Universidade de São Paulo
Faculdade de Saúde Pública

**Ser ou não ser muito citado: fatores associados à citação de
artigos científicos**

Milena Maria de Araújo Lima Barbosa

**Tese apresentada ao Programa de Pós-
graduação em Saúde Pública da
Faculdade de Saúde Pública da
Universidade de São Paulo para obtenção
do título de Doutora em Ciências.**

Área de concentração: Saúde Pública

Orientador: Prof. Dr. Ivan França Junior

São Paulo

2021

Ser ou não ser muito citado: fatores associados à citação de artigos científicos

Milena Maria de Araújo Lima Barbosa

Tese apresentada ao Programa de Pós-graduação em Saúde Pública da Faculdade de Saúde Pública da Universidade de São Paulo para obtenção do título de Doutora em Ciências.

Área de concentração: Saúde Pública

Orientador: Prof. Dr. Ivan França Junior

Versão revisada

São Paulo

2021

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Aos três presentes mais preciosos que Deus já me deu: meu Senhor e Salvador Jesus Cristo, meu amado esposo Gabriel e meu filho querido Felipe, por preencherem todos os meus dias com amor, paz e alegria.

AGRADECIMENTOS

Ao Prof. Ivan França Júnior, cuja dedicação à ciência, paixão pelo ensino e curiosidade eclética me inspiram desde a graduação;

À Profa. Angela Maria Belloni Cuenca, a quem minha gratidão transcende a vida acadêmica, por me fazer perceber a impossibilidade de expressar em poucas palavras o quão importante uma pessoa pode ser;

À Profa. Maria Teresinha Dias de Andrade, por, juntamente com os professores Ivan e Angela, desenvolver em mim a paixão pela escrita científica;

Ao Prof. Rogério Mugnaini, pelas valiosas contribuições neste estudo e por me ensinar os fundamentos da cientometria;

À Profa. Germana Fernandes Barata, pelas valiosas contribuições para a melhoria deste trabalho;

À Maria do Carmo Avamilano Alvarez, Karoline de Oliveira, Fernanda Paranhos Quinta, Letícia Mucci Conceição, Adriana Almeida Santos, Lívia Mari Sato, Larissa Yuko Omae e Helene Mariko Ueno, e ao Renan Predasoli, pelas contribuições na coleta de dados e nas discussões durante nossas reuniões do grupo de pesquisa;

Ao Programa Institucional de Bolsas de Iniciação Científica, por meio da Pró-Reitoria de Pesquisa da USP, e ao Programa Unificado de Bolsas, da Pró-Reitoria de Graduação da USP, pelas bolsas concedidas às alunas Karoline, Adriana e Lívia;

À Rosa Maria Fischi e Renato Machado de Sobral, da biblioteca da FM-USP, pelas instruções quanto à importação dos metadados da Web of Science;

À Carmen de Almeida Castelani, cuja disposição em ajudar tornou os aspectos burocráticos da pós-graduação muito mais leves;

A todos os professores da FSP-USP e funcionários da Biblioteca da FSP, que há 16 anos vêm participando da minha formação.

A todos vocês e a todos os demais que, de alguma maneira, contribuíram para a realização deste trabalho: muito obrigada!!

A quem tem será dado, e este terá em grande quantidade.

De quem não tem, até o que tem lhe será tirado.

[Mateus 13:12](#)

Barbosa MMAL. Ser ou não ser muito citado: fatores associados à citação de artigos científicos [tese]. São Paulo: Faculdade de Saúde Pública da USP; 2021.

RESUMO

OBJETIVOS: Identificar fatores associados à citação de artigos científicos e validar a acurácia do campo financiamento na Web of Science (WoS). **MÉTODOS:** Foram analisados 3.306 artigos da área de Nutrição e Dietética publicados entre 2009 e 2011, indexados na WoS, distribuídos segundo sua condição de citação (pouco ou nada citados, medianamente citados, altamente citados). Foram testadas variáveis independentes relacionadas ao artigo, à pesquisa, ao periódico e ao autor. Para identificar fatores associados às citações, utilizamos regressão logística multinomial. Foram apresentados os valores de *odds ratio* (OR) e seus respectivos intervalos de confiança de 95% (IC95%) para cada variável independente (regressão simples) e para os valores ajustados no modelo final (regressão múltipla). A validade da base de dados foi avaliada pela estimativa da sensibilidade e especificidade. O nível de significância adotado foi de 5%. **RESULTADOS:** Dos 19 fatores testados, 13 permaneceram independentemente associados às citações. Os atributos associados a maior chance de o artigo ser moderada ou altamente citado foram: tipo de título, tamanho do resumo, quantidade de referências, tamanho do artigo, tipo de publicação, declaração de financiamento, quartil do periódico, número de categorias da WoS às quais o periódico pertence, quantidade de autores, engajamento dos autores e quantidade de países em colaboração. Abordar um *hot topic* ou ser de acesso aberto foram associados a menor chance de o artigo receber citações. **CONCLUSÕES:** Identificamos atributos do artigo, da pesquisa, do periódico e do autor que influenciam as chances de um artigo científico ser citado. Dentre os fatores associados, cinco são atributos relacionados ao artigo e seis são modificáveis, podendo ser trabalhados de modo a aumentar as chances de o artigo ser citado. O campo financiamento da WoS mostrou-se confiável para análises bibliométricas sobre financiamento. O idioma de publicação e o local do artigo onde o financiamento é reportado parecem influenciar a confiabilidade desse campo.

Palavras-chave: Análise de Citações; Fatores Associados; Bibliometria; Cientometria; Bases de Dados de Citações.

Barbosa MMAL. To be or not to be highly cited: factors associated with citation of scientific articles [thesis]. São Paulo: Faculdade de Saúde Pública da USP; 2021.

ABSTRACT

OBJECTIVES: To identify factors associated with citation of scientific articles, and validate the accuracy of the funding field in the Web of Science (WoS). **METHODS:** We analyzed 3,306 articles indexed in the WoS related to Nutrition and Dietetics, published between 2009 and 2011, distributed according to their category of citation (never or rarely cited, moderately cited, highly cited). We tested independent variables related to the article, research, journal, and author. In order to identify factors associated with citations, we used multinomial logistic regression. The odds ratio (OR) values and their 95% confidence intervals were calculated for each independent variable (simple regression), and the values adjusted in the final model (multiple regression). The validation of the database was assessed by estimating sensitivity and specificity. The significance level adopted was 5%. **RESULTS:** Of the 19 factors tested, 13 remained independently associated with citations. The attributes associated with a greater chance of the article being moderately or highly cited were as follows: type of the title, abstract size, number of references, article length, publication type, funding statement, quartile of the journal, number of WoS categories to which the journal belongs, number of authors, authors' engagement, and number of countries in collaboration. Addressing a hot topic, or being open access, was associated with less chance of the article receiving citations. **CONCLUSIONS:** We identified attributes of the article, the research, the journal, and the author that affect the chances of a scientific article being cited. Among the associated factors, five are attributes related to the article, and six are modifiable, which can be worked on to increase the chances of receiving more citations. The WoS metadata on funding is reliable for bibliometric analysis. The language of publication and the location where the funding is reported in the article are likely to influence funding information reliability.

Keywords: Citation Analysis; Associated Factors; Bibliometrics; Scientometrics; Citation Databases.

SUMÁRIO

1 INTRODUÇÃO	10
1.1 BASE HISTÓRICA DA BIBLIOMETRIA.....	11
1.2 ANÁLISE DE CITAÇÕES	12
1.3 WEB OF SCIENCE (WoS).....	17
2 OBJETIVOS.....	19
3 MÉTODOS	19
3.1 DELINEAMENTO DO ESTUDO	19
3.2 AMOSTRA	20
3.3 TAMANHO AMOSTRAL.....	21
3.4 AMOSTRAGEM	21
3.5 EXTRAÇÃO DE DADOS	25
3.6 VARIÁVEIS	25
3.7 VALIDAÇÃO DO CAMPO FINANCIAMENTO DA WOS	32
3.8 ANÁLISE DOS DADOS	32
4 RESULTADOS.....	36
5 DISCUSSÃO.....	47
5.1 FATORES ASSOCIADOS A CITAÇÕES	47
5.1.1 Atributos do Artigo	48
5.1.2 Atributos da Pesquisa	56
5.1.3 Atributos do Periódico.....	59
5.1.4 Atributos do Autor	61
5.2 VALIDAÇÃO DA WOS.....	63
5.3 LIMITAÇÕES	66
5.4 CONCLUSÕES.....	71
6 REFERÊNCIAS	72

1 INTRODUÇÃO

Em 1964, Eugene Garfield, fundador do [*Institute for Scientific Information*](#) (ISI), criou o *Science Citation Index*. Essa ferramenta contabiliza o número de citações que um artigo recebe e viabilizou a aplicação em maior escala de estudos, conduzidos até então por documentalistas, utilizando a contagem de citações na análise da produção científica. Com isso, questionamentos têm sido aventados acerca das citações, como quais fatores podem aumentar (ou diminuir) o potencial de um artigo receber citações de outros pesquisadores? Neste estudo, propomos um modelo teórico de fatores associados à citação, a ser aplicado em um campo específico do conhecimento: o da Nutrição e Dietética. Para tanto, testamos o modelo teórico em artigos de periódicos da área de Nutrição e Dietética indexados na base de dados Web of Science (WoS). Muitos dos fatores aqui estudados são características modificáveis e podem orientar a elaboração de projetos de pesquisa ou a redação de um artigo científico, a fim de melhorar a comunicação científica e, conseqüentemente, aumentar o impacto da pesquisa na ciência. Considerando que o produto de uma análise não pode ser mais acurado que os dados que o geraram, optamos também por avaliar um dos campos da WoS testados neste estudo: o financiamento. Sua inclusão nessa base de dados foi iniciada apenas um ano antes do período aqui estudado e a literatura já apresentava dúvidas por parte dos pesquisadores da área quanto à sua validade no período em que coletamos os dados.

1.1 BASE HISTÓRICA DA BIBLIOMETRIA

A bibliometria, uma subárea das Ciências da Informação, é o estudo dos aspectos quantitativos de produção, disseminação e uso da informação documentada (Tague-Sutcliffe, 1992). Desenvolve padrões e modelos matemáticos para medir esses processos, usando seus resultados para elaborar previsões e apoiar tomadas de decisão.

Um dos primeiros estudos bibliométricos encontrados na literatura é o de Francis J. Cole e Nellie B. Eales, publicado em 1917, no qual foi analisado o crescimento da literatura de anatomia comparativa. Os autores usaram o termo “estatística bibliográfica” para descrever o método usado. Esse termo, porém, dava margem à interpretação de que se tratava da bibliografia sobre estatística e foi posteriormente substituído por *bibliometrics*, termo criado por Paul Otlet em 1934 e popularizado por Alan Pritchard em 1969 (Pritchard, 1969; Hood e Wilson, 2001). Entre as décadas de 30 e 40, foram propostas as leis de produtividade dos pesquisadores (Lei de Lotka), de dispersão de periódicos (Lei de Bradford) e de frequência de palavras (Lei de Zipf), que deram base para os três pilares da Bibliometria (Araújo, 2006).

Em 1965, Derek J. de Solla Price descreveu a distribuição assimétrica das citações recebidas pelos autores. Em 1968, Robert K. Merton formulou o fenômeno cunhado de efeito Mateus, que também observa assimetria no reconhecimento creditado aos pesquisadores, de modo que os com mais prestígio tendem a ser ainda mais prestigiados do que aqueles menos conhecidos pela comunidade científica. Alguns anos depois, Price (1976) propõe o uso da bibliometria para estudo da ciência como um sistema amplo, ao que nomeou “ciência da ciência”,

dando origem à Cientometria. A bibliometria, portanto, está inserida na Cientometria; porém, é comum na literatura o uso de ambos os termos como sinônimos.

Posteriormente, Henry Small (1978) surge analisando as referências a partir do significado que elas adquirem ao serem inseridas no texto. Essa teoria de documentos citados como símbolo de ideias serviu de base para pesquisadores em busca de entender o significado e as motivações das citações. Bornmann e Daniel (2006), em revisão sobre comportamento dos citantes, relatam que os estudos empíricos identificaram que as citações não são feitas motivadas apenas pela vontade de reconhecer as influências intelectuais e cognitivas dos pares, mas também por outros fatores não científicos. Entretanto, Bornmann e Daniel (2006) enfatizam que esses estudos não são confiáveis devido a suas fragilidades metodológicas, como a forma que as citações são contabilizadas e a falta de replicabilidade, além de concluírem que há evidências de que as motivações não são tão aleatórias ao ponto de invalidar o uso do fenômeno da citação como indicador de impacto.

1.2 ANÁLISE DE CITAÇÕES

O ramo da bibliometria que analisa os padrões e a frequência de citações feitas e recebidas é denominado análise de citações. O primeiro artigo a reportar o uso de contagem de citações para avaliar a importância do trabalho científico data de 1927, de autoria de Gross e Gross, no

qual os autores analisaram as citações do principal periódico americano de química para ajudar na decisão de quais periódicos as bibliotecas universitárias deveriam adquirir.

A análise de citações parte de algumas premissas relacionadas às funções das citações (Garfield, 1964), a saber: prestar homenagem aos pioneiros; dar crédito para trabalhos relacionados; identificar metodologia, equipamento etc.; oferecer leitura básica; retificar o próprio trabalho; retificar o trabalho de outros; analisar trabalhos anteriores; sustentar declarações; dar destaque a trabalhos pouco disseminados, inadequadamente indexados ou desconhecidos (não citados); identificar publicações originais que descrevam conceitos ou termos epônimos, por exemplo, Mal de Hodgkin; contestar trabalhos ou ideias de outros; debater a primazia das declarações publicadas; entre outras. Resumidamente, as citações são feitas a todos os trabalhos que contribuíram expressivamente para o atual estudo e apenas aos trabalhos que contribuíram expressivamente para o atual estudo e representam a influência, o impacto de um trabalho. Essas premissas são fundamentadas na visão normativa de Merton (1988).

Entretanto, segundo a visão construtivista, o conteúdo do artigo influencia pouco o número de citações recebidas (Knorr-Cetina, 1981 *Apud* Bornmann e Daniel, 2006, p. 49). Surgem, então, algumas limitações desse tipo de análise que devem ser consideradas nos estudos bibliométricos, como influências formais não citadas; citação tendenciosa ou preconcebida; influências informais não citadas; autocitação; diferentes tipos de citação; variações por tipo de publicação, país, período, extensão e especialidade; e limitações técnicas de índices de citação e bibliografias devido a autorias múltiplas, sinônimos, homônimos, erros de edição e cobertura da literatura (Macroberts e Macroberts, 1989; Macias-Chapula, 1998).

Segundo MacRoberts e MacRoberts (2010), grande parte dos trabalhos que influenciam um autor ou não é citada, ou recebe citações de artigos publicados por revistas indexadas em uma base de dados diferente daquela usada pelo pesquisador que analisa as citações. Portanto, segundo os autores, a análise de citações deve levar em consideração também os trabalhos que recebem pouca ou nenhuma citação. De fato, se os artigos não citados representarem artigos não usados, a razão entre investimento em pesquisas e contribuição à ciência mostraria um cenário preocupante. Em 1990, estudos já encontravam altas proporções (55%) de não citação, com variação expressiva entre as áreas do conhecimento. Por exemplo, as áreas de Física e Química possuíam cerca de 30% de não citação, enquanto Artes e Humanidades apresentavam mais de 90% (Hamilton, 1991; Pendlebury, 1991). Em 2009, um estudo (Larivière et al., 2009) constatou taxa de não citação de 12% em artigos do campo da Medicina, 27% em Ciências Naturais e Engenharias e 32% em Ciências Sociais.

Dentre os motivos não científicos que levam um trabalho a ser citado, estão também a citação a fontes autoritativas (altamente citadas) ou “*hot*” (recentemente publicadas) mesmo quando não são relevantes ao trabalho citante, apenas como modo de atrair a atenção para si. Evidências desses comportamentos questionáveis, porém, têm sido refutadas (Moed e Garfield, 2004; Frandsen e Nicolaisen, 2016; Nicolaisen e Frandsen, 2020).

Assim, apesar das críticas (Bornmann e Daniel, 2006; MacRoberts e MacRoberts, 2010), a análise de citações considera que a importância de um trabalho para a ciência é expressa pela quantidade de vezes que esse trabalho foi explicitamente usado, ou seja, citado por outros pesquisadores, e tem sido usada como ferramenta de avaliação de artigos, pesquisadores, revistas científicas, instituições de ensino e pesquisa e até mesmo de países (Schoonbaert e Roelants, 1996). Com efeito, a ausência de indicadores mais compreensivos (agregando análise quantitativa

e qualitativa) faz com que a análise de citações continue influenciando a reputação de autores e instituições, promoções de pesquisadores e financiamento de pesquisas (Schoonbaert e Roelants, 1996). Adicionalmente, dado que as citações são visíveis e mensuráveis, as análises de citações complementam a avaliação qualitativa da ciência.

Com a publicação do primeiro índice de citações, a bibliometria passou a ter duas aplicações distintas: a avaliativa e a relacional. A primeira busca avaliar o impacto do trabalho acadêmico com o fim de comparar as contribuições científicas entre dois ou mais indivíduos ou grupos, enquanto a segunda procura entender as relações dentro da pesquisa, como os padrões de coautoria nacionais e internacionais (Thelwall, 2008).

