

UNIVERSIDADE DE SÃO PAULO
ESCOLA DE COMUNICAÇÕES E ARTES
DEPARTAMENTO DE INFORMAÇÃO E CULTURA

LUZIA SAEKO KANASHIRO

**Estudo de consistência entre as palavras-chave de artigos
científicos de saúde com interface em ciências sociais e a
indexação na base de dados LILACS**

São Paulo
2021

LUZIA SAEKO KANASHIRO

Estudo de consistência entre as palavras-chave de artigos científicos de saúde com interface em ciências sociais e a indexação na base de dados LILACS

Trabalho de Conclusão de Curso apresentado na Universidade de São Paulo, Escola de Comunicações e Artes para obtenção do título de bacharel em Biblioteconomia.

Orientação: Profa. Dra. Cibele Araújo Camargo Marques dos Santos

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.

Catálogo na Publicação
Serviço de Biblioteca e Documentação
Escola de Comunicações e Artes da Universidade de São Paulo
Dados inseridos pelo(a) autor(a)

Kanashiro, Luzia Saeko

Estudo de consistência entre as palavras-chave de artigos científicos de saúde com interface em ciências sociais e a indexação na base de dados LILACS / Luzia Saeko Kanashiro ; orientadora, Cibele Araújo Camargo Marques dos Santos. -- São Paulo, 2021.
52 p.: il.

Trabalho de Conclusão de Curso (Graduação) - Departamento de Informação e Cultura/Escola de Comunicações e Artes / Universidade de São Paulo.
Bibliografia

1. Artigo científico 2. Palavra-chave 3. Indexação em base de dados 4. Índice de consistência 5. Descritores DeCS I. Santos, Cibele Araújo Camargo Marques dos II. Título.

CDD 21.ed. - 020

Elaborado por Alessandra Vieira Canholi Maldonado - CRB-8/6194

KANASHIRO, L. S. **Estudo de consistência entre as palavras-chave de artigos científicos de saúde com interface em ciências sociais e a indexação na base de dados LILACS**. São Paulo, 2021. Monografia (TCC) – Escola de Comunicações e Artes. Universidade de São Paulo.

Aprovado em: 23/04/2021

BANCA EXAMINADORA

Profa. Dra. Cibele Araújo Camargo Marques dos Santos

Instituição: ECA – USP

Profa. Dra. Giovana Deliberali Maimone

Instituição: ECA – USP

Prof. Dr. Marivalde Moacir Francelin

Instituição: ECA – USP

Dedico a meus pais (*in memoriam*), irmãs e irmão.

Ao Raul, companheiro.

À Sophia, Eric e Caio, construindo sempre.

AGRADECIMENTOS

Agradeço à minha orientadora professora Cibele Araújo Camargo Marques dos Santos. Gratidão pela paciência e motivação, pelas orientações e discussões e pela preocupação com o conteúdo. Principalmente pela visão de que “[vai em frente,] se for fazer igual ao que já tem nunca vai existir coisa nova!”.

À minha família, meus pais (*in memoriam*), irmãs e irmão, por tudo a gratidão é eterna. E o Raul, pelo apoio contínuo e discussões de todas as horas, e sua maravilhosa família.

Aos amigos e colegas da Biblioteconomia e do SciELO, pela convivência e energia, meu muito obrigada, o faço nas figuras da Renata Baralle na ECA, do Wilson de Jesus em Marília, hoje Londrina, e de Francine Curivil no SciELO.

À revista Saúde e Sociedade (FSP) e ao grupo Arquigrafia (FAU/ECA/IME) sou mais do que grata pelo que vivenciei e aprendi na realização dos estágios e programas de bolsa.

Aos profissionais das bibliotecas que frequentei em toda a graduação, e com toda minha admiração, às bibliotecárias e bibliotecários. Em particular, às bibliotecas da ECA, FSP, FMUSP e FEA pelos computadores abertos, funcionamento noturno, acesso livre ao acervo, salas e cantos de estudo, meus agradecimentos.

Os limites da minha linguagem significam os limites do meu mundo.

Ludwig Wittgenstein

RESUMO

KANASHIRO, L. S. **Estudo de consistência entre as palavras-chave de artigos científicos de saúde com interface em ciências sociais e a indexação na base de dados LILACS**. São Paulo, 2021. Monografia (TCC) – Escola de Comunicações e Artes. Universidade de São Paulo.

A recuperação da informação em bases de dados e repositórios diversos pode decorrer de uma pesquisa utilizando termos em linguagem natural ou em linguagem controlada. Os autores de artigos científicos usam nas palavras-chave os assuntos que consideram representativos do estudo realizado e os profissionais da informação das bases de dados procuram atribuir na indexação dos artigos termos padronizados, com fins de maximizar a recuperação da informação em buscas realizadas na base. O objetivo deste trabalho é obter em que medida há concordância entre os termos em linguagem natural das palavras-chave de artigos em uma área interdisciplinar e os termos de indexação na linguagem controlada de uma base de dados especializada, através do cálculo de índices de consistência. Uma amostra de 20 artigos do periódico Saúde e Sociedade foram caracterizados com relação ao número de palavras-chave dos autores, número de descritores de sua indexação na base de dados LILACS com o vocabulário controlado DeCS, e número de termos coincidentes entre ambos para o estudo de consistência. Encontramos que o grau de concordância do total de palavras-chave em relação ao total de descritores da amostra é de 47,6% enquanto os índices de consistência variam entre 0 e 35,7%, com uma média de 13,2%.

Palavras-chave: Artigo científico, Palavra-chave, Indexação em base de dados, Índice de consistência, Descritores DeCS, LILACS.

ABSTRACT

KANASHIRO, L. S. **Consistency study between the keywords of scientific health articles with interface in social sciences and the indexing in LILACS database.** São Paulo, 2021. Monografia (TCC) – Escola de Comunicações e Artes. Universidade de São Paulo.

The retrieval of information in different databases and repositories can result from a search using terms in natural language or in controlled language. The authors of scientific articles use in the keywords the subjects they consider representative of the study carried out and the information professionals of the databases seek to assign standardized terms in the indexing of articles, in order to maximize information retrieval in searches carried out in the database. The objective of this work is to obtain to what extent there is agreement between the terms in natural language of the keywords of articles in an interdisciplinary area and the indexing terms in the controlled language of a specialized database, through the calculation of consistency indices. A sample of 20 articles from the journal Saúde e Sociedade were characterized in relation to the number of keywords of the authors, number of descriptors of their indexation in the LILACS database with the DeCS controlled vocabulary, and number of coincident terms between both for the study of consistency. We found that the degree of agreement of the total number of keywords in relation to the total number of descriptors in the sample is 47.6% while the consistency indexes vary between 0 and 35.7%, with an average of 13.2%.

Keywords: Scientific article, Keyword, Indexing in database, Consistency index, DeCS descriptors, LILACS.

LISTA DE SIGLAS

AL&C	América Latina e Caribe
BIREME	Centro Latino-americano de Informação em Ciências da Saúde
BRAPCI	Base de dados de Periódicos em Ciência da Informação
BVS	Biblioteca Virtual em Saúde
DeCS	Descritores em Ciências da Saúde
ISI	Institute for Scientific Information
LILACS	Literatura Latino-Americana e do Caribe em Ciências da Saúde
MEDLINE	Medical Literature Analysis and Retrieval System Online
MeSH	Medical Subject Headings
NLM	National Library of Medicine
OMS	Organização Mundial da Saúde
OPAS	Organização Pan-Americana de Saúde
SciELO	Scientific Electronic Library Online
UNIFESP	Universidade Federal de São Paulo
USP	Universidade de São Paulo
WoS	Web of Science

LISTA DE QUADROS

Quadro 1 – Modelo do quadro de registros no Excel dos artigos coletados.....	27
Quadro 2 – Resumo dos resultados obtidos para a amostra de 20 artigos	39

LISTA DE TABELAS

Tabela 1 – Formatação dos dados dos 20 artigos selecionados	31
Tabela 2 – Número de palavras-chave e de descritores e número de termos coincidentes por artigo e do conjunto	32
Tabela 3 – Frequência do número de palavras-chave atribuídas pelos autores	33
Tabela 4 – Frequência do número de descritores atribuídos pelos bibliotecários indexadores da base LILACS.....	34
Tabela 5 - Distribuição das frequências de coincidência entre palavras-chave e descritores..	35
Tabela 6 – Índices de consistência C entre as palavras-chave e os descritores	36
Tabela 7 – Distribuição de frequência dos índices de consistência	38

LISTA DE FIGURAS

- Figura 1 – Visualização gráfica, por ordem dos artigos, dos índices de consistência C (em %) entre as palavras-chave e os descritores.....37
- Figura 2 – Visualização gráfica da distribuição de frequência dos índices de consistência....38

SUMÁRIO

1	INTRODUÇÃO	15
2	FUNDAMENTAÇÃO	18
2.1	Indexação e recuperação de documentos	18
2.1.1	Consistência da indexação.....	21
2.2	O periódico Saúde e Sociedade	22
2.3	A base de dados LILACS.....	23
2.3.1	Vocabulário controlado DeCS.....	24
3	METODOLOGIA.....	26
3.1	<i>Corpus</i> de estudo.....	26
3.2	Coleta de dados	26
3.3	Tratamento dos dados.....	27
4	RESULTADOS E ANÁLISE.....	31
5	CONSIDERAÇÕES FINAIS.....	41
	REFERÊNCIAS.....	42
	APÊNDICE – <i>Corpus</i> do estudo	44

1 INTRODUÇÃO

A Organização e Representação de Documentos para fins de recuperação da informação tem um processo e um instrumento como pontos chave de sua atividade: a indexação e as linguagens documentárias.

