigos correspondentes



Fonte: dados da pesquisa (2025)

O Gráfico 2 mostra que o pico de citações diluídas ocorre no primeiro mês, enquanto o pico de citações dos artigos acontece no sétimo mês. Nota-se também que, em 24 meses, as citações diluídas caíram 88% (de mais de 5 mil para cerca de 600). No mesmo período, as citações aos artigos diminuíram 50% (de 36 mil para aproximadamente 18 mil). Com isso, demonstra-se que, embora ambos percam citações com o tempo, a queda para os *preprints* é proporcionalmente maior, cerca de 38% mais acentuada que a dos artigos.

Em suma, os *preprints* tendem a parar de ser citados pouco tempo após a publicação do artigo correspondente. Esse período foi, em média, de 8 meses. Esses dados confirmam as

observações de Wang, Glänzel e Chen (2018), que apontaram para uma rápida diminuição nas citações de *preprints* depois que são publicados em revistas científicas. Pode-se afirmar que as citações diluídas acompanham o ciclo de vida das citações dos artigos, mas com um declínio mais rápido e intenso.

É compreensível que um *preprint* seja citado até cerca de seis meses após a publicação do artigo correspondente, pois há um tempo necessário para a indexação em bases de dados e a disseminação do trabalho final entre a comunidade científica. Durante esse período, o *preprint* pode ainda ser a principal fonte acessível ou conhecida pelos pesquisadores. No entanto, quando as citações ao *preprint* ocorrem dois ou três anos após a publicação do artigo revisado, isso pode indicar problemas como desatualização das referências, falta de acesso à versão final publicada ou até mesmo preferência deliberada pelo *preprint*.

### 5.3 Impacto da diluição nos artigos correspondentes (APCit)

---

Aqui foram considerados 9.330 pares, e não 9.412, porque o APCit não pôde ser calculado para 82 artigos (distribuídos em 79 revistas) que não receberam nenhuma citação. Embora seus *preprints* correspondentes tenham acumulado citações diluídas, esses casos foram excluídos, pois a fórmula do APCit, apresentada na metodologia, não admite denominador igual a zero. Esses 82 desconsiderados no cálculo do APCit totalizaram 133 citações diluídas, sendo que 51 deles (cerca de 62%) receberam apenas uma citação cada. Essa quantidade representa apenas 0,88% do total de pares analisados e, portanto, não afeta de forma significativa a interpretação geral das taxas de APCit.

Em média, os 9.330 artigos analisados poderiam ter obtido um aumento de aproximadamente 15% no número de citações, caso as citações diluídas fossem creditadas à versão final publicada. A Tabela 2 apresenta o resumo das estatísticas descritivas do APCit para os 9.330 artigos e demonstra a distribuição e dispersão dos dados.

Tabela 2 – Estatísticas descritivas do APCit dos 9.330 artigos analisados

Medida	Valor	Observações
Contagem (N)	9.330	Total de artigos analisados
Média	15,37%	APCit média
Moda	20,00%	APCit mais frequente
Desvio Padrão	29,22%	Alta dispersão nos dados
Mínimo	0,17%	APCit mais baixo observado
1º Quartil (Q1)	4,55%	25% dos dados $\leq$ 4,55%
Mediana (Q2)	8,70%	50% dos dados $\leq$ 8,70%
3º Quartil (Q3)	16,67%	75% dos dados $\leq$ 16,67%
Máximo	1.200%	APCit mais alto observado

Fonte: dados da pesquisa (2025)

Por meio da Tabela 2 observa-se que a taxa APCit mais frequente foi de 20%, presente em 245 trabalhos, o que representa 2,63% do total. O desvio padrão de 29,22% supera a média de 15,37%, indicando distribuição assimétrica, com valores variando de 0,17% a 1.200%, o que revela a presença de outliers.