Na bibliometria avaliativa, com a comparação dos indivíduos em mãos, instituições direcionam a distribuição de financiamentos e bolsas ou orientam decisões de concursos, por exemplo. A bibliometria avaliativa tem sido amplamente questionada pela supervalorização das citações (Camargo Jr, 2010), demandando forte debate por parte de pesquisadores e editores na busca por métodos mais compreensivos de avaliação, nos quais sejam considerados os impactos da ciência não apenas na comunidade científica, mas na prática profissional e sociedade em geral. Desses esforços, por exemplo, surgem propostas de uso de outras ferramentas, como a contagem de visualizações e *downloads* de artigos ou de suas citações na mídia ou em redes sociais, como proxy do impacto de estudos científicos na academia, nos hospitais, na sociedade. Assim como ocorre com as citações, a utilização de métricas alternativas também falha em captar todo o impacto de resultados científicos sobre a sociedade, dado que nem todo uso gera citações e nem toda visualização gera uso. Nesse contexto, as métricas tradicionais e as alternativas se complementam na avaliação de um trabalho científico para obtenção de uma visão mais ampla, ainda que com limitações, de seu impacto sobre o mundo.

Na bibliometria relacional, por sua vez, as citações podem ser analisadas segundo seu uso pelas instituições ou pelos cientistas, sem a intenção de avaliação. Ressaltamos que o uso das citações com objetivo avaliativo não compõe o objeto desta tese. Aqui, buscamos identificar a maneira como os cientistas usam os artigos científicos – mais especificamente, como se dá parte do processo de citação –, a partir da identificação de fatores que influenciam a frequência de citações que um artigo recebe.

Estudos prévios têm investigado quais características afetam o número de citações recebidas. Os fatores observados vão desde o tamanho do título do artigo (Letchford et al., 2015), formato do resumo (Letchford et al., 2016) e número de autores (Paci et al., 2015) até o nível de evidência do estudo reportado (Okike et al., 2011), entre outros. Tahamtan et al. (2016), em estudo de revisão incluindo 198 publicações, identificaram 28 fatores relacionados ao número de citações e os classificaram em três categorias: fatores relacionados ao artigo, como qualidade do artigo, novidade e interesse do tema abordado, metodologia, tipo de documento, delineamento do estudo, características dos títulos e resumos, características das referências, tamanho do artigo e acessibilidade; fatores relacionados ao periódico, como fator de impacto, idioma e escopo do periódico; e fatores relacionados ao autor, como número de autores, reputação do autor, classificação acadêmica do autor, autocitações, colaboração internacional e nacional, país dos autores, características organizacionais e financiamento. Os autores concluem que alguns fatores como a qualidade do artigo, fator de impacto do periódico, número de autores, visibilidade e colaboração internacional são preditores mais fortes para citações do que sexo, idade e raça dos autores, características dos resultados e características da discussão, por exemplo.

Uma das características tidas como influenciadoras na citação de artigos, por exemplo, é a qualidade do estudo reportado, representada pelo recebimento de financiamento por parte de

agências de fomento à pesquisa. Uma hipótese foi levantada em estudo de Cuenca et al. (2011) de que há relação entre “qualidade da pesquisa, financiamento e publicação em periódicos de prestígio, uma vez que pesquisas financiadas, provavelmente, já tenham passado por um crivo de qualidade e relevância no momento da aprovação do projeto pela agência financiadora. As agências financiadoras, ao exigirem na publicação do artigo (...) a nomeação explícita do financiamento, estariam colaborando para a completeza das informações sobre a produção científica” (p. 109). Conen et al. (2008) mostraram que, na área de cardiologia clínica, os ensaios financiados por organizações com fins lucrativos são mais citados do que os financiados por organizações sem fins lucrativos. Wang e Shapira (2015) observaram que pesquisas financiadas na área de nanotecnologia receberam mais citações do que as que não receberam auxílio algum. Com efeito, o fato de uma pesquisa ter sido subsidiada pode indicar qualidade, pertinência e credibilidade, o que também atrairia um maior número de citações.

1.3 WEB OF SCIENCE (WoS)

Em 1955, Eugene Garfield (Garfield, 1955) propôs um sistema bibliográfico da literatura científica, descrevendo-o como um índice de associação de ideias que indica o relacionamento entre os trabalhos científicos. Em 1964, Garfield cria o primeiro Science Citation Index, publicado inicialmente em papel e que acompanhou os recursos disponíveis no decorrer do tempo, passando para publicação em CD-ROM e posteriormente em plataforma web, em 1997, ao que chamou Web of Science (WoS). Sua empresa, o ISI, foi adquirida pela Thomson, posteriormente

fundida com a Reuters, tornando-se a Thomson-Reuters em 2008 e, após venda da divisão de Intellectual Property and Science, tornou-se em 2016 a atual Clarivate Analytics.

A então WoS da Thomson Reuters, em 2008, passou a inserir em seus registros a menção ao financiamento, coletada do campo *funding acknowledgement* dos artigos e presente no campo FU dos metadados importados da WoS. Essa inovação favoreceu análises bibliométricas sobre agências de fomento a pesquisas (Wang e Shapira, 2011; Tang et al., 2017), uma vez que, anteriormente, a coleta desse dado se dava somente por meio da análise individual do artigo ou rastreando relatórios de pesquisa (Wang e Shapira, 2011; Tang et al., 2017).

Vários estudos discutem a acurácia e completeza da WoS nas informações sobre financiamento de pesquisa (Wang e Shapira, 2011; Tang et al., 2017; Alvarez-Bornstein et al., 2017). Tang et al. (2017) identificaram substanciais inconsistências em registros de artigos publicados em outros idiomas que não o inglês nos campos de *financial acknowledgement*. Shu et al. (2019) encontraram tais limitações em estudo sobre a literatura científica chinesa presente na WoS. Com efeito, uma vez que os registros de financiamento dos artigos foram incorporados mais recentemente na base de dados WoS, é necessário avaliar se esses dados são confiáveis para inclusão nas análises de citações (Paul-Hus et al., 2016).

Assim, dado o contexto em que as pesquisas não parecem integrar um modelo compreensivo dos fatores associados às citações, visto que são analisados apenas um (Letchford et al., 2016) ou poucos (Filion e Pless, 2008) fatores em cada estudo, bem como apenas um periódico (Bhandari et al., 2007), somado ao fato de as análises de citações possuírem muitas limitações, como os problemas técnicos dos índices bibliográficos (Macroberts e Macroberts, 1989; Macias-Chapula, 1998), aventamos as seguintes perguntas: 1) quais atributos do artigo, da

pesquisa, do periódico e do autor estão associados às citações de artigos científicos? 2) podemos considerar válidos os dados sobre financiamento fornecidos pelas bases de dados?

2 OBJETIVOS

Identificar fatores associados à citação de artigos científicos;

Validar a acurácia do campo financiamento na WoS.

3 MÉTODOS

3.1 DELINEAMENTO DO ESTUDO

Neste estudo, comparamos os artigos moderadamente ou altamente citados com aqueles pouco citados. Foi testado um modelo compreensivo de fatores potencialmente associados às citações de artigos científicos. Aqui, também testamos a acurácia do campo financiamento da base de dados WoS.

3.2 AMOSTRA

O modelo foi testado em publicações indexadas pela WoS sob a categoria Nutrição e Dietética (Nutrition & Dietetics). A Nutrição e Dietética é uma área da Medicina, que tem uma forte tradição de alta citação de artigos científicos (Larivière et al., 2009). Essa característica favorece a análise de citações, dada a densidade possível no objeto de estudo.

Para a escolha da base de dados, foram consideradas as características da WoS: indexa mais de 20 mil revistas; possui classificação para os periódicos (categorias às quais pertencem); e fornece os metadados para grande parte das variáveis analisadas neste estudo, incluindo financiamento e número de categorias às quais os periódicos e, conseqüentemente, os artigos pertencem.

Neste estudo, foram incluídas publicações apenas de periódicos cujo processo de avaliação é feito por revisão por pares. Incluímos cartas (Carta ao Editor), artigos de pesquisa original e artigos de revisão, por serem os principais tipos de publicação de pesquisas científicas e terem maior potencial para recebimento de citações. Inicialmente, esses três tipos de publicação foram considerados separadamente nas análises; entretanto, a pequena quantidade de cartas na amostra, sobretudo no grupo de alta citação, poderia superestimar a magnitude do efeito de associação entre as variáveis. Assim, foi decidido agrupar as cartas e os artigos de pesquisa original na mesma categoria e compará-los com as revisões.

3.3 TAMANHO AMOSTRAL

Para o cálculo do tamanho amostral mínimo por grupo, considerou-se os seguintes parâmetros: magnitude de efeito = 0,02; poder do teste = 0,80; α = 0,05; número de preditores = 22. No total, o tamanho amostral mínimo necessário foi de 1.102 artigos/grupo^a.

3.4 AMOSTRAGEM

A data de publicação e a disciplina a qual um artigo pertence são fatores que muito influenciam o número de citações recebidas e devem ser considerados em análises de citações (Craig et al., 2007). Para minimizar esses efeitos, selecionamos uma disciplina para este estudo e padronizamos o tempo após publicação a ser considerado na contagem das citações para todos os artigos.

Wang (2012), em seu estudo sobre escolha da janela de tempo de citação para avaliação do impacto da pesquisa, analisou o comportamento das citações recebidas em 31 anos de todos os artigos indexados na WoS publicados em 1980. O autor recomenda que, para pesquisas de avaliação do impacto geral e incluindo todos os artigos (não apenas os mais citados, por exemplo), uma janela de citação de quatro anos seria suficiente. Larivière et al. (2007), usando uma janela de citação de 100 anos, identificaram que na área das ciências médicas – em que a

^a Soper D. Calculator: a-priori sample size for multiple regression. Disponível em: <https://www.danielsoper.com/statcalc/calculator.aspx?id=1>

Nutrição e Dietética está inclusa –, a idade média dos documentos citados fica entre cinco e seis anos. Abt (1981), em seu estudo com artigos da área de astronomia, identificou que os artigos recebem o máximo de citações/ano após cinco anos de publicação. Paci et al. (2015), buscando identificar fatores associados a citações, porém em amostra de ensaios clínicos randomizados da área de fisioterapia, também utilizaram essa janela de tempo de citações². Similarmente, visando a analisar as citações recebidas após, pelo menos, cinco anos desde a publicação de cada artigo, incluímos neste estudo artigos publicados de 2009 a 2011.

No total, 38.675 registros foram publicados e indexados na base de dados sob a categoria alvo desta pesquisa, a Nutrição e Dietética. A busca pelos registros foi realizada, em 30 de setembro de 2017, na plataforma da WoS, que possibilita a importação dos dados bibliográficos. Foram excluídos livros (n = 515), artigos retratados, biografias, erratas, editoriais, resumos de eventos, notícias, capítulos de livros, *proceedings papers* e reimpressões (n = 10.569), restando 27.591 registros para seleção da amostra final avaliada neste estudo. Após importação desses registros, selecionamos os identificadores únicos da WoS para cada registro e os inserimos no campo de busca da WoS para importação dos relatórios de citações. A Figura 1 esquematiza a seleção da amostra.

Para cada artigo, calculamos o número total de citações recebidas em até cinco anos após a publicação, a saber: citações até 2014 das publicações de 2009; citações até 2015 das publicações de 2010; e citações até 2016 das publicações de 2011. Apesar da diferença do tempo de exposição entre artigos publicados, por exemplo, no início e no fim do mesmo ano, a adoção de uma janela de citações minimiza a incomparabilidade entre o número de citações recebidas pelos artigos.

Figura 1. Seleção da amostra.

Os registros foram listados por ordem decrescente de número de citações recebidas após cinco anos e divididos em percentis. Os artigos acima do percentil 90 foram considerados altamente citados; aqueles entre 45% e 55%, medianamente citados; e aqueles abaixo do percentil 10, pouco ou nada citados. Entretanto, o número de artigos em cada grupo foi maior que 10% da amostra porque consideramos o percentil apenas para definir o número de citações mínimo e máximo; uma vez definido esse número, incluímos em cada grupo todos os artigos que receberam aquele número de citações. Por exemplo, o p45 foi representado por 9 citações; porém, do intervalo entre o p45 até o p42, todos os artigos receberam 9 citações em cinco anos. Já no p55, correspondente a 13 citações, incluímos no grupo todos os artigos que receberam 13 citações em até cinco anos, a saber, todos aqueles entre o p55 e o p57. Logo, foram considerados artigos

medianamente citados aqueles entre o p42 e o p 57. Dessa forma, os 27.591 registros selecionados estavam distribuídos da seguinte maneira: 2.893 artigos no percentil 90 ($p_{90} \geq 38$ citações), 4.496 no percentil 42–57 ($p_{42}-p_{45} = 9$ citações; $p_{45}-p_{57} = 13$ citações) e 2.954 no percentil 10 ($p_{10} \leq 1$), conforme mostrado na Figura 2.

De cada grupo, foram selecionados aleatoriamente 1.102 registros, utilizando-se uma lista de números gerada por meio eletrônico^b.

Figura 2. Seleção da amostra final de acordo com os percentis considerados.

^b Urbaniak GC, Plous S. Research Randomizer (Version 4.0) [Computer software]. 2013 [citado 28 jan 2018]. Disponível em: www.randomizer.org

3.5 EXTRAÇÃO DE DADOS

Para a seleção dos dados a serem extraídos de cada artigo, usamos o modelo teórico elaborado de acordo com a literatura e com hipóteses de associações propostas pelos pesquisadores deste estudo. Para a extração dos dados, foram importados da WoS os seguintes campos: título; resumo; descritores; referências; páginas; idioma; ano de publicação; tipo de publicação; financiamento; periódico; categoria WoS; autores; afiliações; ResearcherID; e Open researcher and contributor ID (Orcid). Em seguida, foi importado o relatório de citações recebidas pelas publicações dos anos 2009 a 2011. Adicionalmente, foram consultados: o Banco Mundial^c, para coleta dos dados de renda dos países; o Ulrich, para consulta do ano de criação dos periódicos; e o Journal Citation Reports – Clarivate^d, para coleta do fator de impacto e quartil^e de cada periódico em cada ano incluído neste estudo.

3.6 VARIÁVEIS

A variável dependente foi a condição de citação com três categorias (citação baixa, citação moderada, citação alta). A inclusão de autocitações nos estudos bibliométricos é criticada sob o argumento de que podem inflar o impacto de um artigo. De acordo com estudo de Glänzel

^c The World Bank Group. DataBank: World Development Indicators. [citado 21 mai 2018]. Disponível em: <https://databank.worldbank.org/source/world-development-indicators>

^d Clarivate. Journal Citation Reports. Disponível em: <https://jcr.clarivate.com>

^e Os quartis dos periódicos foram coletados e gentilmente cedidos pelo Prof. Dr. Rogério Mugnaini.

et al. (2006), nem os modelos “aritméticos puros”, como o aumento de autocitações em função do número de coautores, nem a abordagem “psicológica pura”, a saber, que a autocitação é sujeita à vontade e arbitrariedade dos autores, ocorrem na prática. De fato, podem existir casos abusivos de autocitações, em que o único motivo de um autor citar um artigo reside no fato de este ser de sua autoria, para fins de autopromoção. Entretanto, espera-se que os cientistas desenvolvam suas carreiras dentro de uma ou poucas linhas de pesquisa, fazendo com que seus estudos prévios embasem os subsequentes, o que justificaria autocitações a seus artigos anteriores. Aqui, consideramos que os pesos de ambos os tipos de citações são, a priori, iguais e suficientemente relevantes para serem incluídos nesta análise.

As variáveis independentes estão descritas no Quadro 1. Entre as variáveis que expressavam condições similares, a saber fator de impacto e quartil do periódico, escolheu-se aquela que melhor representava a condição de citação, excluindo a variável fator de impacto, dado que o quartil é mais estável no decorrer do tempo. Após importação das afiliações dos autores na WoS, extraímos as instituições e os países de cada instituição desse campo. Verificamos a existência de colinearidade moderada entre a variável número de instituições e as variáveis quantidade de autores ($\rho = 0,513$; $p < 0,01$) e de países ($\rho = 0,532$; $p < 0,01$), enquanto as variáveis quantidade de autores e de países apresentaram baixa correlação ($\rho = 0,301$; $p < 0,01$). Dessa forma, optamos por excluir a variável quantidade de instituições da análise.

Para as variáveis nível de renda do país do primeiro autor e diversidade de níveis de rendimento dos países dos autores, foram considerados a classificação dos países nos anos de 2009, 2010 e 2011, conforme o ano de publicação de cada artigo. Os dados de rendimento foram coletados no sítio eletrônico do Banco Mundial^c, categorizados em: baixo ($\leq$ US\$ 1.005), médio-baixo (US\$ 1.006–3.955), médio-alto (US\$ 3.956–12.365) e alto (US\$ $>$ 12.365).

Quadro 1. Variáveis independentes consideradas para análise.

Dimensão	Variável	Categorias
Atributos do artigo	Tamanho do título	Curto; médio; longo
	Tipo de título	Assunto ou pergunta; conclusão
	Tamanho do resumo	$\leq$ mediana; $>$ mediana
	Quantidade de referências citadas	$\leq$ mediana; $>$ mediana
	Tamanho do artigo	$\leq$ mediana; $>$ mediana
	Idioma em que o artigo foi escrito	Outro idioma; inglês
	Acesso aberto ao artigo	Não; sim
Atributos da pesquisa	Tipo de publicação	Carta ou artigo original; revisão
	Se o artigo contém declaração de financiamento	Não; sim
	Se a genômica nutricional foi abordada	Não; sim
Atributos do periódico	Ano de criação do periódico	$\leq$ mediana; $>$ mediana
	Impacto do periódico	4Q ou 3Q; 2Q; 1Q
	Número de categorias da WoS às quais o periódico pertence	1; 2; $\geq$ 3
Atributos do autor	Quantidade de autores	Variável contínua
	Quantidade de países em colaboração	Variável contínua
	Diversidade de níveis de rendimento dos países dos autores	1; 2 ou mais
	Renda do país do primeiro autor	Baixa ou média baixa; média alta; alta
	Engajamento dos autores	Nenhum autor com RI ou OI; ao menos 1 autor com RI ou OI; ao menos 1 autor com RI e OI
	Se há alguma declaração sobre conflito de interesses	Não; sim

4Q: 4º quartil; 3Q: terceiro quartil; 2Q: segundo quartil; 1Q: primeiro quartil; WoS: Web of Science; RI: ResearcherID; OI: Orcid.

