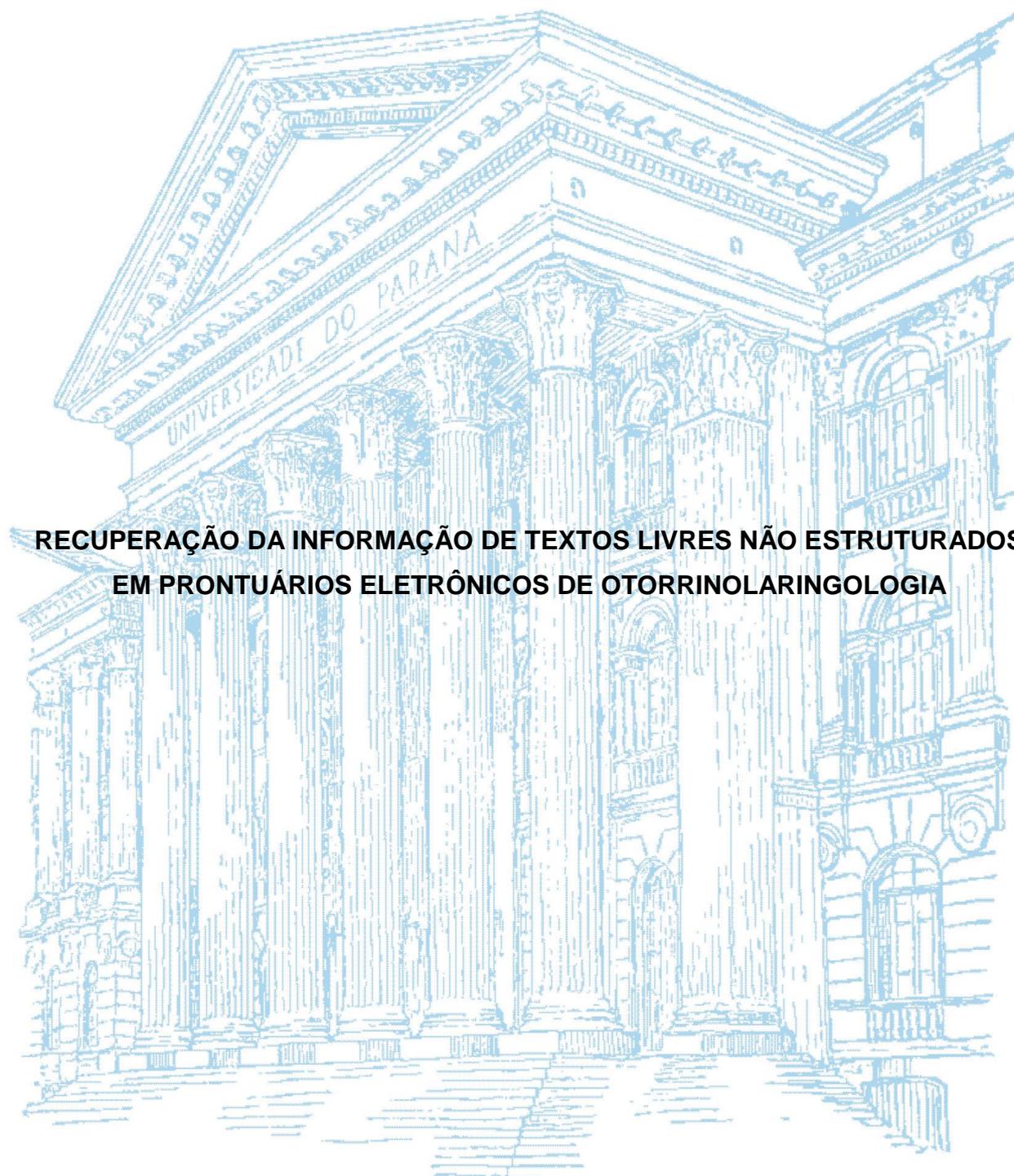


UNIVERSIDADE FEDERAL DO PARANÁ

CLAUDIO JOSE BELTRÃO



**RECUPERAÇÃO DA INFORMAÇÃO DE TEXTOS LIVRES NÃO ESTRUTURADOS
EM PRONTUÁRIOS ELETRÔNICOS DE OTORRINOLARINGOLOGIA**

CURITIBA

2017

CLAUDIO JOSE BELTRÃO

RECUPERAÇÃO DA INFORMAÇÃO DE TEXTOS LIVRES NÃO ESTRUTURADOS
EM PRONTUÁRIOS ELETRÔNICOS DE OTORRINOLARINGOLOGIA

Tese apresentada ao Programa de Pós-Graduação em
Clínica Cirúrgica do Setor de Ciências da Saúde da
Universidade Federal do Paraná, como requisito
parcial à obtenção do grau de Doutor.

Orientador: Prof. Dr. Rogério Hamerschmidt

Coordenador: Prof. Dr. Jorge Eduardo Fouto Matias

CURITIBA

2017

B453 Beltrão, Claudio Jose

Recuperação da informação de textos livres não estruturados em prontuários eletrônicos de otorrinolaringologia / Claudio Jose Beltrão. -- Curitiba, 2017.

99f. : il.

Orientador: Prof. Dr. Rogério Hamerschmidt

Tese (Doutorado) – Programa de Pós-Graduação em Clínica Cirúrgica. Setor de Ciências da Saúde. Universidade Federal do Paraná.

1. Registros eletrônicos de saúde. 2. Protocolos. 3. Otorrinolaringologia. I. Beltrão, Cláudio Jose, II. Hamerschmidt, Rogério, III. Programa de Pós-Graduação em Clínica Cirúrgica. Setor de Ciências da Saúde. Universidade Federal do Paraná. IV. Título.

NLMC: WV27



MINISTÉRIO DA EDUCAÇÃO
SETOR CIÊNCIAS DA SAÚDE
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO MEDICINA (CLÍNICA
CIRÚRGICA)

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em MEDICINA (CLÍNICA CIRÚRGICA) da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **CLAUDIO JOSE BELTRAO** intitulada: "**RECUPERAÇÃO DA INFORMAÇÃO DE TEXTOS LIVRES NÃO ESTRUTURADOS EM PRONTUÁRIOS ELETRÔNICOS DE OTORRINOLARINGOLOGIA**", após terem inquirido o aluno e realizado a avaliação do trabalho, são de parecer pela sua aprovada no rito de defesa.

A outorga do título de doutor está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 08 de Dezembro de 2017.


OSVALDO MALAFAIA
Presidente da Banca Examinadora


ELEN CAROLINO DAVID JOAO DE MASI
Avaliador Externo


JOSÉ SIMÃO DE PAULA PINTO
Avaliador Externo


GUILHERME SIMAS DO AMARAL CATANI
Avaliador Externo


JOÃO LUIZ GARCIA DE FARIA
Avaliador Externo

À minha luz Christianne e às
minhas meninas, Julianna e Marianna,
sem elas nada faria sentido.

AGRADECIMENTOS

Ao **Prof. Dr. Osvaldo Malafaia**, idealizador do Sistema SINPE©, pela visão, apoio e dedicação à linha de pesquisa de Informática em Saúde. Sem seu apoio nada disso teria acontecido.