Para obter uma compreensão mais clara da distribuição e atenuar o impacto de outliers, a análise foi aprofundada com o cálculo dos quartis. O primeiro quartil (Q1) foi de 4,55%, indicando que um quarto dos artigos apresentou uma taxa APCit igual ou inferior a esse valor. A mediana (Q2), que representa o ponto central dos dados, foi de 8,70%. Isso significa que metade dos artigos analisados teria um aumento de até 8,70% no número de citações caso não houvesse a diluição, isto é, se as citações tivessem sido atribuídas ao artigo em vez do *preprint*. Por fim, o terceiro quartil (Q3) de 16,67% mostrou que 75% dos artigos se encontravam abaixo desse valor, enquanto os 25% restantes concentravam as taxas mais elevadas.

#### 5.4 Impacto da diluição nas revistas

Além de calcular o APCit para os artigos, o mesmo foi aplicado às revistas que publicaram os 9.330 artigos, com o objetivo de avaliar o efeito de diluição das citações no nível dos periódicos. Conforme explicado na metodologia, o APCit das revistas foi estimado por duas abordagens: Aumento Potencial Relativo das Citações (APCit-R) e Aumento Potencial Geral das Citações (APCit-G). Entre as revistas analisadas, 1.951 (97,40%) tiveram ao menos um artigo do *corpus*

citado; por outro lado, 52 não receberam nenhuma citação, embora os *preprints* correspondentes tenham sido citados. Importa notar que, das 79 revistas que publicaram os 82 artigos não citados, apenas 52 permaneceram sem citação, porque a classificação de “revista citada” não exige que todos os seus artigos sejam citados. Assim, por exemplo, um periódico com onze artigos (dez citados e um não citado) é contabilizado entre as 1.951 revistas citadas. A Tabela 3 apresenta um resumo estatístico do APCit-R para as revistas analisadas.

Tabela 3 – Resumo das estatísticas descritivas do APCit-R das revistas analisadas

<b>Medida</b>	<b>Valor</b>	<b>Observações</b>
Contagem (N)	1.952	Total de artigos analisados
Média	20,15%	Aumento percentual médio
Moda	100,00%	Valor mais frequente
Desvio Padrão	33,74%	Alta dispersão nos dados
Mínimo	0,42%	APCit-R mais baixo observado
1º Quartil (Q1)	5,69%	25% dos dados $\leq$ 5,69%
Mediana (Q2)	10,00%	50% dos dados $\leq$ 10%
3º Quartil (Q3)	20,00%	75% dos dados $\leq$ 20%
Máximo	520%	APCit-R mais alto observado

Fonte: dados da pesquisa (2025)

A média do APCit-R entre as revistas analisadas foi de 20%, ou seja, em média as revistas teriam um aumento de 20% no número de citações dos artigos originados de *preprints* não fosse a diluição de citações. No entanto, esse valor está bem acima da mediana (segundo quartil), de 10%, o que sugere a presença de outliers que elevam a média, tornando-a menos representativa da tendência central dos dados.

Nesse sentido, a análise por quartis oferece uma análise mais representativa, reduzindo a influência de valores extremos. No primeiro quartil (25% das revistas com menor APCit-R), os periódicos poderiam registrar um aumento de até 5,69% nas citações de artigos originados de *preprints* do corpus caso não houvesse a diluição das citações. No segundo quartil, metade das revistas (50%) teria um aumento de até 10% no número de citações. Já no terceiro quartil, 75%

das revistas alcançariam um APCit-R de até 20%, sendo que aquelas situadas entre os percentis 50% e 75% apresentariam um acréscimo entre 10% e 20%.

Enquanto o APCit-R mede o impacto da diluição considerando apenas o subconjunto de artigos da revista originados de *preprints* do medRxiv (ou seja, o corpus desta pesquisa), o APCit-G mede o impacto da diluição em relação ao volume total de citações recebidas pela revista. Em outras palavras, o APCit-R foca no impacto das citações dentro do grupo específico de artigos derivados de *preprints*, enquanto o APCit-G avalia o efeito das citações diluídas no contexto geral da revista. Nesse sentido, a tabela 4 apresenta a distribuição do APCit-G das revistas, calculado com base no Aumento Potencial Geral das Citações (APCit-G). Nessa abordagem, o APCit foi obtido a partir do total de citações recebidas por todos os artigos publicados por cada revista durante o período de diluição, ou seja, todas as revistas possuem citações, pois são considerados tanto os artigos originados de *preprints* quanto os que não são.