Nos artigos científicos as palavras-chave, quando atribuídas em linguagem natural pelos autores para uma dada temática ou assunto de que trata o artigo, podem ser designadas de diferentes formas pelos pesquisadores, embora atuando dentro da mesma área de especialidade.

O bibliotecário, ao realizar a representação dos assuntos destes artigos para a indexação em uma base de dados, faz uso de uma linguagem documentária, uma ferramenta de controle de vocabulário que padroniza a linguagem de uma especialidade com uma terminologia única que represente os conceitos da área. Isso permite, na recuperação de uma pesquisa, localizar todos os documentos indexados sob os assuntos que constam no vocabulário controlado da base.

Este estudo consiste em comparar os termos que representam um documento na linguagem natural das palavras-chave e na linguagem controlada dos termos indexados de uma base de dados, analisando em que medida coincidem, através de índices de consistência.

É estabelecido em estudos de indexação de documentos (PINHEIRO, 1978; LEININGER, 2000; BOCCATO, 2012) que a recuperação da informação tem uma eficiência maior quando se aplica a linguagem, dita artificial, dos vocabulários controlados no tratamento da informação.

Nesses tempos de Covid-19, eficiência e urgência por informações aceleraram as mudanças no processo de produção e publicação dos artigos científicos, que são os instrumentos que prevalecem na comunicação dos resultados de pesquisa e discussões da comunidade científica. Normalmente é um processo que leva de meses a anos desde a submissão do manuscrito até a publicação do artigo. O surgimento da pandemia de Covid-19 em 2020 demandou que o acesso a eles fosse público e disponibilizado no menor tempo possível, acelerando a instituição dos repositórios de *preprints* de acesso aberto (MAJUMDER e MANDL, 2020).

Um artigo ao ser publicado pode compor um item de uma base de dados em que não é realizado o tratamento de análise conceitual da informação contida no documento, ou pode integrar uma base em que o documento passa pelo processo de indexação realizada por profissionais da informação, os bibliotecários. Estes, após realizarem a análise documentária, disponibilizam os assuntos de que trata o documento de forma padronizada para que entre milhares, o artigo esteja entre os itens recuperados em uma busca por informação específica de alguém com uma demanda informacional.

Em bancos ou bases de dados sem indexação, como são os repositórios, os bancos de dissertações e teses, o ArXiv¹, e bases como o SciELO², a Brapci³, Web of Science⁴ (WOS), como se dá a recuperação de um documento quando se digita um termo no campo de busca do sistema? Em geral, os pontos de acesso são o título, o resumo e as palavras-chave. Ou seja, a recuperação é feita pela linguagem natural, especializada, de determinada área.

No caso das bases indexadas, como são as bases LILACS⁵, Medline⁶, há uso de um vocabulário controlado, de onde indexadores profissionais selecionam os termos precisos, padronizados, que representem o documento. A recuperação então se dá pela linguagem controlada.

O contexto do tema deste trabalho ocorre em função de uma questão surgida na realização de um estágio no periódico Saúde e Sociedade⁷, em 2017. O periódico se situa na interface das áreas de Saúde e das Ciências Sociais e Humanas, e é indexado, entre outras, na base LILACS. Sendo a base especializada na área de saúde, como deveria ser a seleção dos descritores no vocabulário controlado da base, o DeCS? Os artigos do periódico trazem nas palavras-chave termos como “moradores de rua”, “mídia televisiva”, “fronteiras”, “identidades” entre outras. Haveria cobertura suficiente de termos no DeCS?

¹ Repositório digital pioneiro na disponibilização livre, sem revisão por pares, de artigos acadêmicos científicos desde 1991, mantida pela Universidade de Cornell: <https://arxiv.org/>

² Biblioteca digital de acesso aberto a textos científicos integrais, com revisão por pares, principal base de dados das publicações de periódicos científicos brasileiros, da América Latina e Caribe e países lusófonos: <https://scielo.org/>

³ Base de dados referencial nacional, especializada em Ciência da Informação: <https://www.brapci.inf.br/>

⁴ Base de dados referencial multidisciplinar, da Clarivate Analytics, com acesso pago, por assinatura.

⁵ Principal base de dados bibliográfica, de acesso aberto, da literatura científica e técnica latino-americana e do Caribe na área de Ciências da Saúde: <https://lilacs.bvsalud.org/>

⁶ Principal base de dados bibliográfica da Biblioteca Nacional de Medicina dos EUA (National Library of Medicine - NLM), especializada em ciências da vida e biomedicina. Acesso pelo PubMed, a base de dados referencial da NLM: <https://pubmed.ncbi.nlm.nih.gov/>

⁷ Disponível para acesso em: <http://www.scielo.br/sausoc>

Mais adiante, atuando como técnica em biblioteca no SciELO, a entrada em operação de um repositório de *preprints*, em 2020, quando do surgimento da Covid-19, a questão foi como os autores com os trabalhos depositados neste e em repositórios diversos conseguiriam que o artigo comparecesse na lista de itens recuperados em uma pesquisa relacionada à temática, dando sua contribuição nas discussões e resoluções?

Concomitante ao estágio no periódico Saúde e Sociedade, um trabalho em iniciação científica sobre consistência das palavras-chave dos autores dos artigos desse periódico e os termos de indexação na base LILACS, possibilitou estudos que dessem conta de questões como: se os termos do DeCS abrangiam a área de interface com as Ciências Sociais, se os autores indicavam nas palavras-chave termos “adequados” para a recuperação do documento, se os autores teriam essa consciência do papel das palavras-chave etc.

Ao final, focamos aqui em um objetivo: estudar a consistência entre as palavras-chave dos autores dos artigos do Saúde e Sociedade e sua indexação na base de dados LILACS. E com um objetivo específico de obter o quanto as palavras-chave dos autores coincidem com os descritores da base. Muito? Pouco? Mas de quanto? Para essa análise partimos do cálculo dos índices de consistência entre as palavras-chave e os descritores do DeCS de uma amostra de artigos do periódico, realizando uma adaptação de fórmulas de índice de consistência utilizadas encontradas na literatura.

2 FUNDAMENTAÇÃO

2.1 Indexação e recuperação de documentos

A recuperação da informação envolve o uso da linguagem natural e das linguagens documentárias na representação temática de um documento. Na indexação do documento, ambas as linguagens são formas de representar a informação contida nele, portanto, têm papel importante na comunicação científica. A recuperação de um documento pelos mecanismos de busca depende da qualidade do título, resumo e palavras-chave de um artigo, por exemplo (GARCIA, GATTAZ e GATTAZ, 2019). São elementos que auxiliam também a representação temática na indexação do documento em uma base de dados.

A indexação temática

busca identificar os assuntos contidos no texto de um documento, traduzindo-os para uma linguagem de indexação natural - palavras que ocorrem no texto - ou linguagem de indexação controlada - termos/descriptores autorizados em vocabulário controlado (SANTOS, 2019, p. 219).

A linguagem de indexação natural consiste nos termos do vocabulário do autor, como as palavras contidas no texto e, especificamente para este estudo, as palavras-chave dos artigos de periódicos.

Uma linguagem de indexação controlada traz os termos selecionados por profissionais da informação de um vocabulário controlado, após uma análise em que é levada em conta uma gama extensa de requisitos para sua seleção. Dessa forma o controle de vocabulário mostra-se como

[...] um recurso para a consistência do índice de assunto, produto gerado pelo processo de indexação dos artigos científicos e outros documentos com o uso de uma linguagem documentária. Este processo exige conhecimento das características do suporte, da tipologia documental e textual, da política e metodologia aplicadas, compreensão da área temática e do conteúdo documentários, regras claras sobre o processo de indexação e treinamento. (SANTOS, 2017, p. 3).

O vocabulário controlado pode ser definido, segundo Lancaster, como

[...] uma lista de termos autorizados. Em geral, o indexador somente pode atribuir a um documento termos que constem da lista adotada pela instituição para a qual trabalha. [...] Inclui, em geral, uma forma de estrutura semântica

[destinado a] 1. controlar sinônimos, optando por uma única forma padronizada, com remissivas de todas as outras; 2. diferenciar homógrafos. Por exemplo, PERU (país) é um termo bastante diferente de PERU (ave); e 3. reunir ou ligar termos cujos significados apresentem uma relação mais estreita entre si. Dois tipos de relações são identificados explicitamente: as hierárquicas e as não-hierárquicas (ou associativas). [...] (LANCASTER, 2004, p. 19).

Os termos autorizados, aqueles que encabeçam um assunto tendo seus sinônimos incluídos como remissivas para ampliar as possibilidades de recuperação do documento indexado, são os termos preferidos ou descritores.

Já as palavras-chave do autor, que constituem os assuntos principais tratados no documento, são “a indicação realizada pelos autores dos artigos de periódicos, em linguagem natural, embora se refiram à linguagem da especialidade temática da pesquisa, estas palavras pertencem ao vocabulário da língua” (SANTOS, 2017, p. 3).

Na definição do Dicionário de Biblioteconomia e Arquivologia de Cunha e Cavalcanti (2008, p. 274), palavra-chave é a "palavra significativa encontrada no título de um documento, no resumo ou no texto. Essa palavra (ou grupo de palavras) caracteriza o conteúdo temático do item e é usada em catálogos e índices de assunto".

A palavra-chave de autor, no entanto, é apresentada nos artigos e trabalhos acadêmicos pelos próprios autores, como indicado na definição acima, podendo ser utilizada para índices de assuntos em bases de dados.