Ao **Prof. Dr. Rogério Hamerschmidt** pela disposição e orientação na confecção deste trabalho.

Ao **Hospital IPO**, pelo acesso à informação e pela possibilidade da realização deste trabalho.

Ao **Prof. Dr. João Luiz Garcia de Faria**, pela parceria e incentivo ao longo de tantos anos.

Ao **Prof. Dr. Evaldo Dacheux de Macedo Filho** por todo o apoio frente à pesquisa no Hospital IPO.

Ao **Prof. Dr. Jorge Eduardo Fouto Matias**, Coordenador do Programa de Pós-Graduação em Clínica Cirúrgica da Universidade Federal do Paraná, pela oportunidade de realização desta pesquisa.

Ao **Prof. Dr. José Simão de Paula Pinto**, pela disponibilidade e ajuda na utilização do sistema SINPE e demais auxílios.

Ao **Prof. Dr. Carlos Henrique Kuretzki**, pelo auxílio na utilização dos protocolos e sistemas envolvidos.

À **Prof. Dra. Márcia Olandoski**, pela colaboração na análise dos dados e apoio.

À **Universidade Federal do Paraná - UFPR** pela oportunidade estudo, pesquisa e conhecimento científico.

Aos **professores do Programa de Pós-Graduação em Clínica Cirúrgica** da UFPR pelos ensinamentos teóricos.

A todos aqueles que direta ou indiretamente estiveram ligados à coleta, análise e resultados.

Conte-me e eu esquecerei.

Ensine-me e eu lembrarei.

Envolva-me e eu aprenderei.

(Benjamin Franklin)

RESUMO

Introdução: O uso de prontuários eletrônicos na prática médica é uma realidade e os textos livres não estruturados, contêm uma grande quantidade de dados que podem ser utilizados para a formação de conhecimento. Utilizar técnicas de recuperação da informação, contribui para a descoberta dos termos mais utilizados e para compará-los a protocolos existentes. O preenchimento automatizado de coleta pode auxiliar o aumento das amostras em pesquisas.

Objetivos: Extrair palavras, de textos livres não estruturados, em prontuário eletrônico e comparar sua frequência com um protocolo já existente de termos, das doenças de maior prevalência, em uma base de dados. Criar uma interface de ligação para o preenchimento automático dos sintomas e doenças, entre o Sistema CLINIC® e o Sistema SINPE©.

Material e Métodos: O estudo foi realizado com base em um *big data*, de 375.095 prontuários eletrônicos. Destes, foram selecionados os que possuíam CID e campo histórico preenchido, com no mínimo cinco caracteres e excluídos os pós-operatórios e reconsultas inferiores a 30 dias. Do resultado, foram analisados os CID's mais prevalentes e analisados os textos livres não estruturados, com a ferramenta *word cloud* e, após, comparados com termos do sistema de protocolos. Foi elaborada uma interface, entre o sistema de prontuários e o Sistema SINPE©, visando uma coleta de dados automatizada.

Resultados: Foram analisados 125.973 prontuários, contendo textos livres não estruturados. A amostra apresentou 57,24% de pacientes do sexo feminino e 42,76% do sexo masculino, com idade média de 38,53 anos. Os 19 CID's de maior prevalência representaram 70,11% dos casos. Foram desenvolvidas *word cloud's* para os textos livres, nas doenças de maior prevalência. Compararam-se as palavras inseridas no prontuário eletrônico, com os termos do protocolo de otorrinolaringologia, do Sistema SINPE© sendo identificadas as palavras mais frequentes, nas doenças de maior prevalência e construídos intervalos de confiança (IC=95%). A interface desenvolvida, mostrou ser capaz de identificar, nos textos livres, os termos do protocolo e automatizar a coleta de dados.

Conclusões: Foram extraídas, de textos livres não estruturados, as palavras de maior frequência, e comparadas com um protocolo de termos já existente, nas doenças de maior prevalência. Foi desenvolvida uma interface entre os sistemas CLINIC® e SINPE©, capaz de preencher de forma automática os itens do protocolo mestre, tanto em sintomas como em diagnósticos.

Palavras-chave: Registros Eletrônicos de Saúde; Protocolos; Otorrinolaringologia.

ABSTRACT

Introduction: The use of electronic medical records in medical practice is a reality and free unstructured text contains a large amount of information, which may be used to establish knowledge. Using information retrieval techniques contributes for the discovery of the most used terms and the comparison of such terms to existing protocols. Automated data collection may help in increasing samples in research. **Objectives:** Extract words from free unstructured text out of electronic medical records and compare their frequency to an existing protocol of terms for the most prevalent diseases from a database. Create a connecting interface for the automatic filling in of symptoms and diseases between the CLINIC® and the SINPE© systems. **Material and Methods:** The study was performed based on a *big data* made up of 375,095 electronic medical records. Out of these, those which contained the ICD and medical history fields filled in and which had at least five characters were selected, and post-ops and return visits within less than thirty days were excluded. From the results, the most prevalent ICDs and free unstructured text were analyzed with the word cloud tool, and later compared to the terms in the protocol system. An interface was created between the SINPE© system and the medical records system, aiming at automated data collection. **Results:** 125,973 medical records were analyzed which had free unstructured text. The sample was made up of 57.24% of female patients and 42.76% male patients, with an average age of 38.53 years old. The 19 most prevalent ICDs represented 70.11% of cases. Word clouds were developed for the free text in the most prevalent diseases. The words inserted in the electronic medical records were compared to the terms in the otolaryngology protocol in the SINPE© system. The most frequent words were identified in the most prevalent diseases and confidence intervals were built (CI 95%). The interface developed was able to identify the terms of the protocol in free text and automate the data collection. **Conclusions:** The most frequent words were extracted from free unstructured text and compared to an existing protocol of terms for the most prevalent diseases. An interface between the CLINIC® and the SINPE© systems was developed which is able to automatically fill in the items of the master protocol, relative to both symptoms and diagnosis.

Keywords: Electronic Health Records; Protocols; Otorhinolaryngology.