Tabela 4 – Distribuição do APCit das revistas calculado com base na Aumento Potencial Geral das Citações (APCit-G)

Intervalo APCit (%)	Número de Revistas	Percentual de Revistas
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%
30 † 35	0	0,00%
35 † 40	2	0,10%
40 † 45	1	0,05%
(...)	0	0,00%
130 † 135	1	0,05%
<b>Total</b>	<b>2.003</b>	<b>100%</b>

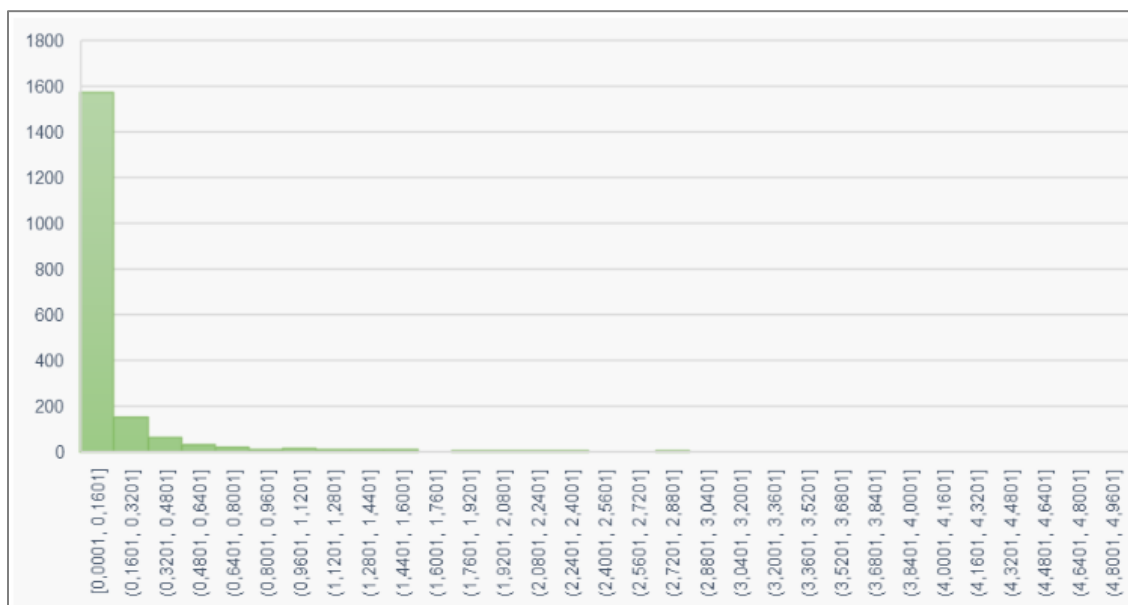
Fonte: elaborado pelo autor (2025)

**Nota:** a notação † representa um intervalo fechado-aberto, ou seja, inclui o valor inicial e exclui o valor final.

Com base na Tabela 4, verifica-se que 98% das revistas registram um APCit-G entre 0 e 4,9%, enquanto apenas 0,8% se situam na faixa de 5% a 9,9%. Diante deste cenário, optou-se por examinar de forma detalhada as revistas compreendidas no intervalo de 0% a 4,9%, buscando uma

análise mais aprofundada. O Gráfico 3 apresenta a distribuição das revistas que possuem um APCit-G nesse intervalo.

Gráfico 3 – Distribuição de revistas por faixa de APCit-G (0% a 5%)



Fonte: elaborado pelo autor (2025)

Como mostra o Gráfico 3, cerca de 79% das 2.003 revistas avaliadas registraram um aumento de até 0,16% no número total de citações ao longo do período de diluição. Apesar de esse efeito parecer pequeno, é importante considerar o contexto individual de cada revista, especialmente em relação ao seu porte, tanto em quantidade de publicações quanto de citações recebidas. Por exemplo, um acréscimo de 0,1% em uma revista com 1 milhão de citações representa 100 mil novas citações, enquanto o mesmo percentual em uma revista com apenas mil citações corresponde a apenas uma.