Para a comunicação científica, que é o processo específico de produção, consumo e transferência de informação no campo científico segundo Cunha e Cavalcanti (2008, p. 97), a importância das palavras-chave é destacada pela sua finalidade, sendo que

A seleção de palavras-chave visa facilitar a recuperação eficiente do conteúdo de um texto para os leitores. Além disso, por serem ferramentas fundamentais para a indexação nas bases de dados, elas atuam como porta de acesso ao texto. Mesmo diante de todas as funcionalidades das palavras-chave, muitos autores subestimam essa etapa da escrita científica. Não fazem, por exemplo, uma reflexão mais acurada das melhores opções, reduzindo significativamente a possibilidade de seus documentos serem localizados. Essa conduta acarreta na sedimentação de seus textos na grande base submersa do iceberg das publicações (GARCIA, GATTAZ e GATTAZ, 2019, p. 6).

As funções das palavras-chave vão além, pois

(1) permitem que os leitores decidam se um artigo contém ou não material relevante aos seus interesses; (2) fornecem aos leitores termos adequados para uso em pesquisas na web para localizar outros materiais sobre os mesmos tópicos ou similares; (3) auxiliam os indexadores/editores a agrupar materiais relacionados em, digamos, edições de final de ano de um determinado periódico ou um conjunto de anais de conferência; (4) permitem que editores/pesquisadores documentem mudanças em uma disciplina ao longo do tempo [...]; e (5) vinculam questões específicas de interesse a questões em níveis mais elevados de meta (abstração) (HARTLEY e KOSTOFF, 2003, p. 433, tradução).

Um exemplo de indexação pela linguagem natural é a que encontramos na base de dados Web of Science. Trata-se de uma base multidisciplinar e o sistema busca os termos das expressões digitadas no campo de busca, em TÓPICO, no título, resumo, palavras-chave e *keywords plus*⁸, ou seja, o termo digitado deve corresponder exatamente ao termo utilizado no texto, lembrando, porém, que essa limitação é contornada pelo recurso de truncagem de palavras para auxiliar na pesquisa.

A base de dados MEDLINE é um exemplo de indexação pela linguagem controlada. É uma base especializada, possibilitando uma indexação objetiva, mais precisa, por meio de uma linguagem controlada. Os descritores são provenientes do vocabulário controlado MeSH (Medical Subject Headings).

A atividade de indexação de assunto pelo profissional bibliotecário, segundo Lancaster (2004), consiste em duas etapas principais: a análise conceitual e a tradução. Strehl (1998, p. 329) resume: “a análise conceitual é a atividade de definição dos assuntos que são tratados no documento, e a tradução corresponde à atividade de conversão dos conceitos identificados na análise para uma linguagem de indexação”.

Essas etapas têm efeito na coerência da indexação (LANCASTER, p. 74), a qual será tratada no próximo tópico.

Por ser uma atividade cognitiva, altamente subjetiva (PINHEIRO, 1978), a atividade de indexação abrange um grau de conhecimentos e uso de documentação com critérios estabelecidos pela base de dados para sua realização.

⁸ “KeyWords Plus® são termos de índice gerados automaticamente a partir dos títulos de artigos citados. [...] O KeyWords Plus aumenta o número de resultados tradicional de palavras-chave ou títulos.” (http://images.webofknowledge.com/WOKRS517B4/help/pt_BR/WOS/hp_full_record.html). Acesso em: 20 fev. 2021).

Lancaster (2004) menciona que uma boa indexação é “a indexação que permite que se recuperem itens de uma base de dados durante buscas para as quais sejam respostas úteis, e que impede que sejam recuperados quando não sejam respostas úteis” (p. 83), e ainda, a qualidade como sendo “definida em termos de eficácia de recuperação – a capacidade de recuperar o que é desejado e de evitar o que não é desejado” (p. 92). Strehl (1998) destaca

a importância do uso de uma política de indexação e de um vocabulário controlado para nortear as atividades do indexador no momento da representação temática dos documentos. A política de indexação, neste contexto, orienta a indexação de acordo com as necessidades de informação dos usuários de um determinado sistema; já o vocabulário controlado estabelece a forma de representar os assuntos que compõem uma área limitada do conhecimento, tornando possível maior coerência entre os termos indexados (STREHL, 1998, p. 330).

Coerência diz respeito à recuperação de documentos que tratam de um mesmo assunto. Se diferentes documentos de mesmo assunto forem indexados com os mesmos termos, eles estarão presentes entre os itens recuperados em uma busca. Para um documento, coerência na indexação, segundo Lancaster (2004) “refere-se à extensão em que há concordância quanto aos termos a serem usados para indexar o documento” (p. 68), e que se este for indexado com diferentes termos por diferentes indexadores (indexação interindexadores) ou pelo mesmo indexador em ocasiões distintas (indexação intraindexador), haverá inconsistência na indexação.

Sendo o foco deste estudo, conceituaremos na próxima seção o termo “consistência de indexação”.

2.1.1 Consistência da indexação

A consistência é um dos elementos que permite avaliar a qualidade da indexação (Gil Leiva, 2002), estando relacionada ao grau de concordância dos termos indexados, expresso por um índice de consistência.

Consistência da indexação, na definição de Gil-Leiva (2002), representa a

[...] busca de semelhanças durante a atribuição de palavras-chave, assuntos ou descritores a um documento. Quando a comparação é feita entre os resultados de um indexador em sua análise de um mesmo documento em períodos diferentes, é conhecida como consistência intraindexador, já, se a comparação é feita entre os resultados de vários indexadores na análise de um

mesmo documento, é denominada consistência interindexadores (GIL-LEIVA, 2002, p. 99, tradução).

Essa ideia foi exposta por Lancaster (2004) em termos de coerência da indexação, como mencionado anteriormente. Lancaster também chama a atenção de que “indexação coerente não é necessariamente o mesmo que indexação de qualidade” (LANCASTER, 2004, p. 82).

Leonard (1977) afirma que “o propósito dos testes de consistência é obter *insights* acerca de métodos que aumentem a eficácia da recuperação [pois] este é o ponto mais importante em qualquer sistema de recuperação” (p. 11, traduzido).

O índice de consistência varia numericamente entre 0,00 e 1,00 (ou 0 e 100%), obtidos a partir dos termos (no original é usado “*codes, keywords and subjects*”) atribuídos pelos indexadores (Leonard, 1977) e calculado por meio de fórmulas variadas. Pinheiro (1978, p. 109), em referência à subjetividade envolvida na atividade da indexação, coloca que a indexação “oscila muito no seu nível de concordância e apresenta discrepâncias”, sendo praticamente impossível obter 100% de consistência, mencionando trabalhos de autores em que os resultados variaram entre 10 e 80%, e observando que “a consistência é consideravelmente aumentada quando são adotadas tabelas de classificação, *thesauri* ou restrições no uso dos termos”.

Para essa questão da impossibilidade de se obter consistência absoluta, retomando o que diz Leonard (1977), o autor reforça:

A consistência entre indexadores é uma medida quantitativa do grau em que dois ou mais indexadores percebem os conceitos de informação importantes contidos em um documento e representam esses conceitos usando códigos e/ou termos idênticos. A razão de ser da indexação de documentos é, no entanto, facilitar a recuperação de documentos relevantes para as consultas de informação feitas pelos usuários da base documental. O objetivo da indexação não é necessariamente atingir 100 por cento de consistência entre um grupo de indexadores, mas recuperar documentos que são relevantes para as solicitações de pesquisa (LEONARD, 1977, p. 2, tradução).

2.2 O periódico Saúde e Sociedade

Para o estudo de consistência entre palavras-chave de artigos e descritores de uma base, o *corpus* de análise consistiu nos artigos publicados no periódico Saúde e Sociedade, que integra a coleção de periódicos da base SciELO (Scientific Electronic Library Online) e está

indexada na base de dados LILACS (Literatura Latino-Americana e do Caribe em Ciências da Saúde).

O periódico Saúde e Sociedade foi criado há 29 anos, em 1992, pela Faculdade de Saúde Pública da USP e pela Associação Paulista de Saúde Pública de São Paulo, parceria mantida até hoje, e tem como escopo a publicação de temas em saúde pública/coletiva na interface com as ciências sociais e humanas. Em 1994 foi indexada na base de dados LILACS e após quase duas décadas de desafios para se consolidar, passou a ser indexada em importantes bases nacionais e internacionais: Cambridge Sociological Abstracts (CSA), Sociological Abstract e Social Services Abstract em 2006; SciELO; Scisearch, Journal Citation Reports/Social Sciences Edition; Web of Science (antiga base ISI) da Thomsom Reuters, atual Clarivate Analytics; Ulrich's International Periodical Directory; Latindex, todos em 2008; EBSCO Publishing em 2009; Library Of Congress Cataloging e base Scopus em 2010 (MARTINS *et al.*, 2015).

O periódico, inicialmente indexado na área de saúde, com o tempo passou a ser indexada em bases de dados da área de Ciências Sociais, o que reforça a característica interdisciplinar do periódico.

2.3 A base de dados LILACS

A LILACS⁹ é uma base de dados especializada em ciências da saúde, disponibilizando com acesso aberto a literatura científica e técnica como artigos, teses e dissertações, livros, capítulos de livros, anais de congressos, publicações governamentais entre outras, produzida pelos países da América Latina e do Caribe (AL&C).

Desenvolvida e coordenada pela BIREME (Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde, antiga Biblioteca Regional de Medicina) em conjunto com uma rede de instituições da AL&C, a base LILACS foi disponibilizada em 1982, constituindo-se em uma das principais fontes de informação da BVS, a Biblioteca Virtual em Saúde (SANTOS, 2010). Atualmente, segundo os dados na página, indexa 903 periódicos de 26 países, com 944 mil registros, sendo 528 mil disponibilizados em texto completo. A indexação

⁹ Disponível para acesso em: <https://lilacs.bvsalud.org/>

de um item é feita utilizando o vocabulário controlado DeCS (a ser tratado no próximo tópico) no nível mais específico possível, atribuindo “quantos descritores forem necessários para descrevê-lo em todos seus aspectos”¹⁰ e “segue o modelo do MEDLINE, utilizando descritores primários e secundários, campo para tipo de publicação, campo para limites de idade, tipo de estudo e gênero e campo para geográfico [entre outros]” (SANTOS, 2010, p. 77-78).