LISTA DE FIGURAS

FIGURA 1 - TELAS DE RESULTADO DA CONSULTA DE TERMOS NA PLATAFORMA DECS.	20
FIGURA 2 - MODELO CONCEITUAL DO SISTEMA SINPE©.....	23
FIGURA 3 - TELA DO APLICATIVO EPI INFO FOR WINDOWS VERSÃO 7.2.1.0..	32
FIGURA 4 - DIAGRAMA PARA INTERFACE ENTRE O SISTEMA CLINIC® E O SISTEMA SINPE©.	34
FIGURA 5 – DEMONSTRAÇÃO DA REDUÇÃO DA BASE ORIGINAL PARA BASE FINAL	35
FIGURA 6 - WORD CLOUD PARA RINITE ALÉRGICA NÃO ESPECIFICADA (CID J30.4) N=17.179.....	37
FIGURA 7 – <i>WORD CLOUD</i> PARA FARINGITE AGUDA (CID J02) N=13.527	37
FIGURA 8 – <i>WORD CLOUD</i> PARA CERUME IMPACTADO (CID H61.2) N=11.755	38
FIGURA 9– <i>WORD CLOUD</i> PARA INFECÇÃO AGUDA DAS VIAS AÉREAS SUP. NÃO ESPECIFICADA (CID J06.9) N=11.233	38
FIGURA 10 - <i>WORD CLOUD</i> PARA AMIGDALITE AGUDA NÃO ESPECIFICADA (CID J03.9) N=11.062	39
FIGURA 11 – <i>WORD CLOUD</i> PARA SINUSITE MAXILAR AGUDA (CID J01.0) N=7.616.....	39
FIGURA 12 – <i>WORD CLOUD</i> PARA NASOFARINGITE AGUDA (CID J00) N=7.432	40
FIGURA 13 – <i>WORD CLOUD</i> PARA TOSSE (CID R05) N=6.615	40
FIGURA 14 – <i>WORD CLOUD</i> PARA DESVIO DE SEPTO NASAL (CID J34.2) N=4.915.....	41
FIGURA 15 – <i>WORD CLOUD</i> PARA EXAME MÉDICO GERAL (CID Z00.0) N=4.502	41
FIGURA 16 – <i>WORD CLOUD</i> PARA LARINGITE AGUDA (CID J04.0) N=4.321.....	42
FIGURA 17 – <i>WORD CLOUD</i> PARA OTITE EXTERNA NÃO ESPECIFICADA (CID H60.9) N=4.292	42
FIGURA 18 – <i>WORD CLOUD</i> PARA TONTURA (CID R42) N=3.880	43

FIGURA 19 – <i>WORD CLOUD</i> PARA OUTRAS OTITES MÉDIAS AGUDAS NÃO-SUPURATIVAS (CID H65.1) N=3.595	43
FIGURA 20 - <i>WORD CLOUD</i> PARA APNEIA DO SONO (CID G47.3) N=3.181	44
FIGURA 21 – <i>WORD CLOUD</i> PARA PERDA DE AUDIÇÃO (CID H91) N=3.162	44
FIGURA 22 – <i>WORD CLOUD</i> PARA DOR DE GARGANTA (CID R07.0) N=2.669..	45
FIGURA 23 – <i>WORD CLOUD</i> PARA OUTRAS RINITES ALÉRGICAS (CID J30) N=2.644.....	45
FIGURA 24 – <i>WORD CLOUD</i> PARA FARINGITE AGUDA DEVIDA A OUTROS MICROORGANISMOS ESPECIFICADOS (CID J02.8) N=2.393.....	46
FIGURA 25 – TELA DO SINPE© COM O PROTOCOLO MESTRE DE OTORRINOLARINGOLOGIA	47
FIGURA 26 – SIMULAÇÃO COM TERMOS E DIAGNÓSTICO.....	56
FIGURA 27 - SISTEMA SINPE© COM SINTOMAS PREENCHIDOS	57
FIGURA 28 - SISTEMA SINPE© COM DIAGNÓSTICO PREENCHIDO E COLETA FINALIZADA.....	57

LISTA DE GRÁFICOS

GRÁFICO 1 – DISTRIBUIÇÃO DOS TERMOS ENCONTRADOS NA ÁREA GARGANTA (N=143.804)	48
GRÁFICO 2 – DISTRIBUIÇÃO DOS TERMOS ENCONTRADOS NA SUBDIVISÃO NARIZ (N=61.704).....	49
GRÁFICO 3 – DISTRIBUIÇÃO DOS TERMOS ENCONTRADOS NA ÁREA OUVIDO (N=35.766)	49
GRÁFICO 4 – DISTRIBUIÇÃO DOS TERMOS ENCONTRADOS EM OUTROS CID´S (N=4.331)	50
GRÁFICO 5 – FREQUÊNCIA E INTERVALO DE CONFIANÇA (95%) DE PALAVRAS NA SUBDIVISÃO GARGANTA (N=143.804).....	51
GRÁFICO 6 - FREQUÊNCIA E INTERVALO DE CONFIANÇA (95%) DE PALAVRAS NA SUBDIVISÃO NARIZ (N=61.704).....	51
GRÁFICO 7 - FREQUÊNCIA E INTERVALO DE CONFIANÇA (95%) DE PALAVRAS NA SUBDIVISÃO OUVIDO (N=35.766)	52
GRÁFICO 8 - FREQUÊNCIA E INTERVALO DE CONFIANÇA (95%) DE PALAVRAS NA SUBDIVISÃO OUTRAS DOENÇAS (N=4.331).....	52
GRÁFICO 9 – DEMONSTRAÇÃO RADIAL DA DISTRIBUIÇÃO DOS TERMOS POR SUBDIVISÃO	55

LISTA DE TABELAS

TABELA 1 - DOENÇAS MAIS PREVALENTES NA AMOSTRA (N=125.973)	36
TABELA 2 - PREVALÊNCIA DE CID'S POR SUBDIVISÃO GARGANTA; NARIZ; OUVIDO E OUTROS (N=125.973).....	48
TABELA 3 – FREQUÊNCIA DE SINTOMAS POR SUBDIVISÃO (N=245.605)	54
TABELA 4 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM RINITE ALÉRGICA (CID J30.4) N=17.179	73
TABELA 5- PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM FARINGITE AGUDA (CID J02) N=13.527	74
TABELA 6 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM CERUME IMPACTADO (CID H61.2) N=11.755.....	75
TABELA 7 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM INFECÇÃO AGUDA DAS VIAS AÉREAS SUPERIORES NÃO ESPECIFICADA (CID J06.9) N=11.233	76
TABELA 8 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM AMIGDALITE AGUDA NÃO ESPECIFICADA (CID J03.9) N=11.062.....	77
TABELA 9 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM SINUSITE MAXILAR AGUDA (CID J01.0) N=7.616	78
TABELA 10 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM NASOFARINGITE AGUDA (CID J00) N=7.432	79
TABELA 11 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM TOSSE (CID R05) N=6.615	80
TABELA 12 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM DESVIO DE SEPTO NASAL (CID J34.2) N=4.915.....	81
TABELA 13 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM EXAME MÉDICO GERAL (CID Z00.0) N=4.502.....	82
TABELA 14 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO LARINGITE AGUDA (CID J04.0) N=4.321	83
TABELA 15 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM OTITE EXTERNA NÃO ESPECIFICADA (CID H60.9) N=4.292 .	84

TABELA 16 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM TONTURA E INSTABILIDADE (CID R42) N=3.880	85
TABELA 17 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM OUTRAS OTITES MÉDIAS GUDAS NÃO-SUPURATIVAS (CID H65.1) N=3.595	86
TABELA 18 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM APNEIA DO SONO (CID G47.3) N=3.181	87
TABELA 19 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM PERDA DE AUDIÇÃO (CID H91) N=3.162	88
TABELA 20 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM DOR DE GARGANTA (CID R07.0) N=2.669	89
TABELA 21 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM OUTRAS RINITES ALÉRGICAS (CID J30) N=2.644	90
TABELA 22 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM FARINGITE AGUDA DEVIDA A OUTROS MICROORGANISMOS ESPECIFICADOS (CID J02.8) N=2.393	91

