



## PRODUÇÃO CIENTÍFICA EM SAÚDE NOS PPGS DE EXCELÊNCIA DO SUDESTE: CLASSIFICAÇÃO HIERÁRQUICA E LEXICOMETRIA

DALMO MACHADO; MARIA LUIZA GUERRA DE TOLEDO, ALINNE DE CARVALHO VEIGA

### RESUMO

Este estudo teve como objetivo identificar os principais campos temáticos em resumos de produções científicas por meio da análise lexical utilizando a Classificação Hierárquica Descendente (CHD), segundo o método de Reinert, com o apoio do software Iramuteq. Método: estudo cientométrico. A fonte primária de informação foi a Plataforma Sucupira (<https://sucupira-v2.capes.gov.br>) contida no próprio sítio da CAPES, a fim de identificar os PPGs que contemplavam o escopo da pesquisa. A partir de então, foi realizada a varredura em cada um dos 30 cursos selecionados, a partir dos seus próprios sítios. A coleta de análise das informações deu-se entre os meses de janeiro e junho de 2025. Foram analisados 35 resumos, totalizando 470 segmentos de texto, dos quais 345 (73,4%) foram classificados em quatro classes distintas, conforme sua coesão lexical. A Classe 1 (20,0%) concentrou-se em ações educativas voltadas para adolescentes e familiares, com forte uso de tecnologias educativas como vídeos. A Classe 2 (30,1%) abordou tratamentos estatísticos e evidências de eficácia, destacando termos como “diferença”, “significância” e “modelo”. A Classe 3 (25,8%) reuniu vocabulário relacionado a ensaios clínicos randomizados, com foco em doenças crônicas e controle metodológico. Já a Classe 4 (24,1%) refletiu intervenções fisioterapêuticas e avaliações funcionais, com destaque para abordagens como exercícios físicos, tratamentos não farmacológicos e reabilitação. Os resultados evidenciam a diversidade de abordagens metodológicas e temáticas nos textos analisados, demonstrando uma tendência à integração entre práticas clínicas, estratégias educativas e rigor metodológico nas produções científicas. A CHD demonstrou ser uma ferramenta eficaz para a categorização semântica de conteúdos complexos em saúde e áreas afins.

**Palavras-chave:** enfermagem; cientometria; programas de pós-graduação.

### 1 INTRODUÇÃO

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) é o órgão federal responsável pela avaliação e recomendação dos programas de pós-graduação (PPG) stricto sensu no Brasil. Ao término de quatro anos, a CAPES emite relatório qualitativo e quantitativo dos PPG, com os respectivos graus de recomendação, que vão de 3 a 7, para todas as áreas de conhecimento. Os cursos ranqueados com nota 7 são aqueles considerados de excelência internacional em todos os aspectos avaliados: produção intelectual, internacionalização, liderança e reconhecimento, formação de recursos humanos, impacto social e infraestrutura e corpo docente qualificado.

Conhecer o perfil das teses e dissertações defendidas nos cursos de excelência pode contribuir para a elevação dos cursos de níveis inferiores, elevando a qualidade da publicação científica brasileira.

Dentre os desenhos metodológicos utilizados em pesquisa em saúde, somente os ensaios clínicos se propõem ao estabelecimento de relações de causa e efeito. Portanto, trata-se dos desenhos preferenciais para análise das intervenções em saúde, essenciais em fases

experimentais e pré-clínicas, mandatórias antes da incorporação de tecnologias pelo Sistema Único de Saúde (SUS).

Segundo a última avaliação quadrienal existem 261 cursos com nota 7 no Brasil, dos quais 182 estão localizados na região Sudeste e, destes, 30 estão incluídos na grande área da saúde, dos quais, os de enfermagem.

Assim, o objetivo da presente é realizar a análise dos resumos de dissertações e teses de ensaios clínicos defendidas no ano de 2024 derivadas de PPG com nota 7, grande área da saúde, região Sudeste.

## 2 MATERIAL E MÉTODOS

Estudo cientométrico que se utilizava da análise lexical utilizando a Classificação Hierárquica Descendente (CHD), pelo método de Reinert, por meio do software Iramuteq.

A fonte primária de informação foi a Plataforma Sucupira ([https://sucupira-v2.capes.gov.br](https://sucupira.v2.capes.gov.br)) contida no próprio sítio da CAPES, a fim de identificar os PPGs que contemplavam o escopo da pesquisa. A partir de então, foi realizada a varredura em cada um dos 30 cursos selecionados, a partir dos seus próprios sítios.