### 5.5 Citações duplicadas

Dos 18.596 pares de *preprints* e artigos analisados, 5,79% (1.076 pares) apresentaram citações duplicadas, totalizando 1.899 ocorrências. Esse número representa apenas 3% do total de citações diluídas, o que sugere que citar simultaneamente o *preprint* e o artigo não é uma prática comum. Embora o motivo exato seja desconhecido, as citações duplicadas podem ser explicadas

pela falta de normas claras para citar *preprints* ou pela necessidade de contextualizar diferenças entre a versão inicial e a final do trabalho. Além disso, essa prática tem o potencial de inflar artificialmente o número de citações de um autor, já que as contagens para as duas versões do mesmo estudo podem ser somadas em favor dos autores.

## 6 Considerações Finais

---

A prática de citar *preprints* em vez de artigos finais publicados pode levar à diluição de citações, prejudicando principalmente os periódicos científicos. Isso afeta métricas importantes para decisões sobre financiamento, indexação e sobrevivência das revistas, enquanto os autores podem receber citações em ambas as versões. Embora a vinculação entre *preprints* e artigos finais seja recomendada, ela não impede totalmente a diluição. Todos os artigos do *corpus* desta pesquisa estavam vinculados aos seus *preprints* na medRxiv, mas mesmo assim sofreram diluição de citações. Faz-se necessário investigar as variações na diluição entre publicações vinculadas e não vinculadas.

Citar *preprints* após a publicação do artigo correspondente vai além da simples perda de citações para os periódicos, levantando questões sobre a confiabilidade e a precisão das informações científicas. Segundo Gehanno *et al.* (2022), muitos *preprints* passam por mudanças significativas antes da publicação final. Isso indica que o conteúdo de um *preprint* pode diferir da versão revisada por pares, gerando incertezas ao ser citado. Silva (2018) reforça que citar *preprints* exige responsabilidade na interpretação do conteúdo, mas promover uma cultura de uso ético é desafiador, especialmente com o aumento de publicações predatórias.

Este estudo focou principalmente em periódicos de ciências da saúde, muitos sob influência da Covid-19, o que limita a generalização dos resultados para outras áreas, já que cada campo possui práticas específicas de publicação e citação. Para verificar se a diluição de citações ocorre de forma semelhante em outros contextos ou fora de emergências, mais pesquisas são necessárias.

Embora não exista uma solução definitiva para a diluição de citações, as revistas podem adotar medidas para minimizar o problema: orientar autores a atualizarem referências de *preprints*

antes da publicação; garantir a conexão entre *preprint* e artigo final nos repositórios; registrar corretamente os metadados em bases como a Crossref, facilitando o acesso; sugerir que *preprints* sejam citados como comunicação pessoal; e incentivar ferramentas como o Google Scholar a priorizarem o artigo publicado como fonte principal.

Por fim, a pesquisa sobre diluição de citações deve continuar. Uma abordagem útil seria analisar em quais partes dos artigos, como introdução, métodos ou resultados, os *preprints* são mais citados, para entender seu papel. Também é importante investigar se a diluição varia entre artigos de acesso aberto e fechado. Estudos futuros podem expandir a análise, incluindo mais dados e *preprints* de outras áreas do conhecimento.

## Agradecimentos

---

Agradecemos o financiamento da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Código de Financiamento 001 e do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), concedido na forma de bolsas.

## Notas

---

- (1) [https://archive.org/details/chart\\_202508](https://archive.org/details/chart_202508)
- (2) Conforme Silva *et al* (2024), o Sistema de Recompensa na Ciência (SRC) baseia-se na valorização da produtividade, citação e prestígio, recompensando pesquisadores com reconhecimento e recursos para dar continuidade às suas carreiras. Ele reflete uma evolução histórica do conceito de recompensa social aplicado ao contexto científico.
- (3) <https://web.archive.org/web/20260506192053/https://www.crossref.org/community/>

## Referências

---

- “About medRxiv.” *medRxiv*, 2024, [www.medrxiv.org/content/about-medrxiv](http://www.medrxiv.org/content/about-medrxiv).
- Akbaritabar, Aliakbar, et al. “A Study of Referencing Changes in Preprint-Publication Pairs across Multiple Fields”. *Journal of Informetrics*, v. 16, n. 2, maio de 2022, p. 101258. *DOI.org (Crossref)*, <https://doi.org/10.1016/j.joi.2022.101258>.