LISTA DE ABREVIATURAS E SIGLAS

ANA	- Agência Nacional das Águas
BIREME	- Biblioteca Regional de Medicina
CDC	- <i>Centers for Disease Control and Prevention</i>
CID	- Classificação Internacional de Doenças
CFM	- Conselho Federal de Medicina
DeCS	- Descritores em Ciências da Saúde
INPI	- Instituto Nacional de Pesquisa Industrial
IR	- <i>Information retrieval</i>
KR	- <i>Knowledge retrieval</i>
MeSH	- <i>Medical Subject Headings</i>
NLM	- <i>U.S. National Library of Medicine</i>
OMS	- Organização Mundial da Saúde
OPAS	- Organização Panamericana de Saúde
PEP	- Prontuário Eletrônico do Paciente
PLN	- Processamento Automático de Linguagem Natural
RC	- Recuperação do Conhecimento
RI	- Recuperação da Informação
SINPE	- Sistema Integrado de Prontuários Eletrônicos
SNOMED	- <i>Systematized Nomenclature of Medicine Clinical Terms</i>
SQL	- <i>Structured Query Language</i>
TI	- Tecnologia da Informação
UMLS	- <i>Unified Medical Language System</i>
XML	- <i>Extensible Markup Language</i>

LISTA DE SÍMBOLOS

© - *Copyright*

® - Marca registrada

SUMÁRIO

1	INTRODUÇÃO	13
1.1	OBJETIVOS	15
2	REVISÃO DE LITERATURA.....	16
2.1	INFORMÁTICA EM SAÚDE	16
2.2	TERMINOLOGIA EM SAÚDE E RECUPERAÇÃO DA INFORMAÇÃO	17
2.3	SISTEMA INTEGRADO DE PROTOCOLOS ELETRÔNICOS (SINPE©)	22
2.4	<i>BIG DATA</i> EM SAÚDE	25
3	MATERIAL E MÉTODOS.....	30
3.1	TIPO DE ESTUDO	30
3.2	ANÁLISE DO COMITÊ DE ÉTICA.....	30
3.3	LOCAL DO ESTUDO.....	30
3.4	AMOSTRA.....	30
3.5	CRITÉRIOS E INCLUSÃO E EXCLUSÃO.....	30
3.6	MÉTODO.....	31
3.6.1	Planejamento da coleta de dados	31
3.6.2	Extração da base de dados	31
3.6.3	Elaboração de <i>word cloud's</i>	32
3.6.4	Análise dos dados	33
3.6.5	Desenvolvimento de interface	33
4	RESULTADOS.....	35
4.1	CONSTRUÇÃO DAS WORD CLOUDS POR DIAGNÓSTICOS MAIS PREVALENTES	36
4.2	CRUZAMENTO DO PRONTUÁRIO COM OS TERMOS DO PROTOCOLO MESTRE DO SISTEMA SINPE©.....	46
4.3	DISTRIBUIÇÃO POR TERMOS ENCONTRADOS	48
4.4	DESENVOLVIMENTO DA INTERFACE ENTRE O SISTEMA CLINIC® E O SISTEMA SINPE©.....	56
5	DISCUSSÃO	58
6	CONCLUSÕES	62
7	REFERÊNCIAS.....	63
	GLOSSÁRIO.....	72

APÊNDICE A – TABELAS DE DISTRIBUIÇÃO POR DOENÇA	73
APÊNDICE B – LISTA DE <i>STOP WORD'S</i> UTILIZADA NA ELABORAÇÃO DAS <i>WORD CLOUD'S</i>.....	92
ANEXO A – ANÁLISE DO COMITÊ DE ÉTICA	93
ANEXO B – CONTRATO DE LICENÇA DE USO DO SOFTWARE SINPE.....	97

1 INTRODUÇÃO

Derivada do latim *promptuarium*, a palavra prontuário é definida como “livro, manual que contém fórmulas e indicações úteis de modo a achar prontamente o que se quer saber” ou ainda “lugar onde se arrecadam coisas que de um momento para outros pode ser precisas” (FERREIRA, 1999).

Os prontuários em papel foram substituídos por registros eletrônicos e o que era guardado em armários, fichários e arquivos, agora repousa em discos rígidos de computadores e/ou servidores virtuais. O acesso e uso dos dados dos pacientes, antes difícil e complicado, passou a ser instantâneo e compartilhado por múltiplos profissionais simultaneamente. Enquanto os profissionais de saúde são encorajados a aderir ao seu uso, com propostas de melhoria do processo de atendimento, alguns colocam em dúvida sua segurança e até, seu conteúdo e a sua recuperação (COLLEN et al., 2015).

O prontuário eletrônico do paciente (PEP) é um meio automatizado que visa substituir o registro manual de: história clínica; solicitação de exame; emissão de receitas; e outras documentações de história de saúde, tendo por objetivo dar continuidade ao tratamento pela equipe multiprofissional (CONSELHO FEDERAL DE MEDICINA (CFM), 2002).

O crescimento e a adoção da tecnologia em saúde, no cuidado direto ao paciente, pode ser visto pelo aumento no percentual de profissionais na área da saúde, nos Estados Unidos. Em 2001, o número de usuários de sistemas de prontuário eletrônico era 18,2% e, em 2012, este número mostrou um crescimento para 71,8% (ONG, 2015).

Ainda que se apresentem incontáveis ganhos, principalmente na última década, no armazenamento digital de textos médicos incluindo: história clínica; resultados e imagens de exames; laudos e integração entre sistemas, grande parte destas informações encontra-se na forma de textos livres, não estruturados, dificultando a recuperação formal destas informações e sua transformação em conhecimento (AUERBUCH et al., 2004).

A grande maioria do conhecimento biomédico é explicitada de forma não estruturada, seja em textos relato de casos ou artigos. Um passo importante para a extração destas informações é o reconhecimento e a normatização de termos que possam ser estruturados nestes textos. Uma das soluções para este problema é a

construção de dicionários de dados. Esses dicionários criam uma relação entre o que está escrito de forma livre e um termo que possa servir como generalização, por exemplo, de um diagnóstico, sintoma ou doença (KANG et al., 2013).

A recuperação e descoberta das informações em textos médicos é uma tarefa difícil, por não existir uma padronização em sua escrita. Isso pode levar a ambiguidades nos textos recuperados e, por isso, a utilização de técnicas que auxiliem esta tarefa é de vital importância. Ainda, segundo os autores, os dados médicos podem ser divididos em três grupos: numéricos; narrativos; imagens e gráficos (SHORTLIFFE; BARNETT, 2001).

Recuperar a informação é criar uma forma para trazê-la novamente ao usuário ou a um conjunto deles. Sua demanda pode surgir de forma espontânea e/ou induzida e seu objetivo pode ser tanto o de somente resgatar um conteúdo, como também o de construir um novo conhecimento (CHOO, 2003).