As publicações são indexadas de forma descentralizada por uma rede de profissionais da informação de 958 centros cooperantes de 37 países da AL&C, coordenada pela BIREME/OPAS/OMS¹¹. No Brasil, participam da rede 260 instituições como: Centro de Documentação, da OPAS/OMS Brasil, CIR - Biblioteca - Centro de Informação e Referência, da Faculdade de Saúde Pública da USP, Biblioteca Central, da Escola Paulista de Medicina da UNIFESP, bibliotecas da Fundação Oswaldo Cruz e de outros centros e universidades do país.

Para assegurar a qualidade da indexação, a LILACS possui política de indexação bem detalhada, disponibilizando guias, manuais, notas técnicas e aplicativos para a atividade e oferece cursos de capacitação para a adoção da metodologia e o uso e domínio das ferramentas por todas as bibliotecas da rede BIREME que participam da indexação dos itens da base¹².

2.3.1 Vocabulário controlado DeCS

Criado pela BIREME, o vocabulário DeCS - Descritores em Ciências da Saúde (<https://decs.bvsalud.org/>) é multilíngue (português, inglês, espanhol e francês), e conforme descrição na página¹³, ele padroniza em uma linguagem única tanto termos a serem utilizados na indexação de artigos de periódicos científicos e demais itens (livros, anais de congressos, relatórios técnicos etc.) como também na pesquisa e recuperação de assuntos da literatura científica da área biomédica nas fontes de informação disponíveis na BVS, como por exemplo LILACS e MEDLINE.

O DeCS foi desenvolvido a partir do MeSH, o vocabulário controlado criado pela National Library of Medicine (NLM), com inclusão de categorias, conforme a página do DeCS,

¹⁰ Disponível em: <https://lilacs.bvsalud.org/metodologia-lilacs/manual-de-indexacao-de-documentos-para-a-base-de-dados-lilacs/#Filosofia>. Acesso em: 12 fev. 2021.

¹¹ Disponível em: <http://red.bvsalud.org/lilacs/pt/sobre-a-lilacs/>. Acesso em: 20 mar. 2021.

¹² Disponível em: <http://red.bvsalud.org/lilacs/pt/sobre-a-lilacs/guias-manuais-e-notas-tecnicas/>. Acesso em: 12 fev. 2021.

¹³ Disponível em: <https://decs.bvsalud.org/sobre-o-decs/>. Acesso em 12 fev. 2021.

“especialmente desenvolvidas para melhor representar a literatura gerada nos países da região [AL&C]”: Saúde Pública, Homeopatia, Ciência e Saúde e Vigilância Sanitária.

Os termos no DeCS estão organizados em uma estrutura hierárquica com grande nível de especificidade e são distribuídos por 20 categorias como Anatomia, Fenômenos e Processos, Ciências Humanas, Assistência à Saúde, para exemplificar (são 16 categorias MeSH e as 4 exclusivas DeCS). Os descritores são termos estabelecidos e definidos a partir de suas características, usos e significado. Cada descritor traz a definição em nota de escopo, fornece dicas em nota de indexação de como deve ser indexado, traz os sinônimos como termos alternativos e uma lista de qualificadores permitidos, que são termos que delimitam um aspecto do assunto, contextualizando a pesquisa (Ex.: Sistema Único de Saúde/história, Sistema Único de Saúde/organização & administração; Obesidade/classificação, Obesidade/cirurgia). São quase 35 mil descritores e qualificadores (34.041 e 77, respectivamente) disponíveis para uso na indexação dos itens e em pesquisas na base de dados LILACS. Nos 4 idiomas e incluídos os termos alternativos, totalizam mais de meio milhão de termos¹⁴.

A recuperação da pesquisa na LILACS traz um destaque nas palavras do texto dos resumos que são descritores ou termos alternativos do DeCS, com um recurso de abrir um *pop-up* com a definição do DeCS ao se passar o mouse sobre o termo, havendo opção de visualizar a página do descritor pelo link na janela. Cada descritor do campo de assuntos do artigo também traz o recurso e ao clicá-lo é listado os artigos que contêm o termo no resumo ou campo de assuntos.

¹⁴ Disponível em: <https://decs.bvsalud.org/P/visaogeral2020.htm>. Acesso em 13 fev. 2021.

3 METODOLOGIA

Para o estudo de consistência entre as palavras-chave dos autores dos artigos do periódico Saúde e Sociedade e sua indexação na base LILACS, utilizamos uma amostra constituindo o *corpus* descrito a seguir.

3.1 *Corpus* de estudo

O *corpus* consiste em um conjunto de 20 artigos do periódico Saúde e Sociedade, indexado na base LILACS, selecionados de um período de 3 anos e utilizando o critério de presença do maior número de descritores atribuídos na indexação nesta base.

O universo dos artigos considerado para o estudo são os números publicados pelo periódico em 2013, 2014, 2015, levantados em projeto de Iniciação Científica em 2017. Foram coletados 338 artigos no total, sendo 103 artigos de 2013, 115 de 2014 e 120 de 2015, dos quais foram extraídos a amostra.

3.2 Coleta de dados

As palavras-chave de cada artigo publicado entre 2013 e 2015 foram extraídas de tabelas disponíveis na editora publicadora do periódico durante o estágio obrigatório do curso de Biblioteconomia realizado no local, organizadas e conferidas com os respectivos artigos na base SciELO.

Para a obtenção dos termos de indexação na LILACS foram feitas tentativas de compilação com os artigos salvos no gerenciador de referências EndNote Web, ou seja, utilizando um método mais automático. Com os testes não satisfatórios e o tempo decorrido já próximo à preparação do primeiro relatório de IC, os fascículos de cada ano foram pesquisados no site da base e os descritores extraídos copiando-os artigo por artigo, organizando os dados em uma tabela Excel, ao lado das palavras-chave.

O quadro de registros, exemplificado no Quadro 1, contém, para cada item, as informações: número do artigo (representado pelo símbolo #), dados de localização (volume e

número), título do artigo, palavras-chave utilizadas pelos autores, termos indexados na base LILACS correspondentes aos descritores do vocabulário controlado DeCS, identificados na planilha como Descritores Lilacs.

Quadro 1 – Modelo do quadro de registros no Excel dos artigos coletados

Item #	Vol. N.	Título do Artigo	Palavras-chave do autor	Descritores Lilacs
1	V.22 N.1 [2013]	A produção do conhecimento na interface entre as ciências sociais e a saúde pública/coletiva [editorial especial]	NT	Ciências Sociais Ciências Humanas Saúde Pública
2	V.22 N.1	Ciências sociais e saúde pública/coletiva: a produção do conhecimento na sua interface	Ciências Sociais e Disciplinas da Saúde; Conhecimento Empírico e Conhecimento Militante	Ciências Sociais Saúde Pública Conhecimentos, Atitudes e Prática em Saúde
3	V.22 N.1	A produção de conhecimentos na interface entre as ciências sociais e humanas e a saúde coletiva	Ciências Sociais e Saúde; Ciências Humanas e Saúde; Saúde Coletiva; Teoria Social	Ciências Sociais Ciências Humanas Saúde Pública
4	V.22 N.1	Temas emergentes em ciências sociais e saúde pública/coletiva: a produção do conhecimento na sua interface	Ciências Sociais; Saúde Pública; Saúde Coletiva; Eixos de Pesquisa; Temas Emergentes; Desafios	Ciências Sociais Saúde Pública Política de Saúde Serviços de Saúde Condições Sociais Condições de Trabalho Reforma dos Serviços de Saúde Recursos Humanos em Saúde Publicações Científicas e Técnicas Comunicação e Divulgação Científica -Pesquisa

Fonte: elaboração própria

Selecionou-se os artigos que visualmente continham o maior número de descritores atribuídos na indexação na LILACS e destes foram extraídos os 20 artigos que iriam passar pela análise de consistência. Os dados dessa amostra, que consiste no *corpus* do estudo de consistência, encontram-se detalhados no Apêndice.

3.3 Tratamento dos dados

Na análise quantitativa da consistência entre as palavras-chave e a indexação da base LILACS foram contabilizados, para cada artigo, o número de palavras-chave, o número de descritores e quantas palavras-chave coincidiam com os descritores, contagens necessárias para o cálculo do índice de consistência.

Foram considerados para a obtenção do índice os pontos expostos a seguir.

Na contagem dos termos coincidentes entre palavras-chave e descritores consideramos apenas os termos exatamente coincidentes entre palavra-chave e descritor. Assim, não foram consideradas na contagem diferenças entre palavras-chave/descriptores como, por exemplo: Atenção Primária em Saúde/Atenção Primária à Saúde, Relações de Trabalho/Relações Trabalhistas, Alta Hospitalar/Alta do Paciente. Notamos que os termos têm uma similaridade, porém para certificar se podemos tratá-los como coincidentes em algum grau, seria necessário verificar o uso dos termos na literatura, comparando com a definição dos termos no DeCS. Como não foram encontradas ocorrências em quantidade que pudesse impactar nos resultados, não realizamos uma verificação minuciosa termo a termo e esses casos não foram contabilizados.