A coleta de análise das informações deu-se entre os meses de janeiro e junho de 2025.

## 3 RESULTADOS E DISCUSSÃO

O delineamento do espaço amostral é visualizado na Figura 1:

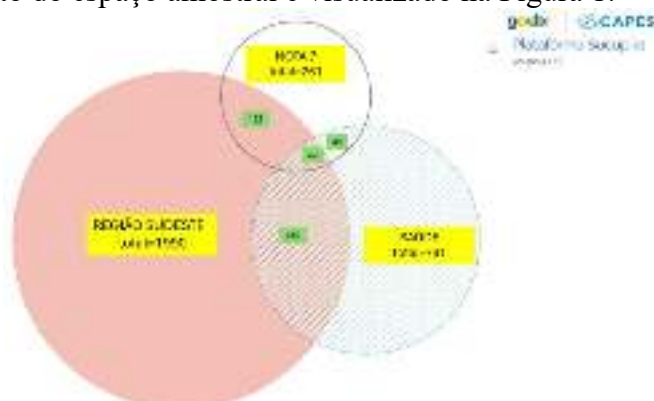


Figura 1: delimitação de espaço amostral

Os 30 cursos geraram 563 dissertações/teses, das quais 35 constituíram-se em ensaios clínicos (Tabela 1).

Tabela 1: Ensaios clínicos encontrados nas dissertações/teses defendidas no ano de 2024

| PROGRAMA           | n (%)     | PROGRAMA               | n (%)     | PROGRAMA             | n (%)     |
|--------------------|-----------|------------------------|-----------|----------------------|-----------|
| 1_UERJ_scoletiva   | 0 (0%)    | 11_UNICAMPP_cilodonto  | 0 (0%)    | 21_USP_nutrispublica | 0 (0%)    |
| 2_UFMG_odonto      | 1 (2,9%)  | 12_UNICAMPP_estomato   | 0 (0%)    | 22_USP_scol          | 2 (5,7%)  |
| 3_UFMG_spublica    | 0 (0%)    | 13_UNICAMPP_odonto     | 0 (0%)    | 23_USP_adffis        | 4 (11,4%) |
| 4_UFRJ_clmed       | 1 (2,9%)  | 14_UNIFESP_psicobio    | 1 (2,9%)  | 24_USP_farm          | 0 (0%)    |
| 5_UFSCAR_fisio     | 4 (11,4%) | 15_UNIFESP_psiqpsi     | 1 (2,9%)  | 25_USPRP_clmed       | 0 (0%)    |
| 6_UFV_nutri        | 2 (5,7%)  | 16_UNIFESP_ortomocovis | 3 (8,6%)  | 26_USPRP_spublica    | 1 (2,9%)  |
| 7_UNESP_biocbiotec | 0 (0%)    | 17_USP_proesa          | 3 (8,6%)  | 27_USPRP_neuro       | 1 (2,9%)  |
| 8_UNESPA_odonto    | 0 (0%)    | 18_USP_psiq            | 0 (0%)    | 28_USPRP_gnecocoba   | 0 (0%)    |
| 9_UNICAMP_tocagin  | 1 (2,9%)  | 19_USP_cardio          | 9 (25,8%) | 29_USPRP_enfund      | 0 (0%)    |
| 10_UNICAMP_fimed   | 0 (0%)    | 20_USP_farm            | 0 (0%)    | 30_USPRP_amcc        | 1 (2,9%)  |

Utilizada a Nuvem de Palavras, a fim de agrupar as palavras e organizá-las graficamente em função da sua relevância, sendo as maiores aquelas que possuíam maior frequência, considerando palavras com frequência igual ou superior a 7.

Foi analisada a nuvem de palavras obtida por meio dos resumos de dissertações e teses, em que verifica-se que as palavras mais prevalentes : “paciente” (f = 92), “avaliar” (f = 71), “tratamento” (f = 47), “redução” (f = 47), “apresentar” (f = 46), “realizar” (f = 42), “clínico” (f = 37), “dor” (f = 34), “desfecho” (f = 34), “saúde” (f = 31), “placebo” (f = 28) e “mulher” (f = 27) (Figura 2).



Figura 2 – Nuvem de palavras de resumos de ensaios clínicos, 2024.

Realizada a Análise de Similitude, que possibilita identificar as ocorrências entre as palavras e seu resultado, traz indicações da conexidade entre as palavras, facilitando a compreensão do corpus textual analisado.