O processo de recuperação de textos surge como uma ferramenta que possibilita uma comunicação entre homem-máquina, extraindo informação de grandes bases narrativas textuais denominadas *corpus*, realizando uma análise estatística de ocorrência de palavras, sequência de palavras, termos e transformando o resultado em conhecimento (VARASCHIN et al., 2001).

Visando a elaboração de protocolos médicos eletrônicos, foi desenvolvido um sistema de arquitetura hierarquizada para protocolos em saúde, denominado Sistema Integrado de Protocolos Eletrônicos - SINPE©. O Sistema está registrado no Instituto Nacional de Propriedade Intelectual – INPI sob número RS 06056-1 (MALAFAIA et al., 2003; PINTO, 2005).

Diversos trabalhos já foram publicados utilizando o Sistema SINPE©, que se mostrou muito eficaz na ordenação e construção do conhecimento médico (NASSIF et al., 2011).

1.1 OBJETIVOS

1. Extrair palavras de textos livres não estruturados em prontuário eletrônico e comparar sua frequência com um protocolo já existente de termos, das doenças de maior prevalência em uma base de dados.
2. Criar uma interface de ligação para o preenchimento automático dos sintomas e doenças, entre os sistemas CLINIC® e SINPE©.

2 REVISÃO DE LITERATURA

2.1 INFORMÁTICA EM SAÚDE

A informática médica é a ciência voltada ao estudo da aplicação dos recursos da informação na área da saúde (SHORTLIFFE, 1984).

Em 1996, Smith citou que, no futuro, a ferramenta de informação do médico poderia ser algum tipo de combinação entre o registro das informações do paciente e a internet e que, provavelmente, não haveria uma única ferramenta, mas uma família delas, sofisticadas e que possibilitassem a interpretação e a busca de informações sobre queixas e associação das doenças (SMITH, 1996).

Para Rezende e Abreu (2013), a Tecnologia da Informação (TI) é um conjunto de recursos tecnológicos e computacionais para a geração e uso da informação que agrega os componentes de *hardware* e *software*. Uma vez que a população esteja na denominada “era da informação”, a TI tornou-se uma parte natural da vida de cada um.

Os registros de pacientes, baseados em sistema de prontuário eletrônico, têm muitos usos, além de seu papel principal no atendimento ao paciente, que inclui extração de dados epidemiológicos e a composição de uma fonte de pesquisa. A precisão destes dados é crítica e afeta diretamente essas aplicações (HOGAN; WAGNER, 1997).

Os sistemas de saúde não se restringem ao nível administrativo, sendo sua grande parte voltada aos cuidados diretos aos pacientes. Os prontuários eletrônicos devem permitir a recuperação rápida e organizada das informações de um indivíduo, ou de um conjunto deles, e podem ser utilizados para a construção de sistemas de apoio e tomada de decisão em saúde, sendo uma de suas definições: “um algoritmo baseado em computação para auxiliar os profissionais em saúde com um ou mais componentes no processo diagnóstico” (MILLER; GEISSBUHLER, 1999; FARIAS et al., 2011; PATRÍCIO et al., 2011).

Segundo a Organização Mundial da Saúde (OMS), um sistema de informação em saúde tem o objetivo de dar suporte à assistência ao paciente, seu tratamento e acompanhamento, otimizando a gestão e os resultados dos programas e serviços de saúde. Ainda para a mesma, o sistema é essencial para monitorar a condição de saúde, o desempenho da promoção, a prevenção e o processo

curativo. Um sistema é composto por componentes inter-relacionados para a aquisição, análise de dados e a provisão das seguintes informações: dados gerenciais; estatísticas de saúde; e literatura em saúde (WORLD HEALTH ORGANIZATION, 2000).

A eficiência dos sistemas informatizados de saúde está relacionada à melhoria do uso de recursos para a realização dos processos desempenhados pelos profissionais de saúde, sejam estes no modelo de gestão ou no cuidado direto ao paciente. O Prontuário Eletrônico do Pacientes (PEP) surge como uma proposta para unir diferentes tipos de dados, produzidos em diferentes formatos, pela equipe multiprofissional (MARIN, 2010).

Ainda segundo a autora, um fator importante nos sistemas de saúde é a comunicação escrita, estimando que os médicos passem aproximadamente 38% do seu tempo escrevendo ou digitando no caso de sistemas informatizados. Para que ocorra a interoperabilidade, é necessária a troca de mensagens padronizadas entre sistemas. A mesma somente acontecerá se houver uma sintaxe padrão, em que a informação compartilhada é “entendida” pelos sistemas, através de um “acordo formal” e do uso de vocabulário padronizado nas mensagens (MARIN, 2010).

O uso e implantação de sistemas de informação hospitalar, que incluam o prontuário eletrônico do paciente devem proporcionar a melhoria da gestão e da qualidade da saúde, garantindo ao paciente a liberdade de acesso a sua informação e a melhoria das rotinas de trabalhos dos profissionais da saúde (PATRÍCIO et al., 2011).

2.2 TERMINOLOGIA EM SAÚDE E RECUPERAÇÃO DA INFORMAÇÃO

Existem dois grandes desafios no uso prontuários eletrônicos: a eficiente transcrição das informações por parte dos médicos, em um formato padronizado; e a identificação de um conjunto mínimo, mas acessível, de variáveis necessárias para avaliar a qualidade e os resultados dos cuidados em saúde (MCDONALD, 1997; ZAHRA et al., 2013).

A palavra “terminologia” especifica um conjunto de termos ou sistema de palavras utilizadas em uma determinada área do conhecimento. Os termos são, portanto, comuns aos envolvidos e possibilitam o crescimento de um assunto

através da troca de experiências, relatadas com palavras semelhantes (CHUTE, 2000).

A expressão Recuperação da Informação (RI), ou *Information Retrieval* (IR), originalmente proposta por Mooers em 1950, abrange o armazenamento, organização e busca de uma coleção de informações. O objetivo de um sistema de RI é a seleção de um conjunto de itens (textos, imagens, vídeos, etc.) que tenha relevância para se referir a um determinado assunto, em um grande conjunto de dados, podendo, neste caso, se referir a um sistema de prontuários de pacientes. Sua utilização é complexa pois, além de aspectos técnicos que envolvem a construção do sistema a ser utilizado, deve-se separar os itens pertinentes dos não pertinentes ao assunto que se está buscando (MOOERS, 1950; BATES, 2012; LARSON, 2012).

As técnicas de RI estão diretamente ligadas aos sistemas eletrônicos de informação, assim como as bases de dados. Sua utilização tem por finalidade equacionar os seguintes itens: a representação da informação; a especificação da busca da informação; e o desenvolvimento de um mecanismo para sua recuperação. As estratégias e modelos de busca podem afetar o resultado esperado, sendo que a disponibilidade de termos que identifiquem o problema a ser pesquisado, assim como a pertinência dos mesmos no assunto, são decisivos para que se obtenha sucesso (MOOERS, 1950; XIE, 2012).