- Alvarez, Gonzalo Rubén, e Sônia Elisa Caregnato. “Preprints na comunicação científica da Física de Altas Energias: análise das submissões no repositório arXiv (2010-2015)”. *Perspectivas em Ciência da Informação*, v. 22, n. 2, junho de 2017, p. 104-17. DOI.org (Crossref), <https://doi.org/10.1590/1981-5344/2830>.
- Avissar-Whiting, Michele. “Downstream Retraction of Preprinted Research in the Life and Medical Sciences”. *PLOS ONE*, organizado por Frederique Lisacek, v. 17, n. 5, maio de 2022, p. e0267971. DOI.org (Crossref), <https://doi.org/10.1371/journal.pone.0267971>.
- Bero, Lisa, 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*, v. 11, n. 7, julho de 2021, p. e051821. DOI.org (Crossref), <https://doi.org/10.1136/bmjopen-2021-051821>.
- Bertin, Marc, e Iana Atanassova. “Preprint Citation Practice in PLOS”. *Scientometrics*, v. 127, n. 12, dezembro de 2022, p. 6895-912. DOI.org (Crossref), <https://doi.org/10.1007/s11192-022-04388-5>.
- Chaleplioglou, Artemis, e Alexandros Koulouris. “Preprint Paper Platforms in the Academic Scholarly Communication Environment”. *Journal of Librarianship and Information Science*, v. 55, n. 1, março de 2023, p. 43-56. DOI.org (Crossref), <https://doi.org/10.1177/09610006211058908>.
- Cobb, Matthew. “The Prehistory of Biology Preprints: A Forgotten Experiment from the 1960s”. *PLOS Biology*, v. 15, n. 11, novembro de 2017, p. e2003995. DOI.org (Crossref), <https://doi.org/10.1371/journal.pbio.2003995>.
- Crespo, Isabel Merlo, e Sônia Elisa Caregnato. “Padrões de comportamento de busca e uso de informação por pesquisadores de biologia molecular e biotecnologia”. *Ciência da Informação*, v. 35, n. 3, dezembro de 2006, p. 30-38. DOI.org (Crossref), <https://doi.org/10.1590/S0100-19652006000300003>.
- Damasio, Edilson. “Preprints na comunicação científica: uma introdução”. *BIBLOS*, v. 32, n. 2, janeiro de 2023, p. 155-68. DOI.org (Crossref), <https://doi.org/10.14295/biblos.v32i2.8635>.
- Davis, Phil. “Journals Lose Citations to Preprint Servers.” *The Scholarly Kitchen*, 2018. <https://web.archive.org/web/20241120183659/https://scholarlykitchen.sspnet.org/2018/05/21/journals-lose-citations-preprint-servers-repositories/>.
- Ettinger, Cassandra L., et al. “A Guide to Preprinting for Early-Career Researchers”. *Biology Open*, v. 11, n. 7, julho de 2022, p. bio059310. DOI.org (Crossref), <https://doi.org/10.1242/bio.059310>.
- Farley, Isaac. “Introduction to Posted Content (Includes Preprints).” *Crossref*, 2024. <https://web.archive.org/web/20241205153313/https://www.crossref.org/documentation/research-nexus/posted-content-includes-preprints/>.
- Flanagin, Annette, et al. “Preprints Involving Medical Research - Do the Benefits Outweigh the Challenges?” *JAMA*, v. 324, n. 18, novembro de 2020, p. 1840. DOI.org (Crossref), <https://doi.org/10.1001/jama.2020.20674>.