A variável engajamento dos autores foi elaborada a partir de duas informações obtidas na WoS: os registros dos identificadores únicos de autores ResearchID e Orcid. O engajamento foi categorizado da seguinte forma: baixo, quando nenhum autor vinculou o artigo ao seu registro no ResearchID ou Orcid; moderado, quando ao menos um autor vinculou o artigo ao registro no ResearchID ou Orcid; e alto, quando ao menos um autor vinculou o artigo em ambos os identificadores únicos de autores.

As variáveis tipo de título, acesso aberto ao artigo e declaração de conflito de interesses foram coletadas de forma manual e independente por dois pesquisadores; nos casos em que houve divergência na classificação, uma nova avaliação foi feita por um terceiro pesquisador ou em conjunto até obtenção de um consenso. Os títulos foram categorizados em: assunto, quando descreve apenas o tema abordado no artigo; pergunta, quando apresentam a pergunta de pesquisa ou hipótese; e conclusão, quando descreve os principais achados. Jamali e Nikzad (2011) apresentam essas categorias com os nomes de, respectivamente, descritivo (ou título neutro), interrogativo e declarativo. Apesar de os termos apresentarem diferenças entre si, seus significados são os mesmos que usamos neste estudo. Dado o baixo número de artigos com títulos do tipo pergunta, as categorias assunto e pergunta foram posteriormente agrupadas. O Quadro 2 apresenta exemplos de cada tipo de título, onde podemos observar no exemplo 9 que, em casos de títulos compostos, consideramos como do tipo conclusão se esta informação estivesse presente. O acesso foi considerado aberto quando o artigo estava disponível na íntegra com acesso gratuito em sítios eletrônicos de busca ou no próprio sítio do periódico. A busca por declaração de conflito de interesses se deu no arquivo PDF de cada artigo e considerou toda a extensão do artigo, como presença da declaração em um campo específico para esse fim ou em notas de rodapé.

Quadro 2. Tipos de título considerados na análise e respectivos exemplos obtidos da amostra.

Tipo de título	Exemplos
Assunto	1) Habitual intake of flavonoid subclasses and incident hypertension in adults
	2) Obesity, diets, and social inequalities
	3) New orientations of research in Chinese Medicines
Pergunta	4) Can infant feeding choices modulate later obesity risk?
	5) Eating style, overeating, and overweight in a representative Dutch sample. Does external eating play a role?
	6) Is sleeping more and working less a new way to control our appetite?
Conclusão	7) Vitamin D supplementation enhances the beneficial effects of weight loss on cardiovascular disease risk markers
	8) Circadian Timing of Food Intake Contributes to Weight Gain
	9) Do all sedentary activities lead to weight gain: sleep does not

A variável paradigma da Nutrição foi operacionalizada considerando-se um assunto em evidência (*hot topic*) no período analisado: a genômica nutricional. O avanço da ciência e tecnologia possibilitou o sequenciamento do genoma humano, concluído no ano de 2003, abrindo caminho para a era pós-genômica nas áreas da saúde. A genômica nutricional engloba três disciplinas – a nutrigenômica, a nutrigenética e a epigenética – que objetivam investigar as interações entre os nutrientes e os genes. O termo *nutrigenomics* foi introduzido no vocabulário

controlado Medical Subject Headings (Mesh) em 2008. Neste estudo, buscamos entre os descritores e títulos dos artigos por termos relacionados ao tema, a saber: *genome*, *genomics*, *genetic*, *genetics*, *nutrigenomics*, *nutrigenetics*, *epigenetics*, *gene*, *genes*, *polymorphism*, *polymorphisms*.

Todos os dados foram registrados em planilha eletrônica. A limpeza dos dados consistiu em excluir trechos finais que não pertenciam ao texto de alguns resumos (trechos adicionados na indexação) e padronizar os nomes dos periódicos e dos países.

Dado que as variáveis foram categorizadas considerando-se o sentido teórico e o comportamento estatístico, algumas variáveis foram recategorizadas após a análise descritiva devido à baixa frequência de algumas categorias, a saber: tipo de título, recategorizada em assunto+pergunta e conclusão; quartil do periódico, recategorizada em 4Q+3Q, 2Q e 1Q; diversidade de níveis de rendimento dos países dos autores, inicialmente categorizada em 1, 2 e 3 e posteriormente recategorizada em 1 e ≥ 2 ; e nível de rendimento do primeiro autor, recategorizada em baixa+média baixa, média alta e alta.

A variável tamanho do título foi categorizada, segundo tercís, em curto (até 12 palavras), médio (13 a 16 palavras) e longo (17 palavras ou mais). Já as variáveis dicotomizadas a partir da mediana foram: tamanho do resumo (≤ 218 palavras; > 218 palavras), quantidade de referências citadas (≤ 34 ; > 34), tamanho do artigo (≤ 7 páginas; > 7 páginas) e ano de criação do periódico (≤ 1976 ; > 1976). A quantidade de autores, de instituições e de países em colaboração foram mantidas como variáveis quantitativas contínuas.

Inicialmente, foram incluídas 3.306 publicações, das quais 61 foram excluídas pelos seguintes motivos: 42 não eram cartas, artigos originais ou revisões, mas sim editoriais, resumos

ou notícias; 19 não foram localizadas. Assim, as análises estatísticas foram realizadas considerando-se 3.245 publicações, cujas distribuições segundo ano de publicação e condição de citação estão apresentadas na Figura 3 e na Tabela 1, respectivamente.

Figura 3. Publicações consideradas na pesquisa, 2009–2011.

Tabela 1. Casos válidos segundo condição de citação.

Condição de citação	Validação			Total
	Caso válido	Excluído - outro tipo de publicação	Excluído - não encontrado	
Baixa	1.041	42	19	1.102
Moderada	1.102	0	0	1.102
Alta	1.102	0	0	1.102
Total	3.245	42	19	3.306

3.7 VALIDAÇÃO DO CAMPO FINANCIAMENTO DA WOS

Para validação da acurácia da WoS, além da exclusão dos artigos que não eram cartas, originais ou de revisão, excluímos também aqueles que não foram encontrados na íntegra, dada a impossibilidade de averiguação manual da declaração de financiamento na publicação. Portanto, foram analisados 3.183 artigos. Na WoS, a busca foi realizada especificando o campo do conhecimento Nutrição e Dietética e, nos registros, os campos codificados como “FU” = Agência do Financiamento e “FX” = Texto do financiamento. Para validar os resultados obtidos na WoS, foi realizada uma verificação em cada artigo por meio de consulta aos PDF em tela, buscando encontrar no texto a menção ao fomento obtido, tanto em notas de rodapé, no campo conflito de interesses e no campo de agradecimentos, como na própria nota sobre financiamentos. Esse processo foi chamado de revisão manual. Em seguida, ambos os resultados, os da base de dados e os da revisão manual, foram comparados.

3.8 ANÁLISE DOS DADOS

A quantidade de autores e a quantidade de países em colaboração foram apresentadas por mediana e percentis 25 e 75. As demais variáveis quantitativas foram categorizadas conforme descrito na seção 3.6 (variáveis). As variáveis qualitativas foram apresentadas pelas frequências absolutas e relativas.

A correlação entre as variáveis número de instituições, de autores e de países foi testada pelo coeficiente de correlação de postos de Spearman (ρ). Na comparação das variáveis quantidade de autores e de países em colaboração, após verificação da normalidade da distribuição dos dados pelo teste Kolmogorov-Smirnov, optou-se pelo teste não paramétrico de Kruskal-Wallis. Associação entre variáveis qualitativas foi verificada pelo teste qui-quadrado de Pearson.

Dado que a variável dependente é categórica e possui mais de duas categorias, utilizamos a regressão logística multinomial para avaliar os fatores relacionados com a condição de citação das publicações. Modelos de regressão permitem a obtenção do efeito de cada preditor de forma independente; assim, se por exemplo o financiamento resultar associado a um maior número de citações, é porque essa associação ocorre independentemente dos demais fatores (como o quartil ao qual o periódico pertence, o número de autores etc).

Na construção do modelo de regressão logística múltiplo, consideramos os fatores que resultaram com valor de $p < 0,20$ na análise univariada. Adicionalmente, excluiu-se a variável idioma por ser uma variável com frequência zero em uma das categorias. Foram apresentados os valores de *odds ratio* (OR) e seus respectivos intervalos de confiança de 95% (IC95%) para cada uma das variáveis independentes (regressão simples) e para os valores ajustados no modelo final (regressão múltipla), bem como os coeficientes beta. O nível de significância adotado foi de 5% para todos os testes de hipóteses. As análises foram realizadas no software *SPSS v.25 for Windows*.

A validade da base de dados foi avaliada pela estimativa da sensibilidade e especificidade. Os registros WoS foram comparados com a revisão manual e calculados os índices de sensibilidade e especificidade da informação sobre financiamento na WoS. A sensibilidade foi

entendida como a capacidade da WoS de detectar casos positivos verdadeiros de revisão manual de cada artigo. Esse índice foi estimado como a proporção de registros positivos nas duas técnicas pelo total de registros positivos na revisão manual. Um banco de dados menos sensível significa maior número de falsos negativos na WoS, resultando na subestimação da prevalência de financiamento dos artigos. A especificidade da WoS é uma medida da proporção de negativos que são identificados corretamente como tais. Uma base de dados menos específica produz um maior número de falsos positivos, aumentando a prevalência de artigos financiados.

A sensibilidade e a especificidade foram utilizadas para descrever quantitativamente o desempenho da WoS por meio da construção de sua respectiva curva ROC. A curva ROC (*receiver operator characteristic curve*) é uma expressão gráfica da relação entre sensibilidade e especificidade, construída em um plano cartesiano. No eixo das ordenadas (eixo Y), são inseridos os valores em decimais da sensibilidade (proporção de verdadeiros positivos) e, no eixo das abscissas (eixo X), os valores complementares da especificidade ($1 - \text{especificidade}$, equivalentes à proporção de falsos positivos). Além de servirem para escolher o melhor ponto de corte de um teste diagnóstico, as curvas ROC são utilizadas também para comparar dois ou mais testes diagnósticos para uma mesma doença. Analogamente, neste estudo, usamos essa curva para comparar as duas formas de identificação do financiamento, a saber, por meio dos metadados da WoS e da revisão manual. A acurácia global pode ser mensurada pelo cálculo da área sob a curva, dado que a dimensão dessa área é diretamente proporcional ao poder discriminatório do teste, ou seja, quanto mais próxima de 1 (ou seja, 100%), mais acurado é o teste.

Todas as análises foram realizadas usando DIAGT, um módulo Stata projetado para relatar estatísticas resumidas para testes de diagnóstico. As diferenças entre os valores de

sensibilidade foram avaliadas por meio do teste de proporções, considerando-se um nível de confiança de 0,05.

Para se obter um diagnóstico, seja de uma doença, seja de qualquer outra condição, é necessário que o resultado da avaliação possa ser reproduzido por quem quer que esteja avaliando. Em outras palavras, o ideal seria que todas as pessoas que tentassem diagnosticar uma condição chegassem ao mesmo resultado, a saber, que a condição está presente ou ausente quando de fato ela estiver presente ou ausente, respectivamente. Entretanto, essa concordância absoluta entre dois avaliadores distintos não ocorre na prática. Assim, uma forma de saber se uma dada classificação pode ser considerada confiável é utilizando o coeficiente Kappa, que descreve a intensidade de concordância entre esses avaliadores. Neste estudo, a estatística Kappa foi calculada para cada local do artigo em que o financiamento foi mencionado, usando o KAPCI, um módulo Stata projetado para medir a concordância entre duas medidas ou avaliadores. O coeficiente Kappa varia de 0^f, quando o grau de concordância é o que se esperava ao acaso, a 1, quando há concordância perfeita. Seguimos a classificação sugerida por Landis e Koch (1977): 0,00–0,20 = leve; 0,21–0,40 razoável; 0,41–0,60 = moderado; 0,61–0,80 = substancial; 0,81–1,00 = quase perfeito.

Adicionalmente, dado que uma das limitações da WoS sugeridas na literatura é o idioma e que a menção de recebimento de financiamento encontra-se geralmente no campo agradecimentos dos artigos, buscamos avaliar se a WoS captura as informações sobre financiamento independentemente do idioma e do local do artigo onde esse dado foi mencionado.

^f O coeficiente Kappa pode ser um valor abaixo de zero e, portanto, negativo, o que sugere que a concordância encontrada foi menor do que aquela esperada ao acaso; porém, esse valor negativo não apresenta interpretação estatística em relação à intensidade de discordância.

Assim, testamos se o idioma e o local do artigo onde consta o financiamento estavam associados ao fato de o financiamento estar reportado no artigo mas não ser identificado pela WoS. Para isso, usamos o teste qui-quadrado de Pearson, considerando nível de significância de 0,05.

3.7 ASPECTOS ÉTICOS

Por se tratar de pesquisa cuja amostra é composta por artigos científicos, e não por humanos ou animais, este estudo está isento de avaliação por comitê de ética.

4 RESULTADOS

4.1 FATORES ASSOCIADOS A CITAÇÕES

As características das publicações segundo a condição da citação são apresentadas na Tabela 2. Todas as variáveis resultaram associadas com a condição de citação e foram estatisticamente significativas ($p < 0,05$).

Tabela 2. Características das publicações incluídas no estudo, 2009–2011.

Variável	Categoria	Condição de citação						Valor de p ^a
		Baixa		Moderada		Alta		
		n = 1041		n = 1102		n = 1102		
		n	%	n	%	n	%	
Atributos do artigo								
Tamanho do título ^c	Curto	522	50,2	358	32,5	422	38,3	< 0,001
	Médio	274	26,3	380	34,5	334	30,3	
	Longo	244	23,5	364	33,0	346	31,4	
Tipo de título ^c	Assunto ou pergunta	918	88,3	838	76,0	886	80,4	< 0,001
	Conclusão	122	11,7	264	24,0	216	19,6	
Tamanho do resumo	≤ mediana	769	73,9	540	49,0	491	44,6	< 0,001
	> mediana	272	26,1	562	51,0	611	55,4	
Quantidade de referências citadas no artigo ^c	≤ mediana	788	76,7	562	51,0	316	28,7	< 0,001
	> mediana	240	23,3	540	49,0	785	71,3	
Tamanho do artigo	≤ mediana	694	66,7	614	55,7	434	39,4	< 0,001
	> mediana	347	33,3	488	44,3	668	60,6	
Idioma em que o artigo foi escrito	Outro idioma	247	23,7	3	0,3	0	0	< 0,001
	Inglês	794	76,3	1.099	99,7	1.102	100	
Acesso aberto ao artigo	Não	375	36,0	544	49,4	427	38,7	< 0,001
	Sim	666	64,0	558	50,6	675	61,3	
Atributos da pesquisa								
Tipo de publicação	Carta ou artigo original	982	94,3	1.040	94,4	881	79,9	< 0,001
	Revisão	59	5,7	62	5,6	221	20,1	
Se o artigo contém declaração de financiamento	Não	764	73,4	352	31,9	326	29,6	< 0,001
	Sim	277	26,6	750	68,1	776	70,4	
Se a nutrigenômica foi abordada no artigo	Não	962	92,4	1.001	90,8	1.043	94,6	0,003
	Sim	79	7,6	101	9,2	59	5,4	
Atributos do periódico								
Ano de criação do periódico	> mediana	440	42,3	575	52,2	669	60,7	< 0,001
	≤ mediana	601	57,7	527	47,8	433	39,3	
Quartil do periódico ^c	4Q/3Q	556	58,3	244	23,0	82	7,6	< 0,001
	2Q	181	19,0	359	33,8	191	17,8	
	1Q	217	22,7	459	43,2	802	74,6	
Número de categorias da WoS às quais o periódico pertence	1 categoria	559	53,7	422	38,3	478	43,4	< 0,001
	2 categorias	366	35,2	366	33,2	380	34,5	
	3 ou mais categorias	116	11,1	314	28,5	244	22,1	
Atributos do autor								
Quantidade de autores ^b		3	(2;5)	5	(3;6)	5	(3;7)	< 0,001
Quantidade de países ^b		1	(1;1)	1	(1;1)	1	(1;2)	< 0,001
Diversidade de níveis de rendimento dos países dos autores ^c	1	948	94,5	993	90,1	1015	92,6	0,001
	2 ou mais	55	5,5	109	9,9	81	7,4	
Renda do país do primeiro autor ^c	Baixa ou média baixa	52	5,2	68	6,2	41	3,7	< 0,001
	Média alta	203	20,2	166	15,1	91	8,3	
	Alta	748	74,6	868	78,8	964	88	
Engajamento dos autores	Nenhum autor tem RI ou OI	744	71,5	556	50,5	470	42,6	< 0,001
	Ao menos 1 autor com RI ou OI	144	13,8	226	20,5	262	23,8	
	Ao menos 1 autor com RI e OI	153	14,7	320	29,0	370	33,6	
Se há alguma declaração sobre conflito de interesses	Não	440	42,3	510	46,3	451	40,9	0,031
	Sim	601	57,7	592	53,7	651	59,1	

4Q: 4º quartil; 3Q: terceiro quartil; 2Q: segundo quartil; 1Q: primeiro quartil; RI: ResearcherID; OI: Orcid.