A evolução da Recuperação da Informação é a Recuperação do Conhecimento (RC) ou *Knowledge Retrieval* (KR). Para que se entenda a sua aplicação, deve-se considerar a hierarquia dos seguintes conceitos: dado; informação; conhecimento; sabedoria. Esta hierarquia é representada pelo nível de complexidade em cada etapa. Enquanto a RI obtém como resultado a informação, a RC o faz na forma de conhecimento. Já a sabedoria requer a aplicação dos conceitos de ética e valores e foge, por um momento, da compreensão da tecnologia da informação (YAO et al., 2007; TIPLE; DHANDE, 2012).

A linguagem médica detém um vocabulário difícil, de grande riqueza e complexidade de termos, e seu emprego ser classificado de preciso a vago. Porém, isso não implica na falta de comunicação entre os profissionais e sim na dificuldade de adaptar sistemas computacionais para o uso na prática da documentação e tomada de decisão. Uma das propostas encontradas para resolver o problema da

ambiguidade foi a criação de vocabulário padronizado, visando a uniformização e codificação da linguagem (DEGOULET; FIESCHI, 1996).

A formação do conhecimento depende da informação e esta, dos dados. Identificar conceitos, criar domínios e catalogar termos, contribui para que este processo possa ser utilizado nas ciências da saúde. Com este objetivo, vários formatos de catalogação de termos vêm sendo empregados ao longo da história (NICOLAS et al., 2010).

Por iniciativa da Biblioteca Nacional de Medicina dos Estados Unidos - *U.S. National Library of Medicine* (NLM), foi desenvolvido a partir de 1954, um sistema de metadados médicos em língua inglesa, com o objetivo de indexar artigos na área da saúde. Desta forma nasceu a lista de vocabulários médicos, *Medical Subject Headings (MeSH)*, composta atualmente de 28.000 descritores e mais de 90.000 termos. Inicialmente com o título de *Subject Heading Authority List*, teve desde o seu início, a proposta de possibilitar a recomendação de novos indexadores, visto que deve acompanhar a evolução do conhecimento e o surgimento de novos conceitos (COLLEN; BALL, 2015; NLM - NATIONAL LIBRARY OF MEDICINE, 2017).

Ainda na busca de uma normatização na área, a mesma biblioteca desenvolveu em 1986, um dicionário de vocábulos médicos denominado *Unified Medical Language System* (UMLS), voltado a facilitar a recuperação da informação, utilizando sinônimos, equivalências e relacionamentos semânticos. É composto por três componentes: *Metathesaurus* – um dicionário composto por mais de um milhão de conceitos; *Semantic Network* – uma rede de categorias e relações semânticas e *Specialist Lexicon and Lexical Tools* – conjunto de ferramentas para processamento de linguagem natural (NICOLAS et al., 2010)(COLLEN; BALL, 2015).

Em 1965, o Colégio Americano de Patologia (CAP) publicou a Lista Sistematizada de Nomenclaturas em Patologia (SNOP), descrevendo termos de anatomia e morfologia. Em 1975, a lista foi expandida para criar o *Systematized Nomenclature of Medicine* (SNOMED). Após várias implementações e melhorias, ela transformou-se na versão atual denominada *SNOMED Clinical Terms*, ou *SNOMED CT* (COLLEN; SHORTLIFFE, 2015).

Em 1986, a Biblioteca Regional de Medicina (BIREME), que faz parte do Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde e é um centro especializado da Organização Pan-Americana da Saúde / Organização Mundial da Saúde (OPAS/OMS), desenvolveu um dicionário denominado

Descritores em Ciências da Saúde (DeCS). Trata-se de um vocabulário trilingue (português-espanhol-ínglês) para servir como uma linguagem única na indexação de artigos científicos e demais materiais de pesquisa, com o objetivo de pesquisar e recuperar assuntos da literatura em saúde (PELLIZZON, 2004; PATRÍCIO et al., 2011).

O DeCS foi desenvolvido a partir dos termos MeSH e teve como objetivo principal a unificação de termos comuns para recuperação de informação na área. O DeCS conta atualmente com 33.136 descritores, sendo 28.552 oriundos do MeSH e 4.584 exclusivos. A terminologia UMLS também faz parte da sua composição, com o envio dos termos em português e espanhol. Sua atualização é anual e através dos termos, é possível otimizar e padronizar as buscas em periódicos e outros instrumentos de pesquisa em saúde (BIREME - CENTRO LATINO-AMERICANO E DO CARIBE DE INFORMAÇÃO EM CIÊNCIAS DA SAÚDE, 2017).

Outra vantagem oferecida pelo DeCS é a disponibilização de sinônimos para os termos catalogados, como exemplo, citou-se quatro termos pesquisados na plataforma: febre; otalgia; espirro e tosse. Como resultado são apresentados seus códigos de categoria, tradução para as outras duas línguas e sinônimos, quando presentes (FIGURA 1).

FIGURA 1 - TELAS DE RESULTADO DA CONSULTA DE TERMOS NA PLATAFORMA DECS.



FONTE: (BIREME - CENTRO LATINO-AMERICANO E DO CARIBE DE INFORMAÇÃO EM CIÊNCIAS DA SAÚDE, 2017)

O crescimento das bases de pesquisa e a dificuldade de processar e categorizar a informação, faz com que os médicos tenham cada vez mais dificuldade de separar os dados significantes, voltados às melhores evidências, daqueles considerados como ruídos no processo de decisão. Um terço do tempo dos médicos é dispendido colhendo e sintetizando informação dos seus pacientes e um terço do custo de uma hospitalização está ligado, de certa forma, a estas informações. O problema é que parte dessa informação fica armazenada somente na memória dos profissionais de saúde, não sendo transcrita nos prontuários (SMITH, 1996).

Em uma revisão bibliográfica de trabalhos publicados entre 1999 e junho de 2010 sobre processamento de linguagem natural e mineração de textos, no mecanismo de busca de artigos PUBMED, foram encontrados treze artigos que descreveram a recuperação de informações em sumários de alta e as principais ferramentas utilizadas. Nesses, destaca-se o trabalho de Rosier et al. (2008), que obteve uma sensibilidade maior que 90% e um valor preditivo superior a 95% na extração de dados de textos livres em relatórios cirúrgicos, relacionados com a estimulação cardíaca e desfibrilação, mostrando-se superior à alcançada com a extração manual (apud BULEGON; MORO, 2010).

Em um estudo com duas bases de prontuários, sendo a primeira com 6.732 registros de histórias clínicas em textos livres e outra com 26.072 registros, no qual se buscou semelhança semântica e ortográfica, com base no dicionário de Descritores em Ciências da Saúde – DeCS e a técnica de *stemming*, foram obtidas, após a indexação de todas as histórias, 200 raízes de termos mais frequentes, indicando a possibilidade do uso de técnicas de recuperação de textos na área médica (MARTHA et al., 2010).

Pesquisadores demonstraram a evolução de um dicionário de observações biomédicas dentro do “*Assistance Publique Hôpitaux Paris*”, maior grupo europeu de hospitais universitários, com a criação de dicionários, mapeamento para nomes e códigos de identificação de observação lógica, além da implementação em um sistema de informação em saúde. Como resultado, foi obtido um dicionário compartilhado por 120 laboratórios incluindo 50.000 códigos (VANDENBUSSCHE et al., 2013).