Um segundo ponto é que para a fórmula do índice de consistência de cada artigo, partimos da análise das expressões de cálculo da consistência entre dois indexadores (consistência interindexação) utilizadas por Lancaster (2004, p. 68) e Gil-Leiva (2002, p. 100):

Lancaster:

$$PC = \frac{AB}{A + B}$$

onde:

PC = valores de coerência de pares de indexadores

(AB) = #termos coincidentes na indexação pelos dois indexadores

(A+B) = #termos atribuídos pelo indexador *a* + #termos atribuídos pelo indexador *b*

Gil-Leiva:

$$C = \frac{T}{(A + B) - T}$$

onde:

C = Consistência entre dois indexadores

T = #termos coincidentes atribuídos pelos dois indexadores

(A+B) = #termos atribuídos pelo indexador *a* + #termos atribuídos pelo indexador *b*

Com base no entendimento dos termos e expressões utilizados por esses autores, aplicamos a lógica da interindexação para este trabalho, a fim de obtermos uma expressão de cálculo de consistência entre palavras-chave do autor e descritores do DeCS. Adaptamos as expressões, considerando o autor do artigo como um dos indexadores (que atribui termos na linguagem natural) e o bibliotecário da base LILACS como o outro indexador (que atribui termos com linguagem controlada).

Primeiro traduzimos as denominações dos termos para este caso:

C = consistência entre as palavras-chave do autor e os descritores na LILACS

A = número de palavras-chave atribuído pelos autores *a*

B = número de descritores atribuídos na indexação LILACS pelo bibliotecário indexador *b*

AB = número de palavras-chave coincidentes com descritores

(A+B) = número de termos de *a* + número de termos de *b* distintos de *a*.

Com isso a consistência C entre as palavras-chave do autor (Pch) e os descritores (Descr) na LILACS, para cada artigo, aplicada à expressão:

$$C = \frac{AB}{A + B}$$

é traduzida para o nosso caso como sendo:

$$C = \frac{\text{\#Pch coincidentes com Descr}}{\text{\#Pch de a} + \text{\#Descr exclusivos de b}}$$

Fazendo uma analogia com os operadores booleanos AND e OR, visualizamos que o quanto os termos palavras-chave/descriptores coincidem (análogo à operação “A AND B”) é um percentual do conjunto todo de termos (análogo à operação “A OR B”).

Simplificando, condensamos as nomenclaturas e obtemos a expressão com a qual serão calculados os índices de consistência para cada artigo deste estudo:

$$C = \frac{\#Coinc}{\#Conj} \quad (1)$$

em que:

C é a consistência entre as palavras-chave do indexador autor e os descritores na base LILACS do indexador bibliotecário;

#Coinc é o número de palavras-chave coincidentes com os descritores;

#Conj é o número de termos distintos do conjunto de palavras-chave e descritores.

4 RESULTADOS E ANÁLISE

Os artigos coletados dos três anos de publicação geraram uma lista de 338 itens dos quais, numa primeira análise, 70 foram destacados como os que apresentavam, visualmente, maior número de descritores. Refinando a seleção, realizando uma contagem dos descritores, extraímos os 20 artigos finais, tabelados conforme o formato apresentado na Tabela 1: número do item da amostra, de 1 a 20; número do artigo da lista inicial de 338 artigos coletados; localização do artigo publicado; título do artigo; palavras-chave do artigo; termos do DeCS atribuídos por bibliotecários na indexação na base LILACS. As palavras-chave do autor e os descritores do DeCS que coincidem estão com uma marca de destaque. A tabela completa encontra-se no Apêndice.

Tabela 1 – Formatação dos dados dos 20 artigos selecionados

Item#	Artigo#	Vol. N.	Título do Artigo	Palavras-chave do autor	Descritores Lilacs
1	13	V.22 N.1	Dimensões da qualidade de vida de idosos moradores de rua do município de São Paulo	1. Idoso; 2. Qualidade de Vida; 3. Moradores de Rua	1. Idoso 2. Habitação para Idosos 4. Instituição de Longa Permanência para Idosos 5. Autonomia Pessoal 6. Política de Saúde 7. Política Social 8. Saúde do Idoso 9. Assistência à Saúde 10. Vulnerabilidade Social 11. Relações Familiares 12. Qualidade de Vida 13. Pesquisa Qualitativa

Item#: número do item da amostra; Artigo#: número do item no total de artigos coletados; Vol. N.: localização do artigo publicado. Fonte: elaboração própria.

Os dados numéricos para os 20 artigos da amostra estão organizados na Tabela 2.

Tabela 2 – Número de palavras-chave e de descritores da amostra e número de termos coincidentes e do conjunto

Item#	#Pch	#Descr	#Coinc	#Conj
1	3	12	2	13
2	2	12	1	13
3	3	11	1	13
4	6	13	5	14
5	3	11	2	12
6	4	16	2	18
7	3	14	3	14
8	3	13	2	14
9	3	16	3	16
10	3	13	0	16
11	4	14	0	18
12	3	11	2	12
13	6	14	2	18
14	3	13	2	14
15	3	12	2	13
16	4	14	1	17
17	6	15	4	17
18	6	14	3	17
19	3	14	0	17
20	6	16	2	20
TOTAL	82	268	39	

#Pch: número de palavras-chave do autor; #Descr: número de descritores LILACS; #Coinc: número de termos coincidentes palavras-chave/descriptores; #Conj: número de termos distintos do conjunto palavras-chave/descriptores. Fonte: elaboração própria.

A tabela mostra, para cada artigo (Item#), o número de palavras-chave atribuído pelos autores (#Pch), o número de descritores atribuídos na indexação na base LILACS (#Descr) e o número de termos coincidentes entre ambos (#Coinc). Para fins de cálculo dos índices de consistência exigido na fórmula (1), também foi incluída uma última coluna referente ao número total de termos distintos do conjunto palavras-chave e descritores (#Conj). As colunas utilizadas para o cálculo dos índices de consistência foram destacadas em cinza.

Na amostra o número total de palavras-chave atribuído pelos autores foi de 82. O número de descritores da LILACS foi de 268 no total. Coincidiram 39 termos, o que representa 47,56% das palavras-chave e 14,55% dos descritores. Deve-se levar em conta que o resultado para os descritores foi limitado pelo número de palavras-chaves admitidas pelo periódico.

O periódico orientava à época que o artigo contivesse entre 2 e 6 palavras-chave. Na Tabela 3 é apresentada a frequência com que cada uma dessas quantidades comparece na amostra de artigos. As linhas preenchidas em amarelo nas tabelas destacam a maior ocorrência.

Tabela 3 – Frequência do número de palavras-chave atribuídas pelos autores

# Palavras-chave	Frequência absoluta (# de artigos)	Frequência relativa (%)
2	1	5
3	11	55
4	3	15
5	0	-
6	5	25
Total	20	100

Nota: a linha em amarelo destaca a maior ocorrência. Fonte: elaboração própria.

A maioria dos autores (de 11 artigos ou 55% dos artigos) incluíram 3 palavras-chave ao submeterem o artigo ao periódico. Dos 20 artigos da amostra, 5 (25% dos artigos) indicaram 6 palavras-chave e apenas 1 indicou o mínimo de palavras-chave requerida.

A média de atribuição de número de palavras-chave por artigo é 3,85.

Na indexação da base LILACS o número de termos do vocabulário DeCS atribuídos a estes artigos variam de 11 a 16. Na Tabela 4 podemos ver a frequência do número de descritores na amostra.

Tabela 4 – Frequência do número de descritores atribuídos pelos bibliotecários indexadores da base LILACS

# Descritores	Frequência absoluta (# de artigos)	Frequência relativa (%)
11	3	15
12	3	15
13	4	20
14	6	30
15	1	5
16	3	15
Total	20	100

Nota: a linha em amarelo destaca a maior ocorrência. Fonte: elaboração própria.

Em 6 artigos (30% da amostra) foram atribuídos 14 descritores, em 3 artigos encontramos o máximo de termos atribuídos da amostra (16 descritores) e apenas 1 artigo contém 15 descritores. Os demais artigos contêm de 11 a 13 descritores.

A média de atribuição pela base LILACS de número de descritores por artigo foi de 13,40.

Quanto ao número de termos coincidentes entre palavras-chave e descritores, este varia entre 0 e 5, distribuídos entre os 20 artigos conforme a Tabela 5.

Tabela 5 - Distribuição das frequências de coincidência entre palavras-chave e descritores

# Termos coincidentes (0 = nenhuma ocorrência)	Frequência absoluta (# artigos)	Frequência relativa (%)
0	3	15
1	3	15
2	9	45
3	3	15
4	1	5
5	1	5
Total	20	100

Nota: a linha em amarelo destaca a maior ocorrência. Fonte: elaboração própria

Em quase metade da amostra, 9 dos artigos (45%), ocorreram coincidência de 2 termos entre palavras-chave e descritores. Em 3 artigos não houve termos DeCS presentes nas palavras-chave. O maior número de termos coincidentes, 5 termos, ocorreu em apenas 1 artigo. Esse artigo, o de número 4, tem 6 palavras-chave e 13 descritores (Tabela 2).

A média de termos coincidentes por artigo é 1,95.

Para uma amostra de 20 artigos do periódico, em que foram constatados 82 palavras-chave, 268 descritores e 39 termos coincidentes, é significativo que quase metade (47,6%) das palavras-chave atribuídas pelos autores são termos do vocabulário controlado DeCS?