^a Teste qui-quadrado de Pearson

^b Mediana (p25; p75), teste Kruskal-Wallis

^c A contagem não totaliza 100% do n devido aos dados omissos.

Para a análise dos fatores associados com a condição de citação, inicialmente foram calculados os valores de *odds ratio* (OR) e respectivos intervalos de confiança de 95% considerando o modelo de regressão logística multinomial. Os resultados da análise univariada são apresentados na Tabela 3. A variável acesso aberto ao artigo apresentou significância só no grupo de citação moderada. As variáveis tipo de publicação, ano de criação do periódico, nível de renda do país do primeiro autor e conflito de interesse apresentaram $p < 0,20$ em uma de suas categorias e, portanto, também foram incluídas na análise múltipla.

O modelo final com os fatores associados com a citação é apresentado na Tabela 4. Apesar de significativa na análise univariada, a variável idioma não foi incluída na análise múltipla devido à frequência zero no grupo com alta citação. Os resultados mostram que publicações cujo tipo de título é apresentado como conclusão têm 1,77 vezes chance de ser moderadamente citadas, comparadas àquelas com título do tipo assunto ou pergunta. No grupo de alta citação, essa variável perdeu a associação na análise múltipla. O tamanho do resumo e a quantidade de referências citadas no artigo são fatores associados tanto com a citação moderada como com a alta. Se o artigo contém declaração de financiamento existe um aumento na chance de serem moderadamente e altamente citados. O quartil do periódico também foi um fator associado com as duas condições de citação (moderada e alta) em relação à baixa citação; quanto menor o quartil do periódico, maior a chance de o artigo ser moderada ou altamente citado. Quanto maior o número de autores e o número de países em colaboração, maiores as chances de o artigo ser moderada ou altamente citado.

Tabela 3. Regressão logística multinomial simples para a condição de citação, 2009–2011.

Variável	Categoria	Moderada		Alta	
		OR (IC95%)	Valor de p	OR (IC95%)	Valor de p
Atributos do artigo					
Tamanho do título	Curto	1		1	
	Médio	2,02 (1,65–2,48)	< 0,001	1,51 (1,23–1,85)	< 0,001
	Longo	2,18 (1,76–2,69)	< 0,001	1,75 (1,43–2,16)	< 0,001
Tipo de título	Assunto ou pergunta	1		1	
	Conclusão	2,37 (1,88–3,00)	< 0,001	1,83 (1,44–2,33)	< 0,001
Tamanho do resumo	≤ Mediana	1		1	
	> Mediana	2,94 (2,45–3,53)	< 0,001	3,52 (2,93–4,22)	< 0,001
Quantidade de referências citadas no artigo	≤ Mediana	1		1	
	> Mediana	3,15 (2,62–3,80)	< 0,001	8,16 (6,71–9,91)	< 0,001
Tamanho do artigo	≤ Mediana	1		1	
	> Mediana	1,59 (1,33–1,89)	< 0,001	3,08 (2,58–3,67)	< 0,001
Idioma em que o artigo foi escrito	Outro idioma	1		NA	
	Inglês	113,96 (36,37–357,07)	< 0,001	NA	NA
Acesso aberto ao artigo	Não	1		1	
	Sim	0,58 (0,49–0,69)	< 0,001	0,89 (0,75–1,06)	0,193
Atributos da pesquisa					
Tipo de publicação	Carta ou artigo original	1		1	
	Revisão	0,99 (0,69–1,43)	0,967	4,18 (3,09–5,64)	< 0,001
Se o artigo contém declaração de financiamento	Não	1		1	
	Sim	5,88 (4,87–7,08)	< 0,001	6,57 (5,44–7,93)	< 0,001
Se a nutrigenômica foi abordada no artigo	Não	1		1	
	Sim	1,23 (0,90–1,67)	0,189	0,69 (0,49–0,98)	0,036
Atributos do periódico					
Ano de criação do periódico	> mediana	1		1	
	≤ mediana	1,49 (1,26–1,77)	< 0,001	2,11 (1,78–2,51)	< 0,001
Quartil do periódico	4Q/3Q	1		1	
	2Q	4,52 (3,58–5,71)	< 0,001	7,16 (5,26–9,74)	< 0,001
	1Q	4,82 (3,87–6,01)	< 0,001	25,06 (19,01–33,03)	< 0,001
Número de categorias da WoS às quais o periódico pertence	1 categoria	1		1	
	Duas categorias	1,32 (1,09–1,61)	0,004	1,21 (1,01–1,47)	0,044
	3 ou mais categorias	3,59 (2,80–4,59)	< 0,001	2,46 (1,91–3,17)	< 0,001
Atributos do autor					
Quantidade de autores	-	1,22 (1,18–1,26)	< 0,001	1,25 (1,21–1,30)	< 0,001
Quantidade de países em colaboração	-	1,71 (1,45–2,02)	< 0,001	1,92 (1,63–2,26)	< 0,001
Número de rendas representadas pelos países das afiliações	1	1		1	
	2 ou mais	1,89 (1,35–2,65)	< 0,001	1,376 (0,97–1,96)	0,077
Renda do país do primeiro autor	Baixa ou média baixa	1		1	
	Média alta	0,63 (0,41–0,95)	0,027	0,57 (0,35–0,92)	0,021
	Alta	0,89 (0,61–1,29)	0,531	1,63 (1,07–2,49)	0,022
Engajamento dos autores	Nenhum autor tem RI ou OI	1		1	
	Ao menos 1 autor com RI ou OI	2,10 (1,66–2,66)	< 0,001	2,88 (2,28–3,64)	< 0,001
	Ao menos 1 autor com RI e OI	2,80 (2,24–3,49)	< 0,001	3,83 (3,07–4,78)	< 0,001
Se há alguma declaração sobre conflito de interesses	Não	1		1	
	Sim	0,85 (0,72–1,01)	0,062	1,06 (0,89–1,25)	0,529

OR: odds ratio; IC95%: intervalo de confiança de 95%; NA: não avaliável; RI: ResearcherID; OI: Orcid.

Tabela 4. Regressão logística multinomial múltipla para a condição de citação dos artigos, 2009–2011.

Variável	Categoria	Moderada		Alta	
		OR (IC95%)	Valor de p	OR (IC95%)	Valor de p
Atributos do artigo					
Tipo de título	Assunto ou pergunta	1		1	
	Conclusão	1,77 (1,31–2,40)	< 0,001	1,24 (0,88–1,75)	0,219
Tamanho do resumo	≤ Mediana	1		1	
	> Mediana	2,77 (2,17–3,54)	< 0,001	2,85 (2,15–3,79)	< 0,001
Quantidade de referências citadas no artigo	≤ Mediana	1		1	
	> Mediana	2,72 (2,11–3,51)	< 0,001	5,95 (4,46–7,94)	< 0,001
Tamanho do artigo	≤ Mediana	1		1	
	> Mediana	1,18 (0,91–1,53)	0,201	2,03 (1,52–2,72)	< 0,001
Acesso aberto ao artigo	Não	1		1	
	Sim	0,51 (0,40–0,66)	< 0,001	0,85 (0,63–1,13)	0,260
Atributos da pesquisa					
Tipo de publicação	Carta ou artigo original	1		1	
	Revisão	2,38 (1,46–3,88)	< 0,001	7,87 (4,79–12,93)	< 0,001
Se o artigo contém declaração de financiamento	Não	1		1	
	Sim	3,20 (2,55–4,03)	< 0,001	3,09 (2,37–4,02)	< 0,001
Se a genômica nutricional foi abordada no artigo	Não	1		1	
	Sim	0,86 (0,57–1,29)	0,459	0,36 (0,22–0,59)	< 0,001
Atributos do periódico					
Quartil do periódico	4Q/3Q	1		1	
	2Q	3,99 (3,00–5,30)	< 0,001	6,62 (4,55–9,63)	< 0,001
	1Q	5,23 (3,92–6,99)	< 0,001	39,41 (27,38–56,73)	< 0,001
Número de categorias da WoS às quais o periódico pertence	1	1		1	
	2	1,32 (1,02–1,71)	0,037	1,52 (1,13–2,04)	0,005
	3 ou +	3,52 (2,52–4,91)	< 0,001	3,89 (2,63–5,76)	< 0,001
Atributos do autor					
Quantidade de autores		1,09 (1,04–1,14)	0,001	1,11 (1,06–1,17)	< 0,001
Engajamento dos autores	Nenhum autor tem RI ou OI	1		1	
	Ao menos 1 autor com RI ou OI	1,27 (0,95–1,70)	0,112	1,63 (1,17–2,25)	0,004
	Ao menos 1 autor com RI e OI	1,47 (1,11–1,96)	0,007	1,73 (1,26–2,36)	0,001
Quantidade de países em colaboração		1,22 (1,02–1,46)	0,027	1,26 (1,04–1,51)	0,016

OR: *odds ratio*; IC95%: intervalo de confiança de 95%; RI: ResearcherID; OI: Orcid.

Nota: Valor de pseudo R² do modelo: Nagelkerke = 0,537.

A Tabela 5 contém os coeficientes de regressão para o modelo de regressão múltipla, para aplicação na equação de probabilidade. Esses valores permitem que se estime a probabilidade de um dado artigo de ser moderada ou altamente citado.

Equação de probabilidade:

$$P(\text{Moderadamente ou altamente citado}) = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_n x_n)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_n x_n)}$$

Tabela 5. Coeficientes de regressão para o modelo de regressão múltipla.

Variável	Categoria	Moderada	Alta
		β	β
Tipo de título	Assunto ou pergunta		
	Conclusão	0,572	0,216
Tamanho do resumo	$\leq$ Mediana		
	$>$ Mediana	1,019	1,048
Quantidade de referências citadas no artigo	$\leq$ Mediana		
	$>$ Mediana	1,000	1,784
Tamanho do artigo	$\leq$ Mediana		
	$>$ Mediana	0,169	0,709
Acesso aberto ao artigo	Não		
	Sim	-0,670	-0,169
Tipo de publicação	Carta ou artigo original		
	Revisão	0,867	2,063
Se o artigo contém declaração de financiamento	Não		
	Sim	1,164	1,127
Se a genômica nutricional foi abordada no artigo	Não		
	Sim	-0,153	-1,016
Quartil do periódico	4Q/3Q		
	2Q	1,383	1,890
	1Q	1,655	3,674
Número de categorias da WoS às quais o periódico pertence	1		
	2	0,276	0,419
	3 ou +	1,258	1,359
Engajamento dos autores	Nenhum autor tem RI ou OI		
	Ao menos 1 autor com RI ou OI	0,239	0,485
	Ao menos 1 autor com RI e OI	0,388	0,549
Quantidade de autores		0,083	0,107
Quantidade de países em colaboração		0,200	0,227

β : Coeficiente de regressão; RI: ResearcherID; OI: Orcid.

Na Tabela 6, são apresentados os percentuais de acerto do modelo para cada condição de citação. O modelo consegue prever corretamente a condição de citação em 65,7% dos casos.

Tabela 6. Percentual de acerto do modelo segundo a condição de citação.

Observado	Previsto			Porcentagem correta
	Baixa	Moderada	Alta	
Baixa	688	174	47	75,7%
Moderada	193	575	294	54,1%
Alta	62	273	733	68,6%
Porcentagem global	31,0%	33,6%	35,3%	65,7%

No Quadro 3, estão apresentadas as variáveis que permaneceram no modelo múltiplo. As variáveis tamanho do título, ano de criação do periódico, condição econômica do país dos autores e presença de declaração de conflito de interesse perderam a significância estatística após ajuste por múltiplos fatores e foram excluídos do modelo final.

Quadro 3. Variáveis independentes analisadas e variáveis mantidas no modelo múltiplo.

Dimensão	Variável	Categorias	Presença no modelo final
Atributos do artigo	Tamanho do título	Curto; médio; longo	Não
	Tipo de título	Assunto ou pergunta; conclusão	Sim
	Tamanho do resumo	≤ mediana; > mediana	Sim
	Quantidade de referências citadas no artigo	≤ mediana; > mediana	Sim
	Tamanho do artigo	≤ mediana; > mediana	Sim
	Idioma em que o artigo foi escrito	Outro idioma; inglês	Não ^a
	Acesso aberto ao artigo	Não; sim	Sim
Atributos da pesquisa	Tipo de publicação	Carta ou artigo original; revisão	Sim
	Se o artigo contém declaração de financiamento	Não; sim	Sim
	Se a genômica nutricional foi abordada	Não; sim	Sim
Atributos do periódico	Ano de criação do periódico	≤ mediana; > mediana	Não
	Impacto do periódico	4Q ou 3Q; 2Q; 1Q	Sim
	Nº de categorias da WoS às quais o periódico pertence	1; 2; ≥ 3	Sim
Atributos do autor	Quantidade de autores	Variável contínua	Sim
	Quantidade de países em colaboração	Variável contínua	Sim
	Diversidade de níveis de rendimento dos países dos autores	1; 2 ou mais	Não
	Renda do país do primeiro autor	Baixa ou média baixa; média alta; alta	Não
	Engajamento dos autores	Nenhum autor com RI ou OI; ao menos 1 autor com RI ou OI; ao menos 1 autor com RI e OI	Sim
	Se há alguma declaração sobre conflito de interesses	Não; sim	Não

4Q: 4º quartil; 3Q: terceiro quartil; 2Q: segundo quartil; 1Q: primeiro quartil; WoS: Web of Science; RI: ResearcherID; OI: Orcid.

^a Não incluída no modelo múltiplo devido à frequência zero em um dos grupos de citação.

4.1 VALIDAÇÃO DA WOS

Quanto à validação do campo financiamento da WoS, 56,5% dos registros de base informam financiamento de pesquisa (Tabela 7). Na revisão manual, 60,5% dos artigos informam financiamento (Tabela 7).

Tabela 7. Distribuição de artigos segundo financiamento, da categoria “Nutrition & Dietetics” da Web of Science, 2008–2010.

	WoS		Revisão manual	
	n	%	n	%
Com financiamento	1.798	56,5	1.927	60,5
Sem financiamento	1.385	43,5	1.256	39,5
Total	3.183	100	3.183	100

Na comparação da WoS com a revisão manual, observou-se que 45 registros da WoS informavam financiamento sem que essa condição fosse identificada na revisão manual. Ainda, 174 registros da WoS não informavam financiamento, mas a revisão manual constatou presença de financiamento no texto do artigo.

A análise da acurácia resultou em 91% de sensibilidade e 96,4% de especificidade para a classificação feita pela base WoS (Tabela 8). Portanto, quando a WoS identifica presença de financiamento, há 91% de chances de o artigo de fato ter mencionado financiamento (verdadeiros positivos); se não identifica, há 96,4% de chances de o artigo não ter mencionado financiamento (verdadeiros negativos). A área sob a curva ROC foi de 93,7%. A estatística Kappa foi de 85,8%. Esses dados representam acurácia e concordância elevadas.

Tabela 8. Distribuição da informação de presença de financiamento nos registros Web of Science (WoS) e na verificação manual do artigo.

		Revisão manual		Total
		Sim	Não	
WoS	Sim	1.753	45	1.798
	Não	174	1.211	1.385
Total		1.927	1.256	3.183

Chama atenção que, na vasta maioria dos artigos, o financiamento não se encontra no local específico destinado a essa informação (Financiamento). Quanto ao campo do artigo onde constava a informação sobre financiamento, houve diferença significativa entre a WoS e a revisão manual. Nos artigos cujo financiamento constava apenas na revisão manual, identificamos maiores proporções de financiamento informado em notas de rodapé e no campo

conflito de interesses do que naqueles cujo financiamento constava tanto na revisão manual quanto na WoS (Tabela 9).

Tabela 9. Distribuição dos locais do artigo onde constaram o financiamento na revisão manual.

Campo do artigo	Verdadeiros +	Falsos -	Total	p^b
Agradecimentos	1.271 (72%)	108 (62%)	1.379	0,001
Nota de rodapé	408 (23%)	54 (31%)	462	
Financiamento	53 (3%)	4 (2%)	57	
Conflito de interesse	25 (1%)	8 (5%)	33	
Total	1.757^a	174	1.931	

Verdadeiros +: financiamento identificado tanto pela WoS quanto pela revisão manual; falsos -: financiamento identificado pela revisão manual, mas não pela WoS.

^a O total é superior aos 1.753 verdadeiros positivos porque quatro artigos mencionaram financiamento em mais de um campo.

^b Teste qui-quadrado de Pearson

Identificamos que, considerando os falsos negativos (sem financiamento na WoS e com financiamento na revisão manual), a WoS contava com maior proporção de artigos escritos em outro idioma que não o inglês. Também foi identificada ausência de falsos positivos em artigos em outro idioma que não o inglês (Tabela 10).

Tabela 10. Distribuição da acurácia da informação sobre financiamento segundo idioma.

Idioma	Verdadeiros + e -	Falsos +	Falsos -	Total	p ^a
Inglês	2.792 (93,4%)	45 (1,5%)	154 (5,2%)	2.991	0,002
Outro	172 (89,6%)	0 (0%)	20 (10,4%)	192	
Total	2.964	45	174	3.183	

Verdadeiros + e -: artigos em que houve concordância quanto à presença ou ausência de financiamento entre a WoS e a revisão manual; falsos +: financiamento identificado pela WoS mas não pela revisão manual; falsos -: financiamento identificado pela revisão manual mas não pela WoS.