Em 2013, em um estudo de comparação entre terminologias que representavam alergias em pacientes, utilizou-se cinco terminologias médicas (SNOMED CT, NDF-RT, MedDRA, UNII e RxNorm) através de uma análise

comparativa incluindo conteúdo, orientação de conceito e estrutura do vocabulário. O resultado demonstrou uma vantagem para a terminologia SNOMED CT, que foi capaz de representar conceitos exclusivos para codificação e alergias não conhecidas (GOSS et al., 2013).

As narrativas em texto livre contidas nos prontuários eletrônicos podem ser utilizadas para categorização do conhecimento e representação das informações de forma ordenada. O uso de um conjunto de terminologias padronizado é uma das formas de se buscar uma padronização de linguagem (SOARES et al., 2010).

Na década de 1940, George Zingley Zipf¹ estudou a frequência em que as palavras apareciam em uma língua, texto ou conjunto de obras literárias. Essa observação resultou na formulação do enunciado chamado Lei de Zipf, também conhecido como Lei do Mínimo Esforço. Segundo o mesmo, a frequência de palavras em um determinado texto apresenta uma uniformidade representada matematicamente onde a segunda palavra de maior aparecimento num texto teria a frequência de metade da primeira. Já a terceira palavra apareceria com uma frequência de um terço da primeira e assim sucessivamente. A equação para esta relação é: $r \times f = k$, onde r é a posição da palavra, f é sua frequência e k é a constante (ZIPF¹, 1949 apud PERES VANTI, 2002; PIANTADOSE, 2014).

O Princípio de Pareto, ou Princípio de Escassez do Fator, em homenagem ao economista Vilfredo Pareto, refere-se à aplicação da regra conhecida como 80/20 em que aproximadamente 80% dos efeitos provêm de 20% das causas. Aplicável na área da saúde, pode ser utilizado como um método para identificar eventos mais relevantes, aumentando a atenção em fenômenos mais prevalentes e descartando grandes concentrações de dados dispersos (WRIGHT; BATES, 2010; MÜLLER et al., 2014).

2.3 SISTEMA INTEGRADO DE PROTOCOLOS ELETRÔNICOS (SINPE©)

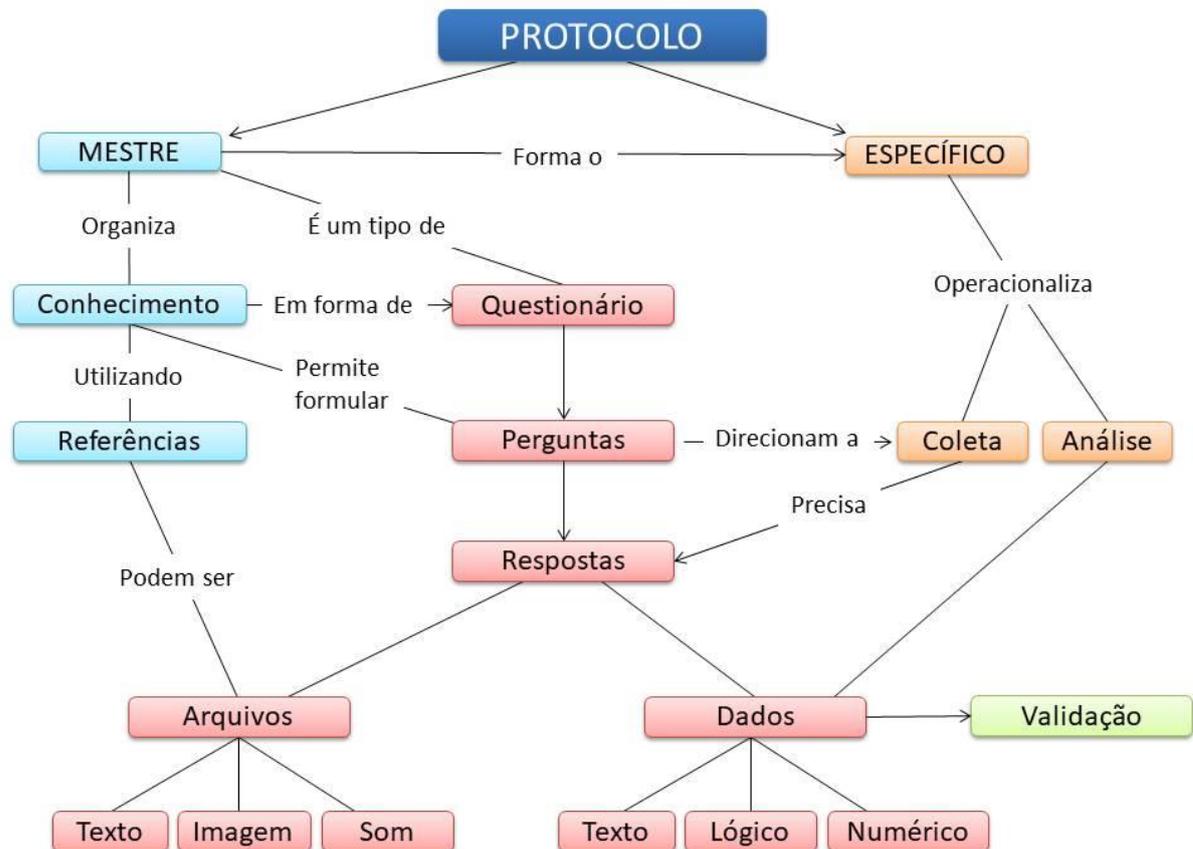
O Programa de Pós-Graduação em Clínica Cirúrgica da Universidade Federal do Paraná vem desenvolvendo, desde a criação da linha de pesquisa Protocolos Eletrônicos em Cirurgia, em 1992, diversas pesquisas na área, com o resultado central no uso de um *software* denominado Sistema Integrado de

¹ ZIPF, G. Human Behavior and the Principle of Least Effort. New York: Addison-Wesley; 1949.

Protocolos Eletrônicos - SINPE©. O sistema conta atualmente com 15 protocolos mestres e 159 protocolos específicos e está registrado no Instituto Nacional de Propriedade Intelectual (INPI) sob o nº 00051543 (MALAFAIA et al., 2003; PINTO, 2005).

O modelo conceitual do Sistema SINPE© (FIGURA 2) é voltado à elaboração destes protocolos e a geração de informação em cinco categorias: item sob análise; características gerais; coleta de dados; gráficos e utilização de coleta (PINTO, 2005).

FIGURA 2 - MODELO CONCEITUAL DO SISTEMA SINPE©



FONTE: Adaptado de PINTO (2006)

Desde o início do seu uso para pesquisa, o Sistema SINPE© foi utilizado em diversos temas de estudos e várias especialidades em saúde, apresentando resultados importantes e expressivos com grande contribuição à pesquisa.

A utilização do Sistema SINPE© demonstrou-se eficaz na coleta, armazenamento e levantamento dos resultados de 407 pacientes analisados quanto

à doença do refluxo gastroesofágico, realizados entre 2001 a 2003 em três hospitais brasileiros (SIGWALT, 2004).