- Fraser, Nicholas, et al. “The Relationship between bioRxiv Preprints, Citations and Altmetrics”. *Quantitative Science Studies*, v. 1, n. 2, junho de 2020, p. 618-38. DOI.org (Crossref), [https://doi.org/10.1162/qss\\_a\\_00043](https://doi.org/10.1162/qss_a_00043).
- Gao, Yong, et al. “Merging the Citations Received by arXiv-Deposited e-Prints and Their Corresponding Published Journal Articles: Problems and Perspectives”. *Information Processing & Management*, v. 57, n. 5, setembro de 2020, p. 102267. DOI.org (Crossref), <https://doi.org/10.1016/j.ipm.2020.102267>.
- Gehanno, Jean-Francois, et al. “Reliability of Citations of medRxiv Preprints in Articles Published on COVID-19 in the World Leading Medical Journals”. *PLOS ONE*, organizado por Venkatesh Shankar Madhugiri, v. 17, n. 8, agosto de 2022, p. e0264661. DOI.org (Crossref), <https://doi.org/10.1371/journal.pone.0264661>.
- George, Christopher H., et al. “Editorial Policy Regarding the Citation of Preprints in the *British Journal of Pharmacology* (BJP)”. *British Journal of Pharmacology*, v. 178, n. 18, setembro de 2021, p. 3605-10. DOI.org (Crossref), <https://doi.org/10.1111/bph.15589>.
- Gil, Antonio Carlos. *Como Elaborar Projetos de Pesquisa*. 4th ed., Atlas, 2002.
- Glänzel, Wolfgang, et al. “Better Late than Never? On the Chance to Become Highly Cited Only beyond the Standard Bibliometric Time Horizon”. *Scientometrics*, v. 58, n. 3, novembro de 2003, p. 571-86. DOI.org (Crossref), <https://doi.org/10.1023/B:SCIE.0000006881.30700.ea>.
- Kim, Jihyun. “Faculty Self-archiving: Motivations and Barriers”. *Journal of the American Society for Information Science and Technology*, v. 61, n. 9, setembro de 2010, p. 1909-22. DOI.org (Crossref), <https://doi.org/10.1002/asi.21336>.
- Klein, Martin, et al. “Comparing Published Scientific Journal Articles to Their Pre-Print Versions”. *International Journal on Digital Libraries*, v. 20, n. 4, dezembro de 2019, p. 335-50. DOI.org (Crossref), <https://doi.org/10.1007/s00799-018-0234-1>.
- Larivière, Vincent, et al. “arXiv E-prints and the Journal of Record: An Analysis of Roles and Relationships”. *Journal of the Association for Information Science and Technology*, v. 65, n. 6, junho de 2014, p. 1157-69. DOI.org (Crossref), <https://doi.org/10.1002/asi.23044>.
- Moshontz, Hannah, et al. “A Guide to Posting and Managing Preprints”. *Advances in Methods and Practices in Psychological Science*, v. 4, n. 2, abril de 2021, p. 25152459211019948. DOI.org (Crossref), <https://doi.org/10.1177/25152459211019948>.
- Ni, Rong, e Ludo Waltman. “To Preprint or Not to Preprint: A Global Researcher Survey”. *Journal of the Association for Information Science and Technology*, v. 75, n. 6, junho de 2024, p. 749-66. DOI.org (Crossref), <https://doi.org/10.1002/asi.24880>.
- Noruzi, Alireza. “arXiv Popularity from a Citation Analysis Point of View.” *Webology*, vol. 13, no. 2, 2016. [www.webology.org/2016/v13n2/editorial22.pdf](http://www.webology.org/2016/v13n2/editorial22.pdf).