^a Teste qui-quadrado de Pearson

5 DISCUSSÃO

5.1 FATORES ASSOCIADOS A CITAÇÕES

Neste estudo, identificamos atributos do artigo, da pesquisa, do periódico e do autor que aumentam as chances de um artigo científico ser citado, em amostra robusta com poder estatístico para avaliar número expressivo de fatores associados e identificar efeitos ainda que pequenos. Ter declaração de financiamento, ser publicado em periódico de alto impacto e ser publicado em periódico multidisciplinar aumentam em mais de três vezes as chances de um artigo ser

moderadamente ou altamente citado. Além dessas características, as chances de o artigo ser altamente citado são maiores ($OR > 3,0$) também quando há maior número de referências e quando o artigo é de revisão.

5.1.1 Atributos do Artigo

Resumos mais longos e maior número de referências aumentam as chances de o artigo receber moderado ou alto número de citações e artigos mais longos têm mais chances de ser altamente citados. Artigos com títulos do tipo conclusão apresentam mais chances de ser moderadamente citados, enquanto o acesso aberto ao artigo diminui essas chances. Vale ressaltar que, apesar de o idioma de publicação não ter entrado no modelo final, essa variável está associada a um maior número de citações. Com efeito, a associação é tão forte que não houve outro idioma no grupo de alta citação, motivo pelo qual a variável precisou ser excluída da análise múltipla.

Uma vez que, para ser citado, um artigo antes precisa ser lido, sua recuperação nas buscas feitas pelos potenciais citantes é o primeiro passo para a conquista de grande número de citações. Por um lado, quanto menor o resumo, mais pesquisadores podem ser atraídos para leitura do texto completo e potencial uso, e consequentemente citação ao artigo. Por outro lado, quanto mais palavras tiverem, mais facilmente os resumos podem ser recuperados nas buscas bibliográficas. Dessa forma, hipotetizamos como mecanismo que um artigo cujo resumo é longo tem vantagens por fornecer abundância de termos passíveis de compor as estratégias de buscas e, consequentemente, ser recuperado e seu resumo ser lido, dando assim a oportunidade de o leitor decidir por ler o artigo na íntegra e eventualmente citá-lo em seu trabalho em andamento. Em

estudo recente, Boyd et al. (2019) compararam os artigos mais citados com os menos citados publicados entre 1997 e 2007 por periódicos da área de urologia. Os autores identificaram que os mais citados continham resumos mais longos, de modo que nossos resultados corroboram, com amostra consideravelmente maior, esses achados. Letchford et al. (2016) observaram que resumos simples (mais curtos e compostos por palavras mais comuns) atraíam mais citações. A hipótese dos autores era de que resumos longos e com palavras incomuns demandariam, para sua compreensão, maior carga cognitiva por parte do leitor e que isso poderia fazer o artigo não receber tantas citações. Apesar do grande tamanho da amostra, o estudo incluiu apenas os artigos mais citados entre 1999 e 2008, de modo que não houve comparação dos mais citados com os menos citados.

Corroborando o estudo de Boyd et al. (2019) na área de urologia, identificamos que os artigos mais citados apresentam maior quantidade de referências e são mais longos. Quanto às referências, maior número de referências aumenta em mais de duas vezes as chances de ser moderadamente citado e em quase seis vezes as chances de ser altamente citado. Temos duas hipóteses para esse resultado. A primeira é que, quanto mais referências citadas, maior a visibilidade do artigo em questão, dado que os autores citados tendem a tomar conhecimento de quais artigos fizeram referência a seus trabalhos. Isso é uma realidade, sobretudo atualmente, em que as tecnologias digitais facilitam não apenas a contagem de citações, mas também a identificação dos citantes. A segunda hipótese é que a referência a diversos trabalhos pode ser um indicador de interdisciplinaridade (Wang e Schneider, 2020), o que aumentaria a audiência desse artigo. A essas explicações, podemos adicionar a observação de Yu et al. (2014) de que o maior número de referências provavelmente é reflexo de um conhecimento mais amplo do autor sobre o seu campo de pesquisa, o que consequentemente melhora a qualidade da pesquisa e seu potencial para ser citada. Literatura recente mostra que o número de referências citadas nos artigos tem

aumentado nas diversas áreas do conhecimento (Nicolaisen e Frandsen, 2020) e que maior número de referências se associa à alta citação (Mammola et al., 2021). Quanto ao tamanho do artigo, Ahmed et al., em 2016, identificaram que artigos com 11–20 páginas eram mais citados do que aqueles com até 10 páginas. Semelhantemente, Boyd et al. (2019) identificaram que artigos mais longos, operacionalizados pelo maior número de palavras, recebiam mais citações. No presente estudo, utilizando métodos estatísticos mais complexos, confirmamos que artigos maiores são 2 vezes mais citados do que aqueles menores. Artigos mais longos podem representar textos mais informativos ou com mais dados, o que fornece aos leitores mais conteúdo suscetível a ser citado. De fato, um artigo não apenas provê uma conclusão, mas também dados primários, procedimentos metodológicos, justificativas e hipóteses, entre outras informações que podem ser usadas para embasar novos estudos, gerando citações àquele trabalho.

Enquanto os títulos que descrevem o tema de pesquisa podem fazer com que sua abrangência torne as buscas bibliográficas menos eficientes, aqueles que informam a hipótese, ou seja, a pergunta de pesquisa, direcionam o leitor durante a busca de trabalhos relevantes. Entretanto, apesar de títulos compostos por perguntas terem o potencial de atrair maior atenção do leitor e obterem mais *downloads*, essa atratividade não se mostra suficiente para atrair mais citações (Jamali e Nikzad, 2011). Nossa hipótese é de que títulos contendo a conclusão do estudo torna a busca bibliográfica ainda mais eficiente, pois são mais informativos e o leitor reconhece com maior facilidade a informação que procura. Jamali e Nikzad (2011) também avaliaram se o tipo de título influencia o número de citações, mas observaram que o tipo declarativo, analogamente ao tipo conclusão no presente estudo, não diferiu significativamente do tipo descritivo (análogo à categoria assunto usada no presente estudo). Essa diferença entre os

achados pode ser decorrente das janelas de citação usadas; como usamos uma janela de cinco anos, dois anos a mais que no estudo de Jamali e Nikzad (2011), nossa amostra pode ter tido tempo suficiente para alcançar resultados distintos entre as citações nas categorias de títulos.

O primeiro estudo que mostrou correlação entre acesso aberto e maior número de citações foi conduzido por Steve Lawrence em 2001 e, desde então, algumas pesquisas têm sido publicadas sobre o tema, porém com resultados conflitantes dado que algumas encontram associação (Eysenbach, 2006; Harnad e Brody, 2004; Gargouri et al., 2010; Hua et al., 2017; Piwowar et al., 2018; Breugelmans et al., 2018; Patel et al., 2019) e outras não (Davis, 2011; Björk e Solomon, 2012; Hua et al., 2016; Dorta-González et al., 2017; Dorta-González e Santana-Jiménez, 2018). No nosso estudo, artigos de acesso aberto apresentaram menores chances de ser moderadamente citados. Aqui, entendemos que não é o fato em si de o artigo ser de acesso aberto que influencia negativamente, mas as características desses artigos quando comparadas àqueles cujo acesso se dá de forma paga.

A princípio, o acesso aberto permite que mais pessoas tenham contato com o artigo, o que aumenta o potencial de este ser lido e citado. Entretanto, se esse acesso não for acompanhado por reconhecimento dos pesquisadores da área, ele poderá não ter o efeito esperado. Eysenbach, em 2006, acompanhou 212 artigos de acesso aberto e 1.280 de acesso fechado, publicados em um periódico científico de alto prestígio, por 10–16 meses após publicação. O autor identificou que aqueles de acesso aberto atraíram mais citações desde logo após a publicação, aumentando para três vezes as chances de serem citados no final do período estudado. Apesar de incluir uma grande amostra e analisar os dados por meio de regressão logística com ajustes para potenciais fatores de confusão, o autor avaliou um periódico de alto impacto, o que pode explicar a diferença nos achados entre seu estudo e a presente pesquisa. Outro fator que pode ter impactado

esses resultados, assim como citado na variável tipo de título, é a janela de citações considerada em ambos os estudos. Craig et al. (2007), em revisão crítica da literatura acerca do tema, argumentam que janelas de tempo curtas podem superestimar o impacto do acesso aberto dado que amplificam o efeito entre artigos publicados com poucos meses de diferença. Sotudeh e Estakhr (2018), entretanto, observaram que a vantagem do acesso aberto se sustenta mesmo considerando janelas de tempo maiores. Em estudo mais recente, Sotudeh (2019) investigou se o acesso aberto apresentaria vantagem mesmo para artigos com temas semelhantes e encontrou resultado positivo. Porém, Hua et al. encontraram diferentes resultados para a área de saúde bucal (Hua et al., 2016) e de oncologia (Hua et al., 2017), sugerindo que o acesso aberto pode ter impacto heterogêneo sobre as diversas áreas.

Estudo recente de Basson et al. (2021) analisou dados bibliométricos de todos os artigos e revisões indexados na WoS e publicados entre 2013 e 2015. Os autores observaram que a vantagem de citação foi encontrada para uma única variável, a saber, se o artigo em acesso aberto foi ou não citado (e não se ele recebeu mais citações que aqueles em acesso restrito, por exemplo). Além disso, os artigos em acesso aberto apresentaram uma vantagem de citação apenas em poucas áreas temáticas (Basson et al., 2021). Similarmente, estudo que testou se artigos em acesso aberto apresentavam maior atividade altmétrica mostrou diferenças entre disciplinas (Holmberg et al., 2020). Os autores identificaram que o acesso aberto resultou em vantagem nas áreas de psicologia, veterinária e geografia social e econômica, enquanto houve uma desvantagem nas áreas de medicina e ciências da saúde. Ambos os estudos embasam nossa hipótese de que os resultados aqui encontrados podem estar sob influência da área analisada.

Além disso, apesar de o acesso a determinados artigos ser condicional mediante pagamento, os leitores e potenciais citantes podem ter acesso a eles por meio de suas instituições

de origem, que pagam assinaturas para que seus pesquisadores tenham acesso gratuito a bases de dados indexadoras ou aos periódicos mais relevantes para suas áreas. Assim, por um lado, o artigo em acesso aberto pode ser aquele com menor qualidade metodológica ou grau de inovação e que, portanto, não foi aceito por periódicos relevantes que cobram taxas de publicação. Por outro lado, o artigo cujo acesso é pago pode ser disponibilizado gratuitamente para seus potenciais citantes, o que aumentaria suas chances de ser citado e, ainda assim, seria contabilizado entre os de acesso fechado nas pesquisas sobre o tema.

Neste estudo, portanto, os resultados mostram que o acesso aberto por si só não leva a um maior número de citações. Ressaltamos, porém, que esse dado deve ser interpretado à luz de dois aspectos importantes. O primeiro refere-se ao fato de que isso não significa que devemos abrir mão do acesso aberto, mas estarmos cientes de que não basta um artigo estar em acesso aberto para que ele tenha grande impacto na comunidade científica. Além do acesso aberto, outras características devem estar presentes para que o potencial de impacto seja alcançado. Este estudo mostra outros fatores nos quais os autores e editores podem trabalhar. O segundo aspecto diz respeito à necessidade de compreender que ser citado não é a meta final de um artigo. Não podemos considerar que a importância do acesso aberto se resume apenas ao seu potencial de aumentar as citações recebidas. De fato, nossos resultados mostram que este nem deveria ser um argumento forte a favor do acesso aberto. A relevância do acesso aberto se dá por inúmeros outros motivos.

Apesar de o acesso a artigos publicados em periódicos brasileiros ser gratuito na vasta maioria das vezes, o restante do mundo enfrenta um panorama bastante distinto, sobretudo em algumas áreas, como a medicina (Alves, 2019). Atualmente, governos e fundações pagam quando financiam as pesquisas e quando precisam acessar os artigos derivados dessas pesquisas. Essa

duplicação do custo se torna ainda mais incômoda diante de dois aspectos: os periódicos não pagam pela condução de pesquisas, nem pela escrita dos artigos, nem pela avaliação destes por parte dos revisores; e a margem de lucro das grandes editoras é exorbitante (Camargo Jr, 2012). Em um cenário de crescente digitalização, espera-se diminuição dos custos e, conseqüentemente, dos valores cobrados para publicação dos achados científicos. Mas o que ocorreu ao longo do tempo foi exatamente o oposto, com expressivo aumento dos preços para acesso aos artigos científicos, o que gerou indignação na comunidade científica e sociedade em geral.

Esse cenário promoveu um movimento a favor do acesso aberto. Lançado em 2018 pela Science Europe, um consórcio de agências nacionais de pesquisa e financiadores de diversos países europeus, o cOAlition S ou Plano S é uma iniciativa de publicação científica de acesso aberto. O plano determinou que, a partir de janeiro de 2020, todas as pesquisas financiadas com dinheiro público fossem publicadas em acesso aberto. Sua implementação foi prorrogada para janeiro de 2021 (Marques, 2021) e, atualmente, passa por um processo de transição, em que o objetivo original é mantido (artigos financiados com verba pública estarem acessíveis ao leitor de forma gratuita), mas com flexibilização dos métodos para atingi-lo (a publicação pode ser feita em periódicos híbridos, por exemplo).

Quanto à variável idioma, esta não permaneceu no modelo final devido à ausência de artigos em outro idioma dentre o grupo de artigos altamente citados. Entretanto, houve associação univariada entre idioma e maiores chances de o artigo ser moderada ou altamente citado, apesar de a magnitude da associação não poder ser mensurada no modelo ajustado. Inúmeros estudos têm mostrado que o idioma em que um artigo é escrito influencia o número de citações recebidas (Van Dalen e Henkens, 2001; Bornmann et al., 2012; Paci et al., 2015; Di Bitetti e Ferreras, 2017). Diekhoff et al. (2013) analisaram 168 periódicos da área médica indexados na WoS e

publicados em mais de um idioma; os autores identificaram que, quanto maior a proporção de artigos publicados em inglês, maior era o reconhecimento internacional, medido pelo menor número de autocitações e maior fator de impacto.

Para avaliar o impacto da linguagem de publicação na frequência de citação, Poomkottayil et al. (2011) analisaram artigos de cirurgia oral indexados na PubMed que foram publicados por faculdades de odontologia suíças entre 2002 e 2007. A frequência de citação foi extraída do ISI e do Google Scholar. Empregando modelo de regressão logística univariada e multivariada, ajustado para fator de impacto do periódico, número de autores e tema de pesquisa, os autores identificaram que, considerando os dados fornecidos pelo ISI, artigos publicados em inglês apresentaram seis vezes maior chance de ser citado do que aqueles publicados em alemão ou francês. Di Bitetti e Ferreras (2017) compararam artigos em inglês e em outro idioma que não o inglês, indexados na base de dados Scopus e usando modelo de regressão logística ajustado pelo periódico, ano de publicação e tamanho do artigo, também encontrando maior citação entre os artigos publicados em inglês. Zhao (2015) identificou que revisões publicadas em outro idioma que não o inglês aumentam a taxa de não citação dos artigos. A língua atual da ciência é o inglês e, apesar da aparente injustiça com autores cujo idioma nativo difere do inglês, ter uma língua franca permite que pesquisadores do mundo todo se comuniquem em uma linguagem comum. Isso simplifica a disseminação do conhecimento, visto que é mais fácil dominar apenas uma língua estrangeira do que duas ou mais.

Dos atributos do artigo analisados, o tamanho do título perdeu sua associação com as citações no modelo múltiplo. Segundo estudo de Letchford et al. (2015), no qual os autores analisaram se o estilo em que o artigo foi escrito influencia as citações recebidas, quanto menor o título, mais fácil de ser entendido, o que atrai mais pesquisadores para sua leitura e potencial uso,

aumentando consequentemente seu número de citações. Estudos anteriores apresentam evidências conflitantes, sugerindo que o comprimento do título de um artigo não está associado ao seu impacto (Jamali e Nikzad, 2011), que títulos mais longos podem estar associados a um maior número de citações (Jacques e Sabire, 2010; Habibzadeh e Yadollahie, 2010) ou mesmo que artigos com títulos mais curtos recebem mais citações (Letchford et al., 2015). Nossos resultados sugerem que o tamanho pode ser um fator de confusão para o tipo de título. Dessa forma, um título maior não atrairia mais citações, mas estaria associado ao tipo conclusão, que, este sim, aumenta a chance do artigo de ser mais citado. De fato, uma conclusão é, na verdade, uma frase completa, contendo sujeito, verbo e complemento, enquanto o assunto é um fragmento de frase; logo, é esperado que títulos do tipo conclusão contenham mais palavras que os do tipo assunto. Mabe e Amin (2002) analisaram pesquisadores quanto a diferenças quando estão no papel de autor e de leitor. O estudo identificou que, como leitores, os pesquisadores querem ler menos; de fato, eles liam em média 1.142 títulos por ano, dos quais 18% os levavam a ler os resumos e apenas 8,5%, os textos completos. Com isso, podemos concluir que o objetivo de ler menos pode ser alcançado mais efetivamente por meio de títulos direcionadores, que poupam leituras de resumos e textos completos não relevantes, do que por meio de títulos curtos, porém pouco informativos.

5.1.2 Atributos da Pesquisa

Estudos de revisão aumentam a chance de o artigo ser altamente citado. Quando há declaração de recebimento de financiamento, os artigos apresentam mais chance de ser moderada ou altamente citados. Já o fato de o artigo abordar um *hot topic* mostrou-se independentemente associado com menor chance de receber citações.