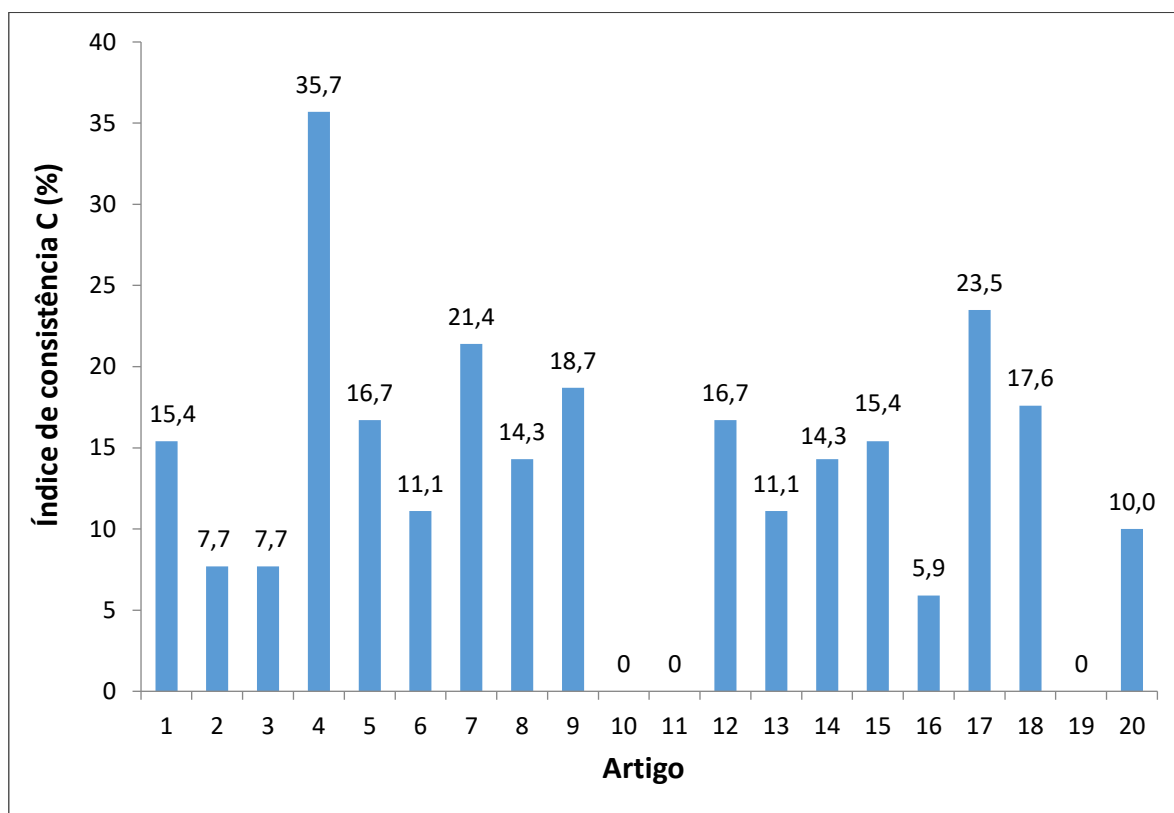
Foram calculados os índices de consistência C para cada artigo, dividindo-se a penúltima coluna da Tabela 2 (#Coinc) pela última coluna (#Conj), conforme a fórmula (1) e os resultados foram dispostos em detalhe na Tabela 6 e visualizados graficamente, em porcentagem, na Figura 1.

Tabela 6 – Índices de consistência C entre as palavras-chave e os descritores

Item#	#Coinc / #Conj	C	C (%)
1	2/13	0,154	15,4
2	1/13	0,077	7,7
3	1/13	0,077	7,7
4	5/14	0,357	35,7
5	2/12	0,167	16,7
6	2/18	0,111	11,1
7	3/14	0,214	21,4
8	2/14	0,143	14,3
9	3/16	0,187	18,7
10	0/16	0	0
11	0/18	0	0
12	2/12	0,167	16,7
13	2/18	0,111	11,1
14	2/14	0,143	14,3
15	2/13	0,154	15,4
16	1/17	0,059	5,9
17	4/17	0,235	23,5
18	3/17	0,176	17,6
19	0/17	0	0
20	2/20	0,100	10,0

#Coinc=número de termos coincidentes palavras-chave/descriptores; #Conj=número de termos distintos do conjunto palavras-chave/descriptores; C=índice de consistência.
Fonte: elaboração própria.

Figura 1 – Visualização gráfica, por ordem dos artigos, dos índices de consistência C (em %) entre as palavras-chave e os descritores



Fonte: elaboração própria.

Os valores de índices variam entre 0 e 35,7%, apenas em um artigo foi obtido um índice de consistência acima de 25%, que foi de 35,7, a maior parte dos valores do índice está localizada na faixa entre 10 e 20% e em 3 artigos não houve nenhuma concordância de palavras-chave e descritores.

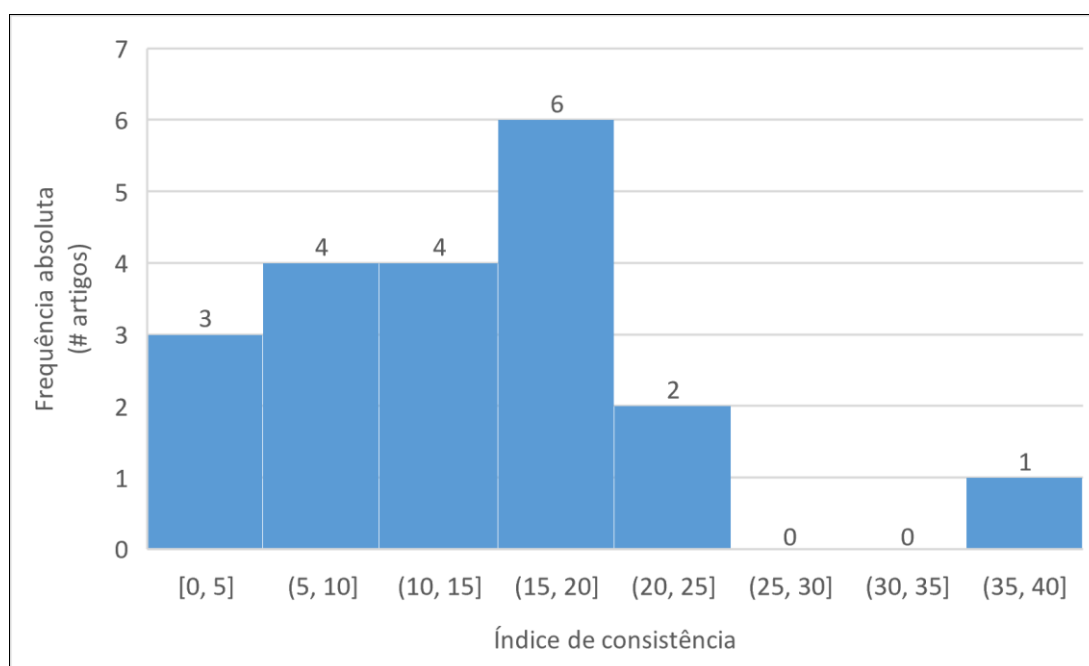
Para melhor análise, realizamos uma contagem por faixa dos valores obtidos, e organizamos numa tabela. A Tabela 7 mostra com que frequência os valores de índice de consistência aparecem na amostra, e a Figura 2, a visualização gráfica da distribuição de frequência. Em vermelho, na tabela, é destacada a faixa de predominância dos índices.

Tabela 7 – Distribuição de frequência dos índices de consistência

Índice de consistência C (%)	Frequência absoluta (# artigos)	Frequência relativa (%)
0 – 5]	3	15
5 – 10]	4	20
10 – 15]	4	15
15 – 20]	6	35
20 – 25]	2	10
25 – 30]	0	-
30 – 35]	0	-
35 – 40]	1	5
Total	20	100

Nota: a linha em amarelo destaca a maior ocorrência; a borda vermelha destaca a faixa de predominância dos valores. Fonte: elaboração própria.

Figura 1 – Visualização gráfica da distribuição de frequência dos índices de consistência



Fonte: elaboração própria.

Na faixa entre 10 e 20, em que constatamos visualmente na Figura 1 que é a faixa onde está localizada a maior parte dos valores dos índices de consistência, verificamos que se trata de metade da amostra, são 10 artigos nessa faixa de valores.

Verificamos também que os índices que aparecem com mais frequência estão no intervalo de 15-20 (em 6 artigos), portanto a moda é 17,5. A mediana, obtida no Excel, é 14,3. O que significa que o índice de consistência representativo da amostra é 17,5% e que metade dos índices estão abaixo de 14,3% e metade acima desse valor.

A média do índice de consistência da amostra, obtida dividindo-se a soma dos índices individuais da Tabela 6 pelo número total de artigos da amostra, nos dá:

$$\frac{2,632}{20} = 0,132, \text{ ou seja, } 13,2\%$$

No Quadro 2 encontra-se agrupado os resultados levantados nesse estudo para a amostra de 20 artigos.

Quadro 2 – Resumo dos resultados obtidos para a amostra de 20 artigos

Variável	Quantidade por artigo	Média	Total
Palavras-chave	2 a 6	3,8	82
Descritores	12 a 16	13,4	268
Termos coincidentes	0 a 5	1,9	39
	Valores calculados		
Índice de consistência	entre 0 e 35,7%	13,2%	—

Fonte: elaboração própria.

De um total de 82 palavras-chave nos 20 artigos, obtivemos que 39 (47,6%) coincidem com os descritores. Encontramos nos trabalhos de Santos (2019) e Miguéis e Neves (2013), que o percentual chega a 59%, e em Gil-Leiva e Alonso-Arroyo (2005), 60%, sendo que os dois últimos analisaram um universo maior de registros e levando em consideração os termos quase coincidentes, adotando uma escala de pontuação nas contagens. Apesar de quase metade das palavras-chave dos autores coincidirem com os termos autorizados do DeCS neste estudo, o índice de consistência do conjunto foi baixo, de 13,2%. A literatura mostra índices variando

de 10 a 80% (PINTO, 2001) e a média dos índices oscilando entre 25 e 60% (GIL-LEIVA, 2012).

Analisando o resultado, o índice de consistência, usando a fórmula (1), foi calculado pontualmente, para cada artigo, enquanto para o conjunto de artigos foi feita uma média dos índices individuais. Lancaster já havia colocado que “um vocabulário controlado deve melhorar a coerência da indexação em relação a um grupo de documentos, mas é bem possível que a diminua no nível de um único documento” (LANCASTER, 2004, p. 74).