- Pagliaro, Mario. “Did You Ask for Citations? An Insight into Preprint Citations En Route to Open Science”. *Publications*, v. 9, n. 3, junho de 2021, p. 26. *DOI.org (Crossref)*, <https://doi.org/10.3390/publications9030026>.
- Pickler, Rita H. “Publishing Preprints”. *Nursing Research*, v. 68, n. 5, setembro de 2019, p. 337-38. *DOI.org (Crossref)*, <https://doi.org/10.1097/NNR.0000000000000378>.
- Price, Derek J. De Solla. “Networks of Scientific Papers: The Pattern of Bibliographic References Indicates the Nature of the Scientific Research Front.” *Science*, v. 149, n. 3683, julho de 1965, p. 510-15. *DOI.org (Crossref)*, <https://doi.org/10.1126/science.149.3683.510>.
- Príncipe, Eloísa. “Prática da Ciência Aberta: os preprints em movimento”. *Páginas a&b : Arquivos & Bibliotecas*, n. esp., 2020, p. 59-70. *DOI.org (Crossref)*, <https://doi.org/10.21747/21836671/pagnespc8>.
- Puebla, Iratxe, et al. *Preprints: Their Evolving Role in Science Communication*. Against the Grain, LLC, 2022. *DOI.org (Crossref)*, <https://doi.org/10.3998/mpub.12412508>.
- Richardson, Roberto Jarry. *Pesquisa Social: Métodos e Técnicas*. Atlas, 1999.
- Rodriguez, Ernesto Galbán. “Preprints and Preprint Servers as Academic Communication Tools.” *Revista Cubana de Información en Ciencias de la Salud*, Havana, 2019. <https://web.archive.org/web/20241120182044/https://www.medigraphic.com/cgi-bin/new/resumenI.cgi?IDARTICULO=87074>.
- Rodrigues, Pedro Henrique da Silva. *Análise da diluição de citações entre preprints do medRxiv publicados em periódicos da área da saúde*. 2025. Dissertação (Mestrado) - Universidade Federal do Rio Grande do Sul, orientadora Sônia Elisa Caregnato. Disponível em: <http://hdl.handle.net/10183/300730>.
- Sarabipour, Sarvenaz, et al. “On the Value of Preprints: An Early Career Researcher Perspective”. *PLOS Biology*, v. 17, n. 2, fevereiro de 2019, p. e3000151. *DOI.org (Crossref)*, <https://doi.org/10.1371/journal.pbio.3000151>.
- Severino, Antônio Joaquim. *Metodologia do Trabalho Científico*. Cortez, 2013.
- Silva, Jaime A. Teixeira da. “The preprint debate: What are the issues?” *Medical Journal Armed Forces India*, v. 74, n. 2, abril de 2018, p. 162-64. *ScienceDirect*, <https://doi.org/10.1016/j.mjafi.2017.08.002>.
- Silva, Maurício Coelho Da, et al. “O Sistema de Recompensa Científico na ótica da Ciência Aberta: dimensões de avaliação, características e desafios”. *Transinformação*, v. 36, 2024, p. e2410680. *DOI.org (Crossref)*, <https://doi.org/10.1590/2318-0889202436e2410680>.
- Silva, Fabiano Couto Corrêa Da, e Lúcia Da Silveira. “O ecossistema da Ciência Aberta”. *Transinformação*, v. 31, 2019, p. e190001. *DOI.org (Crossref)*, <https://doi.org/10.1590/2318-0889201931e190001>.

- Silveira, Murilo Artur Araújo da, Sônia Elisa Caregnato, and Leilah Santiago Bufrem. “Estudo das Razões das Citações na Ciência da Informação: Proposta de Classificação.” *Tendências da Pesquisa Brasileira em Ciência da Informação*, vol. 7, no. 2, 2014.  
<https://ancib.org/revistas/index.php/tpbci/article/view/314>.
- Smart, Pippa. “The Evolution, Benefits, and Challenges of Preprints and Their Interaction with Journals”. *Science Editing*, v. 9, n. 1, fevereiro de 2022, p. 79-84. *DOI.org (Crossref)*,  
<https://doi.org/10.6087/kcse.269>.
- Soderberg, Courtney K., et al. “Credibility of Preprints: An Interdisciplinary Survey of Researchers”. *Royal Society Open Science*, v. 7, n. 10, outubro de 2020, p. 201520. *DOI.org (Crossref)*,  
<https://doi.org/10.1098/rsos.201520>.
- Wang, Z., Wolfgang Glänzel, and Y. Chen. “How Self-Archiving Influences the Citation Impact of a Paper: A Bibliometric Analysis of arXiv Papers and Non-arXiv Papers in the Field of Information and Library Science.” *STI 2018 Conference Proceedings*, 2018, pp. 323-330.  
<https://hdl.handle.net/1887/65329>.

---

Copyright: © 2026 RODRIGUES, Pedro Henrique da Silva; WENDT, Lucas George; CAREGNATO, Sônia Elisa. This is an open-access article distributed under the terms of the Creative Commons CC Attribution-ShareAlike (CC BY-SA), which permits use, distribution, and reproduction in any medium, under the identical terms, and provided the original author and source are credited.

---

Submetido: 18/11/2025

Aceito: 10/05/2026