Quanto ao tipo de publicação, é esperado que as revisões sejam mais citadas, considerando-se a necessidade de embasamento em estudos de revisão para proposição de ensaios clínicos e de conclusão com maior nível de evidência que um artigo original. Adicionalmente, existem citações equivocadas de referências citadas em revisões (Teixeira et al., 2013), quando a informação referenciada não é uma conclusão da revisão, mas sim de um dos artigos primários reportados na revisão; ou seja, quando ocorre citação de citação sem que o autor citante especifique que se trata desse tipo de citação. Estudo na área de psicologia identificou que revisões recebem mais citações do que artigos empíricos (Krampen et al., 2015). Os autores sugerem que essa diferença se deve ao fato de que estudos primários motivam poucas pesquisas subsequentes que, apesar de os citarem, logo os substituem como os novos artigos a serem citados. Já as revisões, defendem os autores, são publicações que descrevem, estruturam, integram e avaliam pesquisas quanto a um domínio mais abrangente, de modo que resulta em maior sucesso de citação, dado que muitas publicações posteriores podem decorrer das conclusões feitas nesses artigos não empíricos. No presente estudo, artigos de revisão tiveram mais de duas vezes as chances de ser moderadamente citados e quase oito vezes as chances de ser altamente citados, quando comparados com artigos de pesquisa original e cartas.

No presente estudo, identificamos que, se algum financiamento for declarado, o artigo tem três vezes mais chance de ser moderada ou altamente citado. Segundo estudo de Amara et al. (2015), projetos de pesquisa que recebem maior nível de financiamento e autores que recebem subsídios recebem mais citações do que artigos não financiados. Essa relação foi rejeitada pelos estudos de Antoniou et al. (2015) na área cardiovascular e de Willis et al. (2011) na área de urologia; neste último, declarar financiamento não mostrou significância estatística nem na análise univariada. Na área médica, artigos que mencionam ter recebido financiamento da

indústria recebem mais citações (OR = 25,7; IC95% 8,5–42,9) se seus resultados favorecem a indústria em relação a artigos que não mencionam ter recebido financiamento (Kulkarni et al., 2007). Na área de nanotecnologia, Wang e Shapira (2015) identificaram que publicações com declaração de financiamento têm maior probabilidade tanto de serem publicadas em periódicos de alto impacto, como de receberem mais citações. Nossa hipótese para os resultados referentes a essa variável é que pesquisas que recebem financiamento passaram por um crivo a mais quanto a características atrativas à ciência, ou seja, presença de financiamento pode ser um indicador de qualidade e aceitabilidade pela comunidade científica, o que aumenta, por sua vez, o potencial de citação dos artigos. Além disso, por ter financiamento, pode ter havido menos limitações para a condução da pesquisa e da publicação. Dessa forma, por exemplo, dispõe-se de recursos para pagamento de *article processing charges* (APC ou taxas de publicação) mais caras em periódicos classificados no Q1, o que atrairia mais citações.

Neste estudo, analisamos se o fato de o artigo abordar um *hot topic*, no caso a genômica nutricional, atrairia mais citações. Quando ajustado pelas variáveis do artigo, do periódico e dos autores, essa variável não perde a associação, mas inverte o sentido da associação, ou seja, o fato de o artigo abordar a genômica nutricional por si só diminui as citações recebidas. Wei et al. (2013) concluíram que os pesquisadores publicam mais artigos sobre *hot topics*. Nosso resultado sugere que esses artigos atraem mais citações devido a outros atributos, e não ao tema em si; por exemplo, seus autores escrevem de uma forma, conduzem pesquisas e publicam em periódicos com características que levam a maior número de citações. Podemos citar, entre os atributos, a variável financiamento. Pesquisas envolvendo genômica nutricional tendem a envolver análises laboratoriais que demandam recursos físicos (equipamentos, reagentes, kits comerciais) dificilmente disponíveis sem que a pesquisa seja financiada. Assim, receber financiamento pode

ser uma das características dos artigos que abordam a genômica nutricional que aumentam suas chances de ser citado. O número de autores também pode influenciar as citações nesses casos, uma vez que análises complexas podem requerer uma equipe maior ou mesmo promover a colaboração entre equipes de laboratórios distintos.

5.1.3 Atributos do Periódico

Se o artigo for publicado em periódico pertencente ao quartil de melhor classificação ou em periódico abrangente quanto à temática, maiores são as chances de ser citado de forma moderada ou alta.

O resultado relativo ao quartil do periódico corrobora os diversos estudos que observam maior citação a artigos publicados em periódicos mais bem avaliados pela comunidade científica. De certo, uma vez que o quartil, assim como o fator de impacto e muitos outros indicadores de qualidade, é formulado com base em citações, espera-se que os artigos publicados por periódicos do Q1 recebam mais citações com o tempo. Porém, essa relação pode não ser tão precisa quanto o previsto (Lozano et al., 2012). Com efeito, o fator de impacto de um periódico pode ser positivamente influenciado por alguns artigos altamente citados, mas a distribuição de citações não é paramétrica, ou seja, não é homogênea para todos os artigos publicados nesse periódico, de modo que o alto número de citações recebidas por um artigo não implica necessariamente na mesma proporção de citações do outro. Adicionalmente, o fator de impacto calculado pela Clarivate Analytics considerada o número de citações recebidas em determinado ano dividido pelo número de itens citáveis publicados nos dois anos anteriores. Esse cálculo superestima o

número médio de citações recebidas por artigo, sobretudo em periódicos que publicam muitos itens considerados pela Clarivate Analytics como não citáveis. Larivière e Gingras (2010a), corrigindo a assimetria entre o denominador e numerador no fator de impacto, identificaram que o valor intrínseco de um artigo não é a única razão pela qual ele é citado ou não; os autores concluem que há um efeito Mateus associado aos periódicos e isso dá ao artigo publicado em periódicos de maior impacto um valor agregado além de sua qualidade intrínseca. Entretanto, esses mesmos autores participaram como coautores de um outro estudo (Lozano et al., 2012), e identificaram que a associação entre fator de impacto do periódico e a taxa de citações de seus artigos vem se enfraquecendo desde 1990, como efeito da era digital na ciência.

Quanto ao resultado relacionado à abrangência de temáticas do periódico, com efeito, quanto mais abrangente, maior o público citante, pois pesquisadores de diferentes áreas terão contato com esse material. Além disso, os leitores de periódicos abrangentes têm maior potencial de também trabalharem em linhas de pesquisa mais inter- ou multidisciplinares, o que aumentaria o interesse desses pelo assunto tratado no artigo publicado. Larivière e Gingras (2010b) mostraram que, embora para todas as disciplinas juntas não haja correlação clara entre o nível de interdisciplinaridade dos artigos e sua taxa de citação, para algumas disciplinas um maior nível de interdisciplinaridade está relacionado a taxas de citação mais altas, enquanto para outras, as citações diminuem à medida que a interdisciplinaridade aumenta. Os autores sugerem que possa haver um nível ótimo de interdisciplinaridade, que fica entre uma interdisciplinaridade tão limitada que não gera alto impacto e uma dispersão grande o suficiente ao ponto de o estudo não encontrar seu nicho. Xie et al. (2019) também não encontraram associação entre interdisciplinaridade e citação. Essas diferenças podem se dever às diferentes formas de operacionalização da variável. Neste estudo, a interdisciplinaridade foi classificada segundo o

número de categorias da WoS às quais os periódicos pertencem e, com efeito, as chances de um artigo ser moderada ou altamente citado aumenta à medida que o número de categorias aumenta.

5.1.4 Atributos do Autor

Autores mais engajados aumentam a chance de o artigo ser moderada ou altamente citado. Quanto maior o número de países e maior o número de autores em colaboração, maiores as chances de o artigo ser moderada ou altamente citado.

Neste estudo, consideramos a presença de registro no ResearcherID ou Orcid, ou ambos, como proxy do engajamento dos autores. As chances de ser citado aumentam à medida que o engajamento dos autores também aumenta; essa relação é mais clara entre os artigos altamente citados. Assim como ocorre com o maior número de autores, a variável engajamento dos autores pode representar uma vantagem para o recebimento de citações. Autores mais engajados podem atuar na divulgação de seus trabalhos, seja na fase pré-publicação, como em congressos e outros eventos científicos ou mesmo em conversas informais, seja na fase pós-publicação, como em mídias sociais ou sites institucionais. Essa divulgação favorece o contato de outros autores com o artigo, aumentando a chance de este ser citado. Quanto maior a visibilidade do autor, maior a visibilidade do artigo, aumentando a chance de este ser lido, usado e, consequentemente, citado.

Se houver colaboração internacional entre os autores, a chance de o artigo ser moderadamente ou altamente citado aumenta. Por um lado, estudos com maior qualidade e impacto na ciência podem atrair mais colaborações de pesquisadores estrangeiros. Por outro lado, presença de colaboração entre países aumenta a chance de financiamento e, consequentemente,

os pesquisadores terão mais recursos, o que também pode refletir na qualidade final da pesquisa e dos artigos que a reportam. Além disso, quanto mais países envolvidos na condução e publicação de um estudo, maior é o potencial de esse estudo atrair o interesse da comunidade científica mundial, o que aumenta o público citante. Xie et al. (2019), em estudo analisando 66 fatores potencialmente associados a citações, identificaram que a frequência de citação não mostrava significância estatística com a colaboração internacional, mas com a colaboração entre instituições. Aqui, testamos ambos os tipos de colaboração; porém, dado que a variável número de instituições apresentou colinearidade com o número de autores e de países, optamos por excluí-la da análise. Ademais, no estudo de Xie et al. (2019), a colaboração institucional esteve presente em 14% da amostra, ao passo que houve apenas um caso (0,2%) de colaboração internacional na amostra ($n = 566$), o que pode ter inviabilizado a identificação de alguma associação com esta última variável. No presente estudo, identificamos que cada país a mais em colaboração aumenta em 22% (OR = 1,22; IC95% 1,02–1,46) e 26% (OR = 1,26; IC95% 1,04–1,51) a chance de o artigo ser moderada ou altamente citado, respectivamente.

Com relação à autoria dos artigos, cada autor aumenta em 9% (OR = 1,09; IC95% 1,04–1,14) e 11% (OR = 1,11; IC95% 1,06–1,17) a chance de o artigo ser moderada ou altamente citado, respectivamente. Com efeito, quanto mais autores, maior o potencial de recebimento de autocitações, divulgação do trabalho – desde o período de condução do estudo até após sua publicação – e de haver mais pesquisadores na linha de pesquisa do artigo, isto é, mais citantes. Todos esses fatores potencializam as chances de um artigo receber citações. Segundo estudo de Fox et al. (2016) com periódicos da área de ecologia, o tamanho do artigo, o número de autores e o número de referências do artigo mostraram-se independentemente associados a um maior número de citações recebidas. Os autores identificaram que, para cada aumento de 10% no

número de autores, o número de vezes que um artigo é citado aumenta em 1,9%. Nabout et al. (2015) identificaram diminuição no número de artigos com autoria única na área de biologia, bem como aumento na interdisciplinaridade, combinando o conhecimento de múltiplos autores e produzindo artigos de alto impacto. De fato, no presente estudo vimos o efeito positivo da interdisciplinaridade sobre as citações, e a interdisciplinaridade decorrente do conjunto de saberes de diversos autores em colaboração também pode ser parte do mecanismo explicativo de como o número de autores aumenta as chances de citação de um artigo.

5.2 VALIDAÇÃO DA WOS

O campo financiamento da WoS mostrou-se confiável para análises bibliométricas sobre financiamento. O idioma de publicação e o local do artigo onde o financiamento é reportado parecem influenciar a confiabilidade desse campo.

Em 2008, a WoS passou a inserir essa informação em seus registros e, desde então, estudos sobre financiamento de pesquisa com dados extraídos da WoS vêm mencionando alguma limitação (Grassano et al., 2017), seja por algoritmos da WoS, seja por falta de informação do editor do periódico ou do autor quanto à reportagem de financiamento no artigo (Alvarez-Bornstein et al., 2017). No presente estudo, o campo financiamento da WoS apresentou-se confiável para análises bibliométricas por ter indicadores de acurácia e avaliação de concordância elevados. Na área de Nutrição e Dietética, identificamos que a sensibilidade foi de 96,4%. Wang e Shapira (2011) identificaram na SCI em 740.000 registros no campo da nanotecnologia 67% de

financiamento. Semelhantemente, identificamos que, segundo a WoS, 56,5% dos artigos de Nutrição e Dietética apresentam declaração de financiamento. Koier e Horlings (2015) identificaram que a WoS deixa de registrar financiamento em 24% dos artigos financiados da área de mudança climática. Aqui, apenas 5,5% da amostra (174/3.183) foram erroneamente classificados como não tendo financiamento.

Ao analisarmos o local do artigo onde constava a informação sobre financiamento, identificamos diferença significativa entre a WoS e a revisão manual. Nos artigos cujo financiamento constava apenas na revisão manual, maiores proporções de financiamento foram informadas em notas de rodapé e no campo conflito de interesses do que naqueles cujo financiamento constava tanto na revisão manual quanto na WoS. A princípio, a WoS coleta a informação sobre financiamento do campo *Acknowledgements* ou *Funding*. Os periódicos que adotam como padrão declarar o financiamento em nota de rodapé ou no campo conflito de interesses podem informar à WoS sobre esse padrão e, caso a WoS passasse a coletar o financiamento nesses campos, para esses periódicos, teríamos o motivo de existirem verdadeiros positivos – corretamente classificados pela WoS como tendo financiamento – nesses campos. Isso explicaria porque para alguns periódicos, i.e., aqueles que não informaram sobre seu padrão à WoS, a informação sobre financiamento reportada em notas de rodapé ou no campo conflito de interesse não foi coletada pela WoS. Quanto aos 45 registros da WoS que informavam financiamento sem que essa condição fosse identificada na revisão manual, aparentemente, alguns periódicos enviam os metadados via FTP (file transfer protocol) às bases indexadoras, mas não publicam a informação sobre financiamento no arquivo em PDF do artigo, que foi o documento usado para a presente análise.

Proporção significativamente maior de falsos negativos foi encontrada nos artigos escritos em outro idioma que não o inglês. Dos artigos que tinham financiamento não identificado pela WoS, 10% eram escritos em outro idioma que não o inglês. Adicionalmente, quando a WoS informa que tem financiamento e o artigo está escrito em outro idioma que não o inglês, não há erro de classificação. Diaz-Faes e Bordons (2014) relataram que a WoS captura e processa apenas financiamentos escritos em inglês e essa inclusão depende da equipe de suporte técnico da Thompson Reuters. Tang et al. (2017), em análise bibliométrica sobre a presença do registro de financiamento nos artigos científicos em outros idiomas além do inglês, verificaram a taxa de presença de registro de financiamento por idioma de publicação. Os autores identificaram cerca de 9,45 milhões de publicações em inglês e, dentre os artigos de todos os idiomas, 4,59 milhões declaravam financiamento. A taxa de presença de financiamento do idioma inglês foi de 49%. Ao contrário do estudo de Diaz-Faes e Bordons (2014), Tang et al. (2017) viram que, dentre publicações de 2009 a 2014, as escritas em outras línguas também têm seus dados de financiamento coletados pela WoS. Contudo, as taxas de presença de financiamento são extremamente baixas para a maioria dos outros idiomas, exceto o chinês, dado que mais de um terço dos artigos chineses continham informações de financiamento. Artigos escritos em alemão, francês, espanhol, português, polonês, japonês, russo e turco apresentavam no máximo 0,15% de registro de financiamento. No presente estudo, vimos que 90% dos artigos escritos em outro idioma que não o inglês foram corretamente registrados pela WoS.

Este estudo mostra a importância de os pesquisadores não apenas buscarem auxílios financeiros, mas também reportarem o financiamento em seus artigos. Com efeito, alguns financiamentos que vão para a estrutura de pesquisa e treinamento de estudantes, por exemplo, podem ser omitidos no artigo. Adicionalmente, financiamentos podem ser mantidos em sigilo no

caso de patentes, por exemplo (Wang e Shapira, 2011). O fato de que nem todos os financiamentos são informados nos artigos também pode decorrer da falta de adoção sistemática dessa prática por alguns periódicos. Portanto, também são necessários incentivos por parte dos editores de revistas científicas, seja enfatizando a importância de os autores reportarem os financiamentos recebidos, seja tornando esses dados públicos.

Neste estudo, 3% dos artigos foram excluídos porque não foram encontrados na íntegra; logo, a concordância entre a WoS e a revisão manual pode estar subestimada ou superestimada. Entretanto, ainda que considerássemos que todos esses artigos fossem falsos negativos ou falsos positivos, a sensibilidade e a especificidade ainda seriam acima de 85%. Portanto, concluímos que os registros de financiamento fornecidos pela WoS são confiáveis. Como o idioma de publicação e o local do artigo onde o financiamento é reportado parecem influenciar a confiabilidade desse campo, a padronização da reportagem de financiamento, por parte tanto dos autores como dos periódicos, pode melhorar a confiabilidade das bases de dados.