Por meio da análise baseada na teoria dos grafos, é possível identificar as ocorrências textuais entre as palavras e as indicações da conexidade entre as palavras, auxiliando na identificação da estrutura do conteúdo de um corpus textual. Observa-se que há quatro palavras que se destacam nos resumos: “paciente”; “avaliar”; “apresentar. Nota-se que as ações (verbos) derivam do componente principal (paciente). Delas se ramificam outras palavras que sugerem significados mais detalhados.

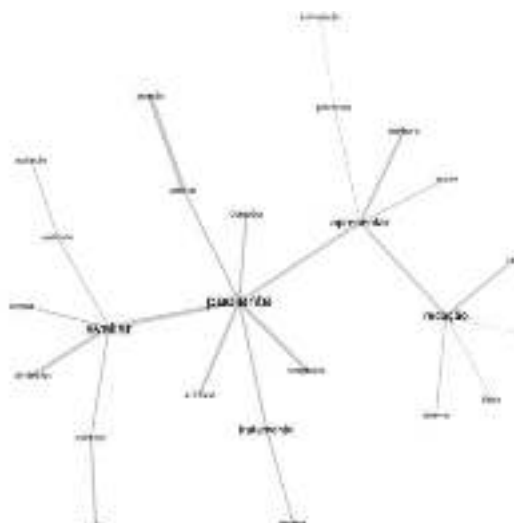


Figura 3 – Análise de similitude de resumos de ensaios clínicos, 2024.

Com o objetivo de identificar os principais campos temáticos presentes nos resumos analisados, foi realizada uma análise lexical utilizando a Classificação Hierárquica Descendente (CHD), segundo o método de Reinert, por meio do software Iramuteq. Esta abordagem permite agrupar segmentos de texto com base na similaridade do vocabulário, produzindo classes de conteúdo semanticamente coesas e distintas entre si.

O corpus analisado foi composto por 35 resumos, resultando em 470 segmentos de texto (STs). Destes, 345 segmentos (73,4%) foram retidos para a análise, respeitando os critérios estatísticos mínimos de frequência. O processamento identificou 3.045 formas (palavras únicas) e um total de 16.577 ocorrências, com uma média de 35,27 palavras por segmento. A CHD classificou os segmentos em quatro classes lexicais distintas, conforme representado no dendrograma da Figura 4.

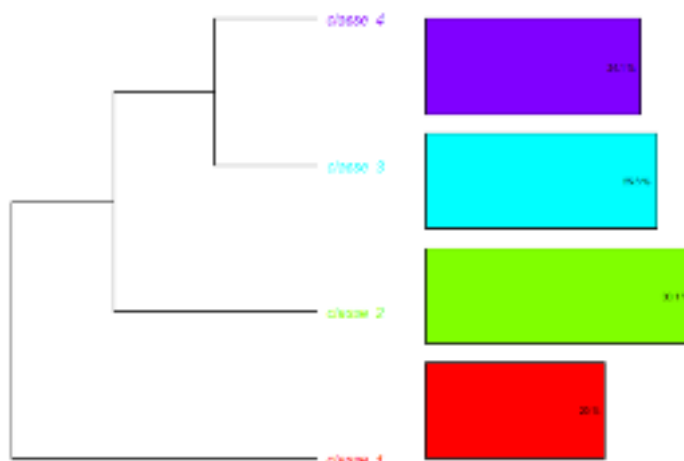


Figura 4 - Dendrograma da Classificação Hierárquica Descendente de resumos de ensaios clínicos, 2024.

A distribuição dos segmentos entre as classes foi a seguinte: Classe 1: 20,0% (n = 69), Classe 2: 30,1% (n = 104), Classe 3: 25,8% (n = 89) e Classe 4: 24,1% (n = 83).

Classe 1 – Educação em saúde com foco em adolescentes e família

Esta classe reúne termos como “saúde”, “educativo”, “adolescente”, “intervenção”, “vídeo”, “mãe”, “pai”, “domiciliar” e “participação”, indicando a predominância de ações

educativas voltadas ao público adolescente e ao núcleo familiar. As estratégias descritas nos segmentos estão fortemente ancoradas em metodologias audiovisuais, como o uso de vídeos informativos, com base em modelos estruturados como o STEP ou sistemas como o EMIRA. A ênfase na educação em saúde e na promoção do cuidado compartilhado sugere iniciativas de prevenção primária, com atuação em ambientes escolares e domiciliares.

#### Classe 2 – Tratamento estatístico e avaliação de eficácia

Esta classe destaca termos técnicos como “diferença”, “estatística”, “significância”, “nível”, “modelo”, “variável”, “plasmático”, “probiótico”, “t1”, “t2”, “anovar” e “porcento”, os quais evidenciam um vocabulário centrado em análises estatísticas, avaliação de eficácia e comparação de resultados entre grupos. Os segmentos indicam que os estudos utilizaram ferramentas estatísticas rigorosas para sustentar a validade dos achados, frequentemente aplicando testes de hipóteses e medidas de efeito. Há uma clara preocupação metodológica com a robustez das evidências e com a interpretação de dados quantitativos.

#### Classe 3 – Ensaios clínicos randomizados e controle de variáveis

Agrupam-se aqui termos como “controlar”, “duplo”, “randomizado”, “cegos”, “dm2”, “peso”, “arterial”, “crm”, “empagliflozina”, “dac”, e “drc”, demonstrando que os textos desta classe referem-se majoritariamente a ensaios clínicos randomizados controlados (RCTs). Há foco na alocação aleatória, controle de viés e avaliação de fármacos, especialmente no contexto de doenças crônicas não transmissíveis como diabetes tipo 2 e doença arterial coronariana. Esses estudos demonstram alto rigor metodológico, com atenção à padronização de procedimentos e monitoramento de variáveis fisiológicas.

#### Classe 4 – Avaliação de intervenções fisioterapêuticas e funcionais

Com palavras associadas como “vascular”, “secundário”, “exercício”, “sessão”, “custo”, “laser”, “compressão”, “global”, e “tratamento”, esta classe reflete estudos voltados à avaliação funcional, reabilitação e respostas a terapias não farmacológicas, como intervenções fisioterapêuticas e exercícios físicos. Também aparecem temas relacionados à função vascular, incontinência urinária de esforço (IUE) e melhoria da qualidade de vida, muitas vezes em associação a custos e desfechos secundários. A presença de vocabulário relacionado a mecanismos fisiológicos sugere avaliações estruturadas e técnicas de mensuração bem definidas.

Tabela 2 - Síntese dos eixos temáticos identificados

| Classe   | Eixo Temático                                        | % Segmentos |
|----------|------------------------------------------------------|-------------|
| Classe 1 | Educação em saúde para adolescentes e famílias       | 20,0%       |
| Classe 2 | Análise estatística e eficácia de intervenções       | 30,1%       |
| Classe 3 | Ensaios clínicos randomizados em doenças crônicas    | 25,8%       |
| Classe 4 | Intervenções fisioterapêuticas e avaliação funcional | 24,1%       |

Os resultados da CHD evidenciam a diversidade temática dos resumos analisados, revelando quatro grandes blocos de interesse: promoção da saúde com foco educacional e familiar (Classe 1), tratamento estatístico e validação de resultados (Classe 2), estudos clínicos rigorosos (Classe 3) e intervenções terapêuticas voltadas à função física e fisiológica (Classe 4).

A organização das classes revela também uma hierarquia de especialização lexical, com a Classe 1 posicionando-se de forma mais autônoma no dendrograma, enquanto as Classes 2, 3 e 4 compartilham um léxico mais técnico e interligado. Tais achados são coerentes com a

crescente multidisciplinaridade da produção científica em saúde, refletindo desde abordagens de base comunitária até metodologias clínicas controladas.

#### **4 CONCLUSÃO**

A análise por CHD revelou quatro núcleos temáticos distintos nos resumos analisados, refletindo a pluralidade metodológica e de enfoque nas produções científicas. A ferramenta demonstrou potencial analítico para estudos de conteúdo, especialmente em contextos interdisciplinares, ao identificar padrões lexicais representativos de diferentes abordagens investigativas em saúde.

#### **REFERÊNCIAS**

BARDIN, L. *Análise de conteúdo*. 1. ed. São Paulo: Edições 70, 2011.

CAMARGO, B. V.; JUSTO, A. M. IRaMuTeQ: Um software gratuito para análise de dados textuais. *Temas em Psicologia*, v. 21, n. 2, p. 513–518, 2013.

LIMA, T. C. et al. O uso da análise lexical na investigação de percepções sobre saúde e cuidado: contribuições metodológicas. *Revista de Pesquisa Cuidado é Fundamental Online*, v. 10, n. 2, p. 482–488, 2018.

MORO, A. R. P. et al. Aplicações do software IRaMuTeQ nas pesquisas qualitativas em saúde: uma revisão integrativa. *Revista Gaúcha de Enfermagem*, v. 42, e20200334, 2021.

REINERT, M. Une méthode de classification descendante hiérarchique: application à l'analyse lexicale par contexte. *Les cahiers de l'analyse des données*, v. 8, n. 2, p. 187–198, 1983.