Um protocolo mestre contendo 6694 itens, sendo 392 no item Admissão e 6302 em Evolução, foi desenvolvido num estudo com o objetivo de criar protocolo para coleta estruturada de dados clínicos para pacientes pediátricos em terapia nutricional utilizando o sistema SINPE. A pesquisa provou ser possível sua utilização e a integração de diferentes centros para realização de estudos multicêntricos (ARANHA JUNIOR et al., 2009).

Maniglia (2011) utilizou o Protocolo SINPE© na avaliação de 191 pacientes submetidos à dacriocistorrinostomia, sendo o uso factível à proposta do estudo.

Foram avaliados pacientes submetidos a septoplastia, através da Técnica Eclética Universal de Maniglia, com a utilização do Sistema SINPE© em 90 pacientes, realizando coleta dos dados na consulta médica, no transoperatório no sétimo, trigésimo e nonagésimo dia pós-operatório, resultando numa base eletrônica de protocolos de septoplastia (AMARAL NETO, 2011).

Foi desenvolvido um protocolo específico de doenças da laringe, adicionado ao protocolo mestre de otorrinolaringologia, num estudo envolvendo 208 pacientes submetidos à microcirurgia de laringe, ocasionando na validação do mesmo e aplicação na pesquisa em fonoaudiologia (MACEDO, 2011).

A partir da criação de um protocolo mestre de desvios posturais foi desenvolvido um protocolo específico da função respiratória de pacientes com escoliose idiopática do adolescente, com o processamento e análise do sistema SINPE©, utilizando uma base piloto de 10 pacientes (CANO et al., 2011).

Em um estudo com 4.281 pacientes sobre a incidência das doenças do aparelho digestivo, o sistema SINPE© foi utilizado com sucesso para unificar os dados clínicos em um único protocolo mestre, além de possibilitar a análise epidemiológica (KALIL FILHO, 2012).

Catani et al. (2013) coletaram dados de 245 pacientes submetidos à microcirurgia de laringe a fim de criar um protocolo mestre em otorrinolaringologia e outro específico de doenças laríngeas.

Berger et al. (2013) utilizaram o Sistema SINPE© para elaborar e validar um protocolo específico para pacientes com indicação de rinoplastia. Em 170 casos, submetidos a 148 rinoplastias primárias, obtiveram como resultado a incorporação de 954 itens ao protocolo mestre de otorrinolaringologia.

Visando a criação de um protocolo eletrônico para atendimento de pacientes em terapia nutricional enteral domiciliar, foram desenvolvidos, após revisão de material teórico, 1.793 subitens agrupados em nove itens principais sendo: identificação; avaliação nutricional; prescrição dietética; indicações de via enteral; vias de acesso para nutrição enteral; composição da nutrição enteral; sistema de administração; complicações; e reinternamentos. Foi utilizado o sistema SINPE© para organização e criação do protocolo, tendo o mesmo demonstrado sucesso em sua utilização (SCHIEFERDECKER et al., 2013).

Fagundes (2014) avaliou 27 pacientes submetidos à rinoplastia primária e utilizou com sucesso o prontuário mestre de otorrinolaringologia para avaliar a eficácia das suturas intercrura mediais e de rotação da ponta nasal no aumento da rotação e projeção.

Em 2009 e 2016, Kurestzki et al. adicionaram ao sistema SINPE© um recurso para realização automática dos principais métodos estatísticos utilizados em medicina (qui-quadrado, Mann-Whitney, exato de Fisher e t-Student), resultando na incorporação de análise estatística automática no aplicativo e validando seu uso como ferramenta para pesquisa na área médica (KURETZKI, 2009; KURETZKI et al., 2016).

Um protocolo de doenças oftalmológicas, com ênfase em estrabismo, contendo 2.141 itens no protocolo mestre e 20 protocolos específicos na doença, foi desenvolvido e testado com a entrada de dados de 50 pacientes. A pesquisa foi dividida em três fases, sendo: a criação da base teórica; informatização da base utilizando o sistema SINPE© e interpretação das informações com demonstração dos resultados (RIBEIRO et al., 2016).

2.4 *BIG DATA* EM SAÚDE

O aumento exponencial das informações em saúde prenuncia a necessidade de novas formas de manipulação, análise e acesso a estes dados. Não só cresce o número de periódicos, trabalhos e artigos como também os dados de pacientes em bases, atualizadas em tempo real. Algumas ações se fazem necessárias para que toda essa informação não se perca, sendo: a troca dessas informações entre periódicos e bancos de dados; o desenvolvimento de estruturas de reconhecimento

de dados e o aumento da visibilidade; o acesso e suporte científico em pesquisa; e o desenvolvimento de carreira nesta área (HOWE et al., 2010).

A pesquisa clínica está focada na busca de respostas a problemas clínicos que impactam tanto nos pacientes, como no sistema de saúde. Um dos problemas para tornar confiáveis seus resultados é o tamanho da amostra. Além da possibilidade de vieses, a baixa taxa de amostragem pode levar a interpretações errôneas e conclusões incorretas. Bases que contenham os dados de toda uma população podem ser a resposta para o aumento da quantidade de indivíduos em um estudo, desde que se leve em consideração os mesmos cuidados de seleção, qualidade e evidência dos dados nos arquivos selecionados para a pesquisa (SLOBOGEAN et al., 2015).

A informática médica e a tecnologia aplicada à saúde apresentaram um crescimento significativo nas últimas décadas, oferecendo aos pacientes uma oportunidade de maior promoção às suas condições de vida. O crescimento de bases de dados contendo informação em saúde, aliado ao desenvolvimento de novas tecnologias de armazenamento e interpretação destes dados, tanto na captura como no uso em tempo real, contribuiu para o surgimento do termo *Big Data* (KUZIEMSKY et al., 2014).

Big Data pode ser definido de diversas formas, abrangendo as características denominados “4 V’s”: volume; variedade; velocidade; e veracidade. Nestas bases, dados são coletados e armazenados durante a execução de uma atividade ou área em que estão sendo utilizados. Apesar de ser considerada uma grande oportunidade para solução de problemas e abertura de novas oportunidades, alguns desafios devem ser considerados no seu uso, tais como: as técnicas estatísticas a serem empregadas; a validação do seu conteúdo e os cuidados quanto à privacidade e confidencialidade das informações contidas nos registros (ATKINSON, 2014; LEE; YOON, 2017).

Segundo Sinha et al. (2009) o termo *large dataset* - “grande conjunto de dados” - vem sendo referenciado desde meados de 1975, porém, somente apareceu em discussões sobre o contexto de sua efetividade em 1990. Grandes conjuntos podem ser representados pela grande quantidade, mas não necessariamente pela informação que representam, portanto, este volume não garante que os dados possam conter respostas a perguntas feitas dentro do contexto a que pertencem.

Com o objetivo de qualificar um *Big Data* quanto ao seu tamanho, foi proposta uma classificação por tamanho em número de registros: 10^2 - minúsculo; 10^4 - pequeno; 10