5.3 LIMITAÇÕES

Algumas limitações deste estudo devem ser ressaltadas. Embora o artigo científico possua uma estrutura complexa, nesta pesquisa, incluímos apenas as variáveis que estavam disponíveis ou que poderiam ser obtidas manualmente de forma apurada e durante o período da pesquisa. Além disso, como o potencial de citação pode variar significativamente entre os campos, aqui

analisamos artigos limitados a periódicos de Nutrição e Dietética; nosso modelo, portanto, se aplica a essa área e outras que apresentem características semelhantes. Dessa forma, a despeito de usarmos uma área específica como amostra, entendemos que as conclusões podem ser inferidas e testadas para qualquer área em que os artigos científicos sejam a principal forma de divulgação do conhecimento e a citação, uma ferramenta importante de reconhecimento do trabalho dos pares. Adicionalmente, apesar da ampla diversidade de variáveis preditoras analisadas, outras variáveis podem ser incluídas em pesquisas futuras. O modelo possui um poder preditivo de 65,7%, o que sugere que há espaço teórico para o teste de outras variáveis. O fato de o artigo estar ou não escrito de acordo com as diretrizes de publicação para seu tipo de estudo – por exemplo, Strobe para estudos observacionais, Prisma para revisões e meta-análises e Consort para ensaios clínicos randomizados – pode influenciar a qualidade do artigo e, conseqüentemente, seu impacto. Dessa forma, recomendamos a condução de pesquisas em que esse e outros fatores, sobretudo os passíveis de serem modificados pelos autores mesmo após o término das pesquisas, sejam incluídos no modelo múltiplo.

Outro aspecto deste estudo que deve ser enfatizado é que não temos uma resposta contundente dos motivos que levam os cientistas a citar um trabalho e não citar um outro, uma vez que analisamos dados secundários como variáveis proxy de potenciais razões. Pesquisas futuras podem incluir entrevistas aos cientistas de modo a clarificar os mecanismos pelos quais os fatores aqui identificados, bem como outros não abordados, afetam a comunicação científica. Entretanto, questionar um autor pode não revelar as verdadeiras razões para citar, já que o autor pode não se lembrar (viés de memória) ou não ter boa compreensão delas, ou mesmo responder de forma a enganar deliberadamente o entrevistador quanto às suas reais motivações (Bornmann

e Daniel, 2006). Dessa forma, o melhor momento para discussão com o autor sobre seus motivos é a fase de redação do artigo, o que envolve muitos desafios, mas pode ser factível.

Outra ressalva deve ser feita em relação à variável engajamento dos autores. Essa variável foi operacionalizada a partir dos registros de ResearcherID e Orcid, identificadores únicos criados em 2008 e 2012, respectivamente. Esses campos não são indexados ativamente pela WoS, por exemplo por coleta nos metadados ou PDF do artigo, mas sim registrados a partir da vinculação dos artigos aos identificadores únicos por parte dos próprios autores. Isso significa que essas vinculações foram feitas em qualquer momento após 2008, i.e., no decorrer do período de coleta de dados da presente pesquisa, que ocorreu em 2017. Assim, por um lado, para o Orcid, pode ter havido um período expressivo de não exposição dos autores a essa variável, causando subestimação da associação, sobretudo para os artigos publicados em 2009 e 2010. Por outro lado, para o ResearcherID, como a coleta foi feita em 2017, os autores podem ter feito a vinculação até dois anos depois do final da janela de citação considerada, o que pode ter superestimado a associação, sobretudo para os artigos publicados em 2009 e 2010. Entretanto, a informação de quando o autor faz a vinculação não está disponível nos metadados, o que impossibilita a averiguação do efeito dessa limitação sobre a associação.

Ressaltamos, ainda, que o fato de este estudo considerar artigos de 2009 a 2011 pode ter conferido maior sentido ao uso dos dados de identificadores como proxy do engajamento, uma vez que nesse período a obtenção desses dados dependia mais da iniciativa dos autores. Atualmente, pode ser conveniente escolher outra informação para operacionalizar a variável engajamento, pois a exigência, por parte dos periódicos ou das instituições de pesquisa, de declaração de identificadores nos artigos científicos mascara o quanto dessa declaração se deve ao engajamento do autor. Com efeito, variáveis operacionais distintas podem satisfatoriamente

representar uma mesma variável teórica. A conclusão da pesquisa, entretanto, é feita com base na variável teórica, ou seja, é o maior engajamento por parte dos autores que aumenta as citações.

As análises estatísticas usadas nesta pesquisa permitem identificar fatores que sejam independentemente associados à variável dependente, ou seja, a associação entre uma variável preditora e as citações recebidas pelo artigo, independentemente das demais variáveis também analisadas; não permitem, porém, concluir sobre possíveis interações entre as variáveis. Por exemplo, receber financiamento mostrou-se associado a ser moderada ou altamente citado, assim como o número de autores e quartil do periódico. Entretanto, o fato de um artigo mencionar financiamento e, concomitantemente, ser publicado em periódico do Q1 pode ter um impacto ainda maior nas citações. Interações como essa não foram testadas neste estudo, mas poderão fazer parte de pesquisas futuras. Da mesma forma, apesar de o acesso aberto poder estar relacionado ao quartil do periódico, essa relação foi ajustada no modelo múltiplo, de modo que a associação entre acesso aberto e menor chance de ser moderada ou altamente citado se faz presente independentemente do quartil do periódico. É possível, entretanto, que haja interação entre ambos os fatores (acesso aberto e quartil do periódico) e pode ser conveniente que, em análises futuras, essas interações sejam testadas no modelo múltiplo. Apesar dessas ressalvas, podemos prever com relativa precisão o impacto de diversos fatores sobre a citação nos primeiros cinco anos após a publicação dos artigos científicos.

Quanto à validação da WoS, optamos por analisar apenas os metadados do campo financiamento, devido à sua importância na avaliação científica e ao fato de sua indexação ser mais recente e ter gerado dúvidas quanto a sua validade. Para pesquisas futuras semelhantes a esta, em que sejam analisadas informações sobre financiamento, pode-se optar pelo uso da base

de dados Scopus^g, que em 2013 passou a coletar e indexar esse dado. Diferentemente da WoS, porém, a Scopus não fornece dados para download das categorias às quais os periódicos pertencem.

Em 2018, foi lançada uma nova base de dados que fornece relatório de citações, a Dimension^h. Uma vez que essa plataforma fornece dados de citações e métricas alternativas derivados de várias fontes, pesquisas contemplando análise de citações podem usá-la para identificar o impacto dos artigos de uma forma mais abrangente. Atualmente, dados sobre acesso aberto são fornecidos pela WoS, Scopus e Dimensions, o que substitui a coleta manual e, conseqüentemente, possibilita o processamento de dados em maior escala.

Adicionalmente, acaba de ser lançado pela Clarivate, na plataforma do InCites, um novo esquema de categorização de áreas, chamado *Citation Topics*ⁱ. Nele, os artigos são classificados segundo suas temáticas com base na atividade de citações, tanto as recebidas quanto as feitas (contidas em suas listas de referências). Essa maneira de medir a multidisciplinaridade de um artigo já vinha sendo aplicada em algumas pesquisas, mas a classificação era feita de forma manual. Essa nova funcionalidade do InCites permite que as futuras análises que levem em consideração a classificação temática do artigo, e não do periódico no qual o artigo foi publicado, sejam realizadas com menor investimento de tempo por parte dos pesquisadores e em maior escala.

^g Elsevier B.V. Scopus. Disponível em: <https://www.scopus.com/search/form.uri?display=basic#basic>

^h Digital Science & Research Solutions, Inc. Dimensions. 2021. Disponível em: <https://app.dimensions.ai/discover/publication>

ⁱ Clarivate. Citation Topics. 2020. Disponível em: <https://incites.help.clarivate.com/Content/Research-Areas/citation-topics.htm>

5.4 CONCLUSÕES

Em resumo, nossos resultados sugerem que o impacto da citação de um artigo pode ser previsto por indicadores cientométricos. Características dos artigos, das pesquisas, dos periódicos e dos autores estão associados ao número de citações recebidas. Adicionalmente, concluímos que a WoS é uma base confiável para estudos bibliométricos e recomendamos algumas práticas por parte dos autores e dos editores dos periódicos para aumentar a qualidade dos metadados, bem como para as próximas pesquisas nesse tema (Quadro 4).

Quadro 4. Principais recomendações a partir das conclusões deste estudo.

Alvo	Principais recomendações
Autores	1) Prioridade à escrita de títulos contendo a conclusão do estudo;
	2) Minimização das restrições quanto ao tamanho dos resumos, da lista de referências e do artigo como um todo;
	3) Declaração de financiamentos recebidos, caso se aplique;
	4) Engajamento dos autores nas atividades pós-publicação;
Editores	5) Adoção sistemática da coleta e publicação dos dados de financiamento e identificadores únicos;
	6) Minimização de restrições quanto ao tamanho dos resumos, da lista de referências e do artigo como um todo;
Pesquisas futuras	7) Inclusão no modelo múltiplo de outros fatores potencialmente associados às citações;
	8) Inclusão de entrevistas aos autores com o objetivo de complementar a compreensão dos motivos que os levam a citar um dado trabalho;
	9) Análise de interações entre as variáveis no modelo múltiplo.

6 REFERÊNCIAS

Abt HA. Long-term citation histories of astronomical papers. Publications of the Astronomical Society of the Pacific. 1981; 93:207-10.

Ahmed A, Adam M, Ghafar NA, Muhammad M, Ebrahim NA. Impact of article page count and number of authors on citations in Disability related fields: a systematic review article. Iran J Public Health. 2016; 45(9):1118-1125.

Alvarez-Bornstein B, Morillo F, Bordons M. Funding acknowledgments in the Web of Science: completeness and accuracy of collected data. Scientometrics. 2017;112:1793-1812.
<https://doi.org/10.1007/s11192-017-2453-4>

Alves MC. Plano S expõe conflitos no mercado de edição de artigos científicos. Ciência e Cultura. 2019;71. <https://doi.org/10.21800/2317-66602019000100007>

Amara N, Landry R, Halilem N. What can university administrators do to increase the publication and citation scores of their faculty members? Scientometrics. 2015;103:489-530.
<https://doi.org/10.1007/s11192-015-1537-2>

Antoniou GA, Antoniou SA, Georgakarakos EI, Sfyroeras GS, Georgiadis GS. Bibliometric analysis of factors predicting increased citations in the vascular and endovascular literature. Annals of Vascular Surgery. 2015;29(2):286-292. <https://doi.org/10.1016/j.avsg.2014.09.017>

Araújo CA. Bibliometria: evolução histórica e questões atuais. *Em Questão*. 2006;12:11-32.

Basson I, Blanckenberg JP, Prozesky H. Do open access journal articles experience a citation advantage? Results and methodological reflections of an application of multiple measures to an analysis by WoS subject areas. *Scientometrics*. 2021;126:459-484.
<https://doi.org/10.1007/s11192-020-03734-9>

Bhandari M, Busse J, Devereaux PJ, Montori VM, Swiontkowski M, Tornetta P, et al. Factors associated with citation rates in the orthopedic literature. *Canadian Journal of Surgery*. 2007;50(2):119-23.

Björk B-C, Solomon D. Open access versus subscription journals: a comparison of scientific impact. *BMC Medicine*. 2012;10:73. <https://doi.org/10.1186/1741-7015-10-73>

Bornmann L, Daniel H-D. What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*. 2006;64:45-80.

Bornmann L, Schier H, Marx W, Daniel H-D. What factors determine citation counts of publications in chemistry besides their quality? *Journal of Informetrics*. 2012;6:11-18.
<https://doi.org/10.1016/j.joi.2011.08.004>

Boyd CJ, Gentry ZL, Martin KD, Rais-Bahrami S. Factors associated with the highest and lowest cited research articles in urology journals. *Urology*. 2019;124:23-27.

<https://doi.org/10.1016/j.urology.2018.11.034>.

Camargo Jr KR. O rei está nú, mas segue impávido: os abusos da bibliometria na avaliação da Ciência. *Saúde & Transformação Social*. 2010;1:3-8. Disponível em:

<https://www.redalyc.org/pdf/2653/265319560003.pdf>

Camargo Jr KR. A indústria de publicação contra o acesso aberto. *Revista de Saúde Pública*. 2012;46(6):1090-4. <https://doi.org/10.1590/S0034-89102013005000006>

Conen D, Torres J, Ridker PM. Differential citation rates of major cardiovascular clinical trials according to source of funding - a survey from 2000 to 2005. *Circulation*. 2008;118:1321-1327.

<https://doi.org/10.1161/CIRCULATIONAHA.108.794016>

Craig ID, Plume AM, McVeigh ME, Pringle J, Amin M. Do open access articles have greater citation impact? A critical review of the literature. *Journal of Informetrics*. 2007;1:239-248.

Cuenca AMB, Noronha DP, Andrade MTD, Ueno HM, Kobayashi KM. Periódicos brasileiros de saúde pública: a questão do financiamento. *InCID: Revista De Ciência da Informação e Documentação*. 2011;2(2):101-110.

Davis PM. Open access, readership, citations: a randomized controlled trial of scientific journal publishing. *The FASEB J*. 2011;25:2129-34. <https://doi.org/10.1096/fj.11-183988>

de Solla Price D. A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*. 1976;27(5-6):292-306.

de Solla Price DJ. Networks of scientific papers. *Science*. 1965;149(3683):510-515.

Di Bitetti MS, Ferreras JÁ. Publish (in English) or perish: The effect on citation rate of using languages other than English in scientific publications. *Ambio*. 2017;46:121-127.

<https://doi.org/10.1007/s13280-016-0820-7>

Diaz-Faes AA, Bordons M. Acknowledgments in scientific publications: presence in Spanish science and text patterns across disciplines. *Journal of the Association for Information Science and Technology*. 2014;65(9):1834-1849.

Diekhoff T, Schlattmann P, Dewey M. Impact of article language in multi-language medical journals - a bibliometric analysis of self-citations and impact factor. *Plos One*.

2013;8(10):e76816. <https://doi.org/10.1371/journal.pone.0076816>

Dorta-González P, González-Betancor SM, Dorta-González MI. Reconsidering the gold open access citation advantage postulate in a multidisciplinary context: an analysis of the subject categories in the Web of Science database 2009–2014. *Scientometrics*. 2017;112:877-901.

<https://doi.org/10.1007/s11192-017-2422-y>

Eysenbach G. Citation Advantage of open access articles. *Plos Biology*. 2006;4(5):e157.

Dorta-González P, Santana-Jiménez Y. Prevalence and citation advantage of gold open access in the subject areas of the Scopus database. *Research Evaluation*. 2018;27:1-15.

Filion KB, Pless IB. Factors related to the frequency of citation of epidemiologic publications. *Epidemiologic Perspectives & Innovations*. 2008;5:3.

Fox CW, Paine CET, Sauterey B. Citations increase with manuscript length, author number, and references cited in ecology journals. *Ecology and Evolution*. 2016;6:7717-7726.

<https://doi.org/10.1002/ece3.2505>

Frandsen TF, Nicolaisen J. Citation behavior: a large-scale test of the persuasion by name-dropping hypothesis. *Journal of the Association for Information Science and Technology*. 2016;68(5):1278-1284. <https://doi.org/10.1002/asi.23746>

Garfield E. Can citation indexing be automated? In: Stevens ME, Guiliano VE, Heilprin LB, editors. *Statistical association methods for mechanized documentation, Symposium proceedings*, 1964:189-192.

Garfield E. Citation indexes for science: a new dimension in documentation through association of ideas. *Science*. 1955;1(3159):108-111.

Gargouri Y, Hajjem C, Larivière V, Gingras Y, Carr L, Brody T, et al. Self-selected or mandated, open access increases citation impact for higher quality research. *Plos One*. 2010;5(10):e13636.

<https://doi.org/10.1371/journal.pone.0013636>

Glänzel W, Debackere K, Thijs B, Schubert A. A concise review on the role of author self-citations in information science, bibliometrics and science policy. *Scientometrics*. 2006;67(2):263-277. <https://doi.org/10.1556/Scient.67.2006.2.8>

Grassano N, Moon J, Hopkins M. Funding data from publication acknowledgements: coverage, uses and limitations. *SSRN Electronic Journal*. 2016. <https://doi.org/10.2139/ssrn.2767348>.

Gross PLK, Gross EM. College libraries and chemical education. *Science*. 1927;66(1713):385-389. <https://doi.org/10.1126/science.66.1713.385>

Habibzadeh F, Yadollahie M. Are shorter article titles more attractive for citations - cross-sectional study of 22 scientific journals. *Croatian Medical Journal*. 2010;51:165-170.

Hamilton DP. Research papers: who's uncited now? *Science*. 1991;251:25.

Harnad S, Brody T. Comparing the impact of open access (OA) vs. non-OA articles in the same journals. *D-Lib Magazine*. 2004;10(6). [citado 20 jul 2020]. Disponível em: <http://www.dlib.org/dlib/june04/harnad/06harnad.html>

Holmberg K, Hedman J, Bowman TD, Didegah F, Laakso M. Do articles in open access journals have more frequent altmetric activity than articles in subscription-based journals? An investigation of the research output of Finnish universities. *Scientometrics*. 2020;122:645-659. <https://doi.org/10.1007/s11192-019-03301-x>

Hood WW, Wilson CS. The literature of bibliometrics, scientometrics, and informetrics. *Scientometrics*. 2001;52(2):291-314. <https://doi.org/10.1023/A:1017919924342>

Hua F, Sun H, Walsh T, Glenny A-M, Worthington H. Open access to journal articles in oncology: current situation and citation impact. *Annals of Oncology*; 2017;28:2612-2617. <https://doi.org/10.1093/annonc/mdx398>

Hua F, Sun H, Walsh T, Worthington H, Glenny A-M. Open access to journal articles in dentistry: prevalence and citation impact. *Journal of Dentistry*. 2016;47:41-48. <https://doi.org/10.1016/j.jdent.2016.02.005>

Jacques TS, Sebire NJ. The impact of article titles on citation hits: an analysis of general and specialist medical journals. *Journal of the Royal Society of Medicine Short Reports*. 2010;1:1-5. <https://doi.org/10.1258/shorts.2009.100020>

Jamali HR, Nikzad M. Article title type and its relation with the number of downloads and citations. *Scientometrics*. 2011;88:653-661. <https://doi.org/10.1007/s11192-011-0412-z>

Koier E, Horlings E. How accurately does output reflect the nature and design of transdisciplinary research programmes? *Research Evaluation*. 2015;24:37-50. <https://doi.org/10.1093/reseval/rvu027>

Krampen G, Weiland P, Wiesenhütter J. Citation success of different publication types: a case study on all references in psychology publications from the German-speaking countries (D–A–CH–L–L) in 2009, 2010, and 2011. *Scientometrics*. 2015;104:827-840.