Também podemos considerar a possibilidade de os indexadores da base, na tradução dos conceitos levantados, não encontrarem descritores suficientes no vocabulário controlado que correspondam a eles (PINTO, 2001). Borba (2012, p. 30) coloca a problemática da atualização dos sistemas de recuperação da informação, uma vez que existe um processo natural de modificação da linguagem dos especialistas conforme a evolução da ciência, que dependendo da área de conhecimento se dá em maior ou menor velocidade, sendo necessário monitorar os discursos dos especialistas e efetuar atualizações no sistema com maior ou menor constância e a área das ciências da saúde, em particular, é uma área cada vez mais interdisciplinar, conforme Meneghini (2010):

As ciências da saúde representam atualmente uma área científica de ponta no mundo, devido à sua crescente presença em distintos campos de conhecimento, desde as chamadas ciências duras até a medicina, epidemiologia, sociologia e ciências políticas. Não surpreendentemente, as bases de dados científicos enfrentam dificuldades em classificar as ciências da saúde em categorias de conhecimento mais amplas, como ciência ou sociologia (MENEGHINI, 2010, p. 1058).

Além disso, o limite de número de palavras-chave na submissão do artigo pode ser uma fonte de dificuldade para que o autor expresse todos os assuntos de que trata no artigo. O que pode limitar a cobertura de análise do indexador por meio das palavras-chave do autor, podendo alguns conceitos passar despercebidos ou serem indexados sem necessidade ou indevidamente. Por sua vez, os autores, apesar de identificarem os conceitos importantes e relevantes do artigo, na transformação destes em palavras-chave podem estar representando incorretamente os assuntos para fins de recuperação da informação (BORBA, 2012).

5 CONSIDERAÇÕES FINAIS

A consistência entre as palavras-chave dos artigos do periódico e a indexação na base LILACS mostrou-se baixa, com um valor de índice de consistência calculado, para o grupo de análise, de 13,2%. O valor mais alto foi de 35,7% obtido para um único artigo entre os 20 artigos da amostra; em três deles não houve nenhuma concordância de termos (índice zero). Avaliamos que o estudo sugere que para as publicações na interface da Saúde e Ciências Humanas pode existir um vácuo que pode estar na representação dos termos no DeCS.

Quase metade das palavras-chave coletadas (47,6%) são termos do vocabulário controlado DeCS. Em um estudo com uma amostra maior talvez pudéssemos confirmar se é uma característica da indexação de documentos na área da interface da Saúde e Ciências Sociais em uma base de dados especializada na área da Saúde e analisar melhor as causas para que se mantenha em torno de 50% ou se o grau de concordância de termos aumenta de modo significativo.

Uma análise dos termos do DeCS poderia ser feita para dar uma visão da relação entre a linguagem natural e a terminologia do domínio com a finalidade de aperfeiçoar a linguagem documentária. Borba (2012) destaca que “um sistema de recuperação da informação, no qual seja representada a informação em consonância com a linguagem dos usuários, possibilitará uma recuperação de informação com maior precisão e pertinência” (p. 26).

Este trabalho também possibilitou um olhar aos estudos envolvendo a melhoria da recuperação com buscas híbridas, com palavras livres e controladas. Estudos indicam que o uso em conjunto das duas linguagens na estratégia de busca proporciona melhores resultados na recuperação da informação (LOPES, 2002). Uma avaliação mais específica a ser feita, na direção da temática estudada, é até que ponto o autor de uma publicação com característica mais interdisciplinar como o periódico deste estudo recorre ao DeCS na definição das palavras-chave dos artigos ou se prevalece a consulta aos pares (linguagem natural).

Um estudo que consideramos importante seria na direção de estratégias de análise de assuntos e definição das palavras-chave a serem atribuídas nos artigos pelos autores visando a recuperação da informação em sistemas sem indexação profissional ou visando seu uso no trabalho dos indexadores das bases de dados.

REFERÊNCIAS

- BOCCATO, V. R. C. A linguagem documentária em catálogos on-line para política de indexação. In: GIL-LEIVA, I.; FUJITA, M. S. L. (ed.). **Política de indexação**. São Paulo: Cultura Acadêmica, 2012. p. 139-151.
- BORBA, D. S.; VAN DER LAAN, R. H.; CHINI, B. R. Palavras-chave: convergências e diferenciações entre a linguagem natural e a terminologia. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 17, n. 2, p. 26-36, abr./jun. 2012.
- CUNHA, M. B.; CAVALCANTI, C. R. O. **Dicionário de Biblioteconomia e Arquivologia**. Brasília, DF: Briquet de Lemos Livros, 2008.
- GARCIA, D. C. F.; GATTAZ, C. C.; GATTAZ, N. C. A relevância do título, do resumo e de palavras-chave para a escrita de artigos científicos. **Revista de Administração Contemporânea**, Maringá, v. 23, n. 3, maio/jun. 2019.
- GIL-LEIVA, I. Consistency in the indexing of documents among new indexers. **Anales de Documentación**, v. 5, p. 99-111, 2002.
- GIL-LEIVA, I. Aspectos conceituais da indexação. In: GIL-LEIVA, I.; FUJITA, M. S. L. (ed.). **Política de indexação**. São Paulo: Cultura Acadêmica, 2012. p. 31-105.
- GIL-LEIVA, I.; ALONSO-ARROYO, A. Keywords given by authors of scientific articles in database descriptors. **Journal of the American Society for Information Science and Technology**, v. 58, n. 8, p. 1175-1187, 2007.
- HARTLEY, J.; KOSTOFF, R. N. How useful are “key words” in scientific journals? **Journal of Information Science**, v. 29, n. 5, p. 433–438, 2003.
- LANCASTER, F. W. **Indexação e resumos: teoria e prática**. 2. ed. Brasília, DF: Briquet de Lemos Livros, 2004.
- LEININGER, K. Interindexer consistency In PsycINFO. **Journal of Librarianship and Information Science**, v. 32, n. 1, p. 4-8, 2000.
- LEONARD, L. E. **Inter-indexer consistency studies, 1954-1975: a review of the literature and summary of study results**. Champaign: University of Illinois, 1977. (Occasional Papers, 131).
- LOPES, I. L. Uso das linguagens controlada e natural em bases de dados: revisão da literatura. **Ciência da Informação**, Brasília, DF, v. 31, n. 1, p. 41-52, 2002.
- MAJUMDER, M. S.; MANDL, K. D. Early in the epidemic: impact of preprints on global discourse about COVID-19 transmissibility. **The Lancet Global Health**, v. 8, n. 5, e627-e630, Mar. 2020.

MARTINS, C. L.; RIBEIRO, H.; ALVARENGA, A. T.; CARVALHEIRO, J. R. Saúde e Sociedade: parceria e abertura para novas abordagens. **Ciência & Saúde Coletiva**, Rio de Janeiro, v. 20, n. 7, p. 2069-2080, 2015.

MENEZHINI, R. Visibilidade internacional da produção brasileira em saúde coletiva. **Cadernos de Saúde Pública**, Rio de Janeiro, v. 26, n. 6, p. 1058-1059, 2010.

MIGUÉIS, A.; NEVES, B. Uma abordagem à linguagem de indexação dos artigos científicos depositados no repositório científico da Universidade de Coimbra. **Ponto de Acesso**, Salvador, v. 7, n. 1, p. 116-131, 2013.

PINHEIRO, L. V. R. Medidas de consistência da indexação: interconsistência. **Ciência da Informação**, Brasília, DF, v. 7, n. 2, p. 109-114, 1978.

PINTO, V. B. Indexação documentária: uma forma de representação do conhecimento registrado. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 6, n. 2, p. 223-234, jul./dez. 2001.

SANTOS, C. A. C. M. **Convergência temática entre produção científica e política nacional de pesquisa em Saúde Pública**: estudo com base em análise de artigos de periódicos indexados. 2010. 165 f. Tese (Doutorado em Ciência da Informação) – Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2010.

SANTOS, C. A. C. M. **Consistência entre as palavras-chave dos artigos da revista Saúde e Sociedade e a indexação da base de dados LILACS**. Projeto de Iniciação Científica, Departamento de Informação e Cultura, Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2017.

SANTOS, F. C. L.; MOLLICA, M. C. M.; GUEDES, V. L. S. Coerência na representação temática de artigos científicos na área de saúde pública. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 24, n. 3, p. 214-232, jul./set. 2019.

STREHL, L. Avaliação da consistência da indexação realizada em uma biblioteca universitária de Artes. **Ciência da Informação**, Brasília, DF, v. 27, n. 3, p. 329-335, set./dez. 1998.

APÊNDICE – *Corpus do estudo*

Item#	Artigo #	Vol. N.	Título do Artigo	Palavras-chave do autor	Descritores Lilacs
1	13	V.22 N.1	Dimensões da qualidade de vida de idosos moradores de rua do município de São Paulo	Idoso; Qualidade de Vida; Moradores de Rua	Idoso Habitação para Idosos Instituição de Longa Permanência para Idosos Autonomia Pessoal Política de Saúde Política Social Saúde do Idoso Assistência à Saúde Vulnerabilidade Social Relações Familiares Qualidade de Vida Pesquisa Qualitativa
2	37	V.22 N.2	O que os adolescentes preferem: os alimentos da escola ou os alimentos competitivos?	Adolescente; Preferências Alimentares; Alimentação Escolar	Nutrição do Adolescente Comportamento Alimentar Estudantes Alimentação Escolar Hábitos Alimentares Nutrição em Saúde Pública Programas e Políticas de Nutrição e Alimentação Necessidades Nutricionais -Sujeitos da Pesquisa Estudos Transversais Inquéritos e Questionários Coleta de Dados

3	51	V.22 N.2	Quem participa em quê?: experiências de construção compartilhada no âmbito da cooperação Brasil - Canadá para o aperfeiçoamento da gestão na Atenção Primária à Saúde	Construção Compartilhada; Atenção Primária em Saúde; Cooperação Internacional	Atenção Primária à Saúde Cooperação Internacional Cooperação Técnica Políticas, Planejamento e Administração em Saúde Políticas Públicas Direito à Saúde Sistema Único de Saúde -Formulação de Projetos Avaliação Pesquisa Qualitativa Análise Quantitativa
4	61	V.22 N.3	Inatividade nos deslocamentos para o trabalho e fatores associados em industrializados	Atividade Motora; Deslocamento Residência-Trabalho; Caminhada; Ciclismo; Trabalhadores; Brasil	Atividade Motora Caminhada Ciclismo Exercício Trabalhadores Indústrias Fatores Socioeconômicos Fatores de Risco Estilo de Vida Sedentário -Estudos Transversais Entrevistas como Assunto Modelos Logísticos Brasil

5	114	V.23 N.1	Etilismo na jornada laboral: peculiaridades da vida naval	Alcoolismo; Saúde Ocupacional; Condições de Trabalho	Alcoolismo Consumo de Bebidas Alcoólicas Militares Características Culturais Bebidas Alcoólicas Punição Ambiente de Trabalho Condições de Trabalho Transtornos Induzidos por Álcool -Antropologia Cultural Brasil
6	145	V.23 N.2	As experiências de mulheres jovens no processo do aborto clandestino: uma abordagem sociológica	Aborto; Saúde Reprodutiva; Sexualidade; Juventude	Saúde da Mulher Aborto Criminoso Saúde da Família Condições Sociais Relações Familiares Saúde Reprodutiva Sexualidade Vulnerabilidade em Saúde Atenção Primária à Saúde Serviços de Saúde Reprodutiva -Pesquisa Qualitativa Pobreza Desemprego Promoção da Saúde Sistema Único de Saúde Mortalidade Materna

7	168	V.23 N.3	Violência intrafamiliar e as estratégias de atuação da equipe de Saúde da Família	Violência; Violência Doméstica; Saúde da Família	Violência Violência/prevenção & controle Violência Doméstica Saúde da Família Estratégia Saúde da Família Pessoal de Saúde Agentes Comunitários de Saúde Relações Familiares Vulnerabilidade em Saúde -Centros de Saúde Saúde Pública Assistência à Saúde Serviços de Saúde Sistema Único de Saúde
8	170	V.23 N.3	A produção científica sobre acesso no âmbito do Sistema Único de Saúde do Brasil: avanços, limites e desafios	Acesso aos Serviços de Saúde; Sistemas de Saúde; Atenção à Saúde	Acesso aos Serviços de Saúde Gestão em Saúde Sistemas de Saúde Sistema Único de Saúde/organização & administração Determinantes Sociais da Saúde Direito à Saúde Acesso Universal a Serviços de Saúde -Saúde Pública Equidade em Saúde Assistência Integral à Saúde Publicações Periódicas como Assunto Resumos e Indexação como Assunto Publicações Científicas e Técnicas

9	181	V.23 N.3	A gente é um passador de informação: práticas educativas de agentes de combate a endemias no serviço de controle de zoonoses em Belo Horizonte, MG	Educação em Saúde; Zoonoses; Agentes Comunitários de Saúde	Zoonoses/prevenção & controle Doenças Transmissíveis Doenças Endêmicas Políticas Públicas de Saúde Educação em Saúde Agentes Comunitários de Saúde Capacitação de Recursos Humanos em Saúde Desenvolvimento Ecológico Impactos na Saúde Controle de Vetores -Fatores Socioeconômicos Fatores Culturais Pesquisa Qualitativa Estudos de Casos Processo Saúde-Doença Urbanização
10	184	V.23 N.3	Leishmaniose tegumentar americana entre os indígenas Xakriabá: imagens, ideias, concepções e estratégias de prevenção e controle	Leishmaniose Tegumentar Americana; Grupos Étnicos; Pesquisa Participativa Baseada na Comunidade	Leishmaniose Cutânea Leishmaniose Cutânea/prevenção & controle População Indígena Educação em Saúde Pessoal de Saúde Perfis Sanitários Consentimento Livre e Esclarecido Participação Comunitária -Pesquisa Sistema Único de Saúde Fatores Socioeconômicos Fatores Culturais Fatores de Risco

11	206	V.23 N.4	As relações de trabalho dos fisioterapeutas na cidade de Salvador, Bahia	Fisioterapia; Reestruturação Produtiva; Relações de Trabalho; Precarização	Saúde do Trabalhador Fisioterapeutas Condições de Trabalho Ambiente de Trabalho Riscos Ocupacionais Relações Trabalhistas Modalidades de Fisioterapia Jornada de Trabalho Esgotamento Profissional -Fatores Socioeconômicos Mercado de Trabalho Pesquisa Qualitativa Prática Profissional Políticas Públicas
12	209	V.23 N.4	Comunicação nas práticas em saúde: revisão integrativa da literatura	Comunicação; Comunicação em Saúde; Educação em Saúde	Conhecimentos, Atitudes e Prática em Saúde Prática de Saúde Pública Comunicação em Saúde Sistema Único de Saúde/legislação & jurisprudência Educação em Saúde Pessoal de Saúde Participação Comunitária -Prática Profissional Relações Interpessoais Serviços de Saúde Atenção à Saúde

13	215	V.23 N.4	Segurança do paciente no uso de medicamentos após a alta hospitalar: estudo exploratório ¹	Medicamentos; Erros de Medicação; Eventos Adversos; Alta Hospitalar; Segurança do Paciente; Saúde Pública	Preparações Farmacêuticas Segurança do Paciente Erros de Medicação Alta do Paciente Medicamentos sob Prescrição Quimioterapia Continuidade da Assistência ao Paciente Assistência Domiciliar Reconciliação de Medicamentos -Adesão à Medicação Hospitais Universitários Pessoal de Saúde Serviços de Saúde Atenção Primária à Saúde
14	220	V.24 N.1	Eficiência dos gastos públicos em saúde: desafio para municípios de Santa Catarina, Brasil	Eficiência; Saúde; Investimentos em Saúde	Gestão em Saúde Assistência Integral à Saúde Equidade em Saúde Gastos em Saúde Controle de Custos Sistema Único de Saúde Eficiência Investimentos em Saúde -Sistemas de Saúde Serviços de Saúde Participação Social Determinantes Sociais da Saúde Descentralização

15	262	V.24 N.2	Avaliação de resultados da atenção aos agravos cardiovasculares como traçador do princípio de integralidade	Avaliação em Saúde; Gestão em Saúde; Assistência Integral à Saúde	Avaliação em Saúde Assistência Integral à Saúde Serviços de Saúde Doenças Cardiovasculares Sistemas de Saúde Regionalização Políticas Públicas Descentralização -Sistema Único de Saúde Acesso Universal a Serviços de Saúde Integralidade em Saúde Equidade em Saúde
16	269	V.24 N.2	Olhar da gestão sobre a implantação da ficha de notificação da violência doméstica, sexual e/outras violências em uma metrópole do Brasil	Violência; Notificação; Vigilância em Saúde; Crianças e Adolescentes	Violência Doméstica Violência Sexual Violência Notificação de Abuso Políticas Públicas Capacitação de Recursos Humanos em Saúde Vigilância em Saúde Pública Criança Adolescente -Pesquisa Qualitativa Perfil de Saúde Defesa da Criança e do Adolescente Direitos Humanos Sistemas de Informação

17	295	V.24 N.3	Discursos de médicos de família brasileiros e italianos sobre autonomia na perspectiva bioética	Autonomia; Atenção Primária à Saúde; Brasil; Itália; Prevenção ao Risco; Bioética	Medicina de Família e Comunidade Atenção Primária à Saúde Autonomia Profissional Política de Saúde Bioética Tomada de Decisões Integralidade em Saúde -Pesquisa Qualitativa Estratégias Risco Promoção da Saúde Direito à Saúde Qualidade de Vida Brasil Itália
18	301	V.24 N.4	Justiça social e equidade em saúde: uma abordagem centrada nos funcionamentos	Justiça Social; Políticas Públicas; Equidade em Saúde; Abordagem das Capacitações; Perspectiva dos Funcionamentos	Justiça Social Equidade em Saúde Políticas Públicas Capacitação de Recursos Humanos em Saúde Recursos em Saúde Alocação de Custos Problemas Sociais Processo Saúde-Doença Condições Sociais -Equidade na Alocação de Recursos Necessidades e Demandas de Serviços de Saúde Eficiência Tomada de Decisões Qualidade de Vida

19	303	V.24 N.4	As múltiplas fronteiras presentes no atendimento à saúde do estrangeiro em Corumbá, Brasil	Fronteiras; Identidades; Atendimento à Saúde	Serviços de Saúde Áreas de Fronteira Saúde na Fronteira Centros de Saúde Sistemas de Saúde Pessoal de Saúde Relações Interpessoais Políticas Públicas de Saúde -Pesquisa Qualitativa Assistência à Saúde Sistema Único de Saúde Antropologia Cultural Sociologia Preconceito
20	314	V.24 N.4	Mulheres, saúde e uso de crack: a reprodução do novo racismo na/pela mídia televisiva	Psicologia Social; Crack; Drogas; Gênero; Mídia Televisiva; Racismo	Saúde da Mulher Drogas Ilícitas Cocaína Crack Gravidez Meios de Comunicação de Massa Direitos Sexuais e Reprodutivos Racismo Internação Compulsória de Doente Mental Direitos da Mulher Autonomia Pessoal -Atenção à Saúde Psicologia Social Discriminação Social Iniquidade Social Problemas Sociais Pessoal de Saúde

Fonte: elaboração própria.