<https://doi.org/10.1007/s11192-015-1573-y>

Kulkarni AV, Busse JW, Shams I. Characteristics associated with citation rate of the medical literature. *PLoS One*. 2007;2(5):e403. <https://doi.org/10.1371/journal.pone.0000403.t001>

Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-174.

Larivière V, Archambault É, Gingras Y. Long-term variations in the aging of scientific literature: from exponential growth to steady-state science (1900–2004). *Journal of the American Society for Information Science and Technology*. 2008;59(2):288-296. <https://doi.org/10.1002/asi.20744>

Larivière V, Gingras Y, Archambault É. The decline in the concentration of citations, 1900–2007. *Journal of the American Society for Information Science and Technology*. 2009;60(4):858-862.

Larivière V, Gingras Y. On the relationship between interdiscinarity and scientific impact. *Journal of the American Society for Information Science and Technology*. 2010b;61:126-131. <https://doi.org/10.1002/asi.21226>

Larivière V, Gingras Y. The impact factor's Matthew effect: a natural experiment in Bibliometrics. *Journal of the American Society for Information Science and Technology*. 2010a;61(2):424-427. <https://doi.org/10.1002/asi.21232>

Lawrence S. Free online availability substantially increases a paper's impact. *Nature*. 2001. <https://doi.org/10.1038/nature28042>

Letchford A, Moat HS, Preis T. The advantage of short paper titles. *Open Society Open Science*. 2015;2:150266. <https://doi.org/10.1098/rsos.150266>

Letchford A, Preis T, Moat HS. The advantage of simple paper abstracts. *Journal of Informetrics*. 2016;10(1):1-8.

Lozano GA, Larivière V, Gingras Y. The weakening relationship between the Impact Factor and papers' citation in the digital age. *Journal of the American Society for Information Science and Technology*. 2012;63(11):2140-2145. <https://doi.org/10.1002/asi.22731>

Mabe MA, Amin M. Dr. Jekyll and Dr. Hyde: author-reader asymmetries in scholarly publishing. *Aslib Proceedings*. 2002;54(3):149-157. <https://doi.org/10.1108/00012530210441692>

Macias-Chapula CA. O papel da informetria e da cienciometria e sua perspectiva nacional e internacional. *Ciência da Informação*. 1998;27(2):134-140.

MacRoberts MH, MacRoberts BR. Problems of citation analysis: a critical review. *Journal of the American Society for Information Science*. 1989;40(5):342-349.

[https://doi.org/10.1002/\(SICI\)1097-4571\(198909\)40:5<342::AID-ASI7>3.0.CO;2-U](https://doi.org/10.1002/(SICI)1097-4571(198909)40:5<342::AID-ASI7>3.0.CO;2-U)

MacRoberts MH, MacRoberts BR. Problems of citation analysis: a study of uncited and seldom-cited influences. *Journal of the American Society for Information Science and Technology*. 2010;61(1):1-13.

Mammola S, Fontaneto D, Martínez A, Chichorro F. Impact of the reference list features on the number of citations. *Scientometrics*. 2021;126:785-799. <https://doi.org/10.1007/s11192-020-03759-0>

Marques F. O labirinto do Plano S. *Revista Pesquisa Fapesp*. 2021;299 [citado 6 mar 2021]. Disponível em: <https://revistapesquisa.fapesp.br/o-labirinto-do-plano-s/>

Merton RK. The Matthew effect in science II: cumulative advantage and the symbolism of intellectual property. *JSTOR*. 1988;79(4):606-623.

Merton RK. The Matthew Effect in Science. *Science*. 1968;159(3810):56-63.

Mike Thelwall. Bibliometrics to webometrics. *Journal of Information Science*. 2008;34(4):605-621. <https://doi.org/10.1177/0165551507087238>

Moed HF, Garfield E. In basic science the percentage of “authoritative” references decreases as bibliographies become shorter. *Scientometrics*. 2004;60:295-303.

Nabout JC, Parreira MR, Teresa FB, Carneiro FM, Cunha HF, Ondeí LS, et al. Publish (in a group) or perish (alone): the trend from single- to multi-authorship in biological papers.

Scientometrics. 2015;102:357-364. <https://doi.org/10.1007/s11192-014-1385-5>

Nicolaisen J, Frandsen TF. Acting hot or not? Testing the citing to show-off hypothesis. *Journal of Documentation*. 2021;77(2):461-478. <https://doi.org/10.1108/JD-06-2020-0097>

Nicolaisen J, Frandsen TF. Number of references: a large-scale study of interval ratios. 2020; *Scientometrics*. 2021;126:259-285. <https://doi.org/10.1007/s11192-020-03764-3>

Okike K, Kocher MS, Torpey JL, Bhandari M. Level of evidence and conflict of interest disclosure associated with higher citation rates in orthopedics. *Journal of clinical epidemiology*. 2011;64(3):331-338.

Paci M, Landi N, Briganti G, Lombardi B. Factors associated with citation rate of randomized controlled trials in physiotherapy. *Archives of Physiotherapy*. 2015;5:9.

<https://doi.org/10.1186/s40945-015-0009-6>

Patel RB, Vaduganathan M, Mosarla RC, Venkateswaran RV, Bhatt DL, Bonow RO. Open access publishing and subsequent citations among articles in major cardiovascular journals. *The*

American Journal of Medicine. 2019;132(9):1103-1105.

<https://doi.org/10.1016/j.amjmed.2019.02.009>

Paul-Hus A, Desrochers N, Costas R. Characterization, description, and considerations for the use of funding acknowledgement data in Web of Science. *Scientometrics*. 2016;108:167–182.

<https://doi.org/10.1007/s11192-016-1953-y>

Pendlebury DA. Science, Citation, and Funding. *Science*. 1991; 251:1410-1411.

Piwowar H, Priem J, Larivière V, Alperin JP, Matthias L, Norlander B, et al. The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ*. 2018;6:e4375.

<https://doi.org/10.7717/peerj.4375>

Poomkottayil D, Bornstein MM, Sendi P. Lost in translation: the impact of publication language on citation frequency in the scientific dental literature. *Swiss Medical Weekly*. 2011;141:w13148.

<https://doi.org/10.4414/smw.2011.13148>

Pritchard A. Statistical bibliography or bibliometrics? *Journal of Documentation*.

1969;25(4):348-349.

Schoonbaert D, Roelants G. Citation analysis for measuring the value of scientific publications: quality assessment tool or comedy of errors? *Tropical Medicine and International Health*.

1996;1(6):739-752.

Shu F, Julien C-A. Does the web of science accurately represent chinese scientific performance?

Journal of the Association for Information Science and Technology. 2019;70(10):1138-1152.

<https://doi.org/10.1002/asi.24184>

Small HG. Cited documents as concept symbols. Social Studies of Science (SAGE, London and

Beverly Hills). 1978;8(1978):327-40.

Sooryamoorthy R. Do types of collaboration change citation - Collaboration and citation patterns

of South African science publications. Scientometrics. 2009;81(1):177-193.

Sotudeh H, Estakhr Z. Sustainability of open access citation advantage - the case of Elsevier's

author-pays hybrid open access journals. Scientometrics. 2018;115:563-576.

Tague-Sutcliffe J. An introduction to informetrics. Information Processing & Management 1992;

28:1-3.

Tahamtan I, Afshar AS, Ahamdzadeh K. Factors affecting number of citations - a comprehensive

review of the literature. Scientometrics. 2016;107:1195-1225. [https://doi.org/10.1007/s11192-](https://doi.org/10.1007/s11192-016-1889-2)

[016-1889-2](https://doi.org/10.1007/s11192-016-1889-2)

Tang L, Hu G, Liu W. Funding acknowledgment analysis - Queries and caveats. Journal of the

Association for Information Science and Technology. 2017;68(3):790-794.

<https://doi.org/10.1002/asi.23713>

Teixeira MC, Thomaz SM, Michelin TS, Mormul RP, Meurer T, Fasolli JVB, et al. Incorrect citations give unfair credit to review authors in ecology journals. *Plos One*. 2013;8(12):e81871.

<https://doi.org/10.1371/journal.pone.0081871>

van Dalen HP, Henkens K. What makes a scientific article influential? The case of demographers. *Scientometrics*. 2001;50(3):455-482.

Wang J, Shapira P. Funding acknowledgement analysis: an enhanced tool to investigate research sponsorship impacts: the case of nanotechnology. *Scientometrics*. 2011;87:563-586.

<https://doi.org/10.1007/s11192-011-0362-5>

Wang J, Shapira P. Is there a relationship between research sponsorship and publication impact? An analysis of funding acknowledgments in nanotechnology papers. *Plos One*.

2015;10(2):e0117727. <https://doi.org/10.1371/journal.pone.0117727>

Wang M, Yu G, An S, Yu D. Discovery of factors influencing citation impact based on a soft fuzzy rough set model. *Scientometrics*. 2012;93:635-644. [https://doi.org/10.1007/s11192-012-](https://doi.org/10.1007/s11192-012-0766-x)

[0766-x](https://doi.org/10.1007/s11192-012-0766-x)

Wang Q, Schneider JW. Consistency and validity of interdisciplinarity measures. *Quantitative Science Studies*. 2020;1:239-263. https://doi.org/10.1162/qss_a_00011

Wei T, Li M, Wu C, Yan X-Y, Fan Y, Di Z, et al. Do scientists trace hot topics? *Scientific Reports*. 2013;3:2207. <https://doi.org/10.1038/srep02207>

Willis DL, Bahler CD, Neuberger MM, Dahm P. Predictors of citations in the urological literature. *BJU International*. 2011;107:1876-1880. <https://doi.org/10.1111/j.1464-410X.2010.10028.x>

Xie J, Gong K, Li J, Ke Q, Kang H, Cheng Y. A probe into 66 factors which are possibly associated with the number of citations an article received. *Scientometrics*. 2019;119:1429-1454. <https://doi.org/10.1007/s11192-019-03094-z>

Yu T, Yu G, Li P-Y, Wang L. Citation impact prediction for scientific papers using stepwise regression analysis. *Scientometrics*. 2014;101:1233-1252. <https://doi.org/10.1007/s11192-014-1279-6>

Zhao SX. Uncitedness of reviews. *Current Science*. 2015;109(8):1377-1378.

CURRÍCULO LATTES DO AUTOR

Milena Maria de Araújo Lima Barbosa

Endereço para acessar este CV: <http://lattes.cnpq.br/5290589267213966>

ID Lattes: **5290589267213966**

Última atualização do currículo em 07/03/2021

Doutoranda em Saúde Pública. Mestre em Nutrição em Saúde Pública pela Faculdade de Saúde Pública da Universidade de São Paulo. Formada em Nutrição pela mesma instituição em 2010. Participa do grupo de pesquisa em Comunicação Científica em Saúde Pública da mesma instituição. Atua como editora executiva adjunta da Revista de Saúde Pública desde 2013. (Texto informado pelo autor)

Identificação

Nome Milena Maria de Araújo Lima Barbosa

Nome em citações bibliográficas ARAÚJO, M. M.; BARBOSA, M.M.A.L.; DE ARAÚJO LIMA BARBOSA, MILENA MARIA; Barbosa, Milena Maria de Araújo Lima

Lattes ID <http://lattes.cnpq.br/5290589267213966>

Endereço

Endereço Profissional Faculdade de Saúde Pública da USP.
Faculdade de Saúde Pública
Pacaembu
01246904 - São Paulo, SP - Brasil
Telefone: (11) 30617826

Formação acadêmica/titulação

2015 Doutorado em andamento em Saúde Pública.
Faculdade de Saúde Pública - Universidade de São Paulo, FSP - USP, Brasil.
Título: Ser ou não ser citado: fatores associados à citação de artigos científicos.
Orientador: Ivan França Junior.
Coorientador: Angela Maria Belloni Cuenca.

2013 - 2015 Mestrado em Nutrição em Saúde Pública.
Faculdade de Saúde Pública da USP, FSP, Brasil.
Título: Adiponectina modifica a resposta à suplementação com ômega-3 em humanos com fatores de risco cardiovascular, Ano de Obtenção: 2015.
Orientador: Nágila Raquel Teixeira Damasceno.
Palavras-chave: Ácidos Graxos Ômega-3; Adiponectina; Doenças Cardiovasculares; Resistência à Insulina.

2012 - 2013 Aperfeiçoamento em Treinamento Técnico Nível III. (Carga Horária: 40h).
Faculdade de Saúde Pública da USP, FSP, Brasil.
Título: Propriedades físico-químicas da LDL e da HDL, marcadores cardiometabólicos e oxidativos podem ser modulados pelo consumo de ômega-3, ômega-6 e ômega-9 em indivíduos com risco cardiovascular intermediário?. Ano de finalização: 2013.
Orientador: Nágila Raquel Teixeira Damasceno.
Bolsista do(a): Fundação de Amparo à Pesquisa do Estado de São Paulo, FAPESP, Brasil.

2005 - 2010 Graduação em Nutrição.
Faculdade de Saúde Pública, FSP, Brasil.

Formação Complementar

2019 - 2019 Extensão universitária em Análise de Regressão Múltipla. (Carga horária: 40h).
Faculdade de Saúde Pública da USP, FSP, Brasil.

Fonte: <http://lattes.cnpq.br/5290589267213966>

CURRÍCULO LATTES DO ORIENTADOR

Ivan França Junior

Endereço para acessar este CV: <http://lattes.cnpq.br/5796097952164848>

ID Lattes: **5796097952164848**

Última atualização do currículo em 26/02/2021

Graduado em Medicina pela Universidade Federal do Rio Grande do Sul (1984), é mestre (1993) e doutor (1998) em Medicina Preventiva pela Faculdade de Medicina da Universidade de São Paulo. É Livre Docente pelo Departamento de Saúde Materno-Infantil (2010) da Faculdade de Saúde Pública (FSP-USP). É professor titular no Depto de Saúde, Ciclos de Vida e Sociedade (antigo Saúde Materno-Infantil) na FSP/USP, com dedicação integral à docência e à pesquisa. Em pesquisa, coordena e integra equipes brasileiras ou internacionais na área de saúde pública/saúde coletiva que, por meio de abordagens epidemiológicas e de direitos humanos, focam temas de saúde de crianças, adolescentes e jovens, da epidemia de HIV e aids e também em violência física/sexual e gênero. Vem incrementando a pesquisa e ensino na área de de informação e comunicação científicas. Participa de cinco grupos de pesquisa cadastrados no CNPq. Atuou como Editor Associado na Revista de Saúde Pública (2007-2016) e da Revista Brasileira de Epidemiologia (ad hoc). Atua como editor sênior da revista Global Public Health. No ensino de pós-graduação, além de orientador de mestrado e doutorado, participou da coordenação, pelo Programa de Pós-Graduação em Saúde Pública da FSP/USP (PPG-SP). Foi vice-coordenador do PPG-SP por dois mandatos. Foi membro da Comissão de Avaliação da Pós-Graduação em Saúde Coletiva da CAPES no triênio 2007-2009. No ensino de graduação, ministra disciplinas, para estudantes de nutrição e saúde pública, relacionadas a ciclos de vida e saúde, evidências em saúde pública e introdução à pesquisa científica. De 2016 a 2020, exerceu a presidência da Comissão de Graduação da FSP/USP. (Texto informado pelo autor)

Identificação

Nome	Ivan França Junior
Nome em citações bibliográficas	FRANÇA JUNIOR, I.; França-Junior, Ivani; França I; França Junior, Ivani; França-Jr, Ivani; FRANÇA JÚNIOR, IVAN; FRANCA-JUNIOR, IVAN; FRANÇA, IVAN; FRANÇA-JUNIOR, IVAN; FRANÇA-JUNIOR, IVAN; FRANÇA-JUNIOR, I.
Lattes ID	http://lattes.cnpq.br/5796097952164848
Orcid iD	https://orcid.org/0000-0002-7004-7301

Endereço

Endereço Profissional	Universidade de São Paulo, Faculdade de Saúde Pública, Departamento de Saúde Materno-Infantil. Av. Dr. Arnaldo 715 2o. andar sala 218 Cerqueira Cesar 01246-904 - São Paulo, SP - Brasil Telefone: (11) 30617127 Fax: (11) 30850240 URL da Homepage: http://www.fsp.usp.br/site/docentes/index/7
------------------------------	--

Formação acadêmica/titulação

1995 - 1998	Doutorado em Medicina (Medicina Preventiva) [Sp-Capital]. Universidade de São Paulo, USP, Brasil. Título: Mudança secular das estaturas de jovens na cidade de São Paulo, 1950-1976: uma abordagem para discutir a saúde, Ano de obtenção: 1998. Orientador: Guilherme Rodrigues da Silva. Bolsista do(a): Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CAPES, Brasil. Palavras-chave: Epidemiologia; Estatura Corporal; Tendência Secular. Grande área: Ciências da Saúde Setores de atividade: Saúde Humana.
1988 - 1993	Mestrado em Medicina (Medicina Preventiva) [Sp-Capital]. Universidade de São Paulo, USP, Brasil. Título: A antropometria como prática social de saúde: uma abordagem histórica, Ano de Obtenção: 1993. Orientador: Hillegonda Maria Dutihl Novaes.

Fonte: <http://lattes.cnpq.br/5796097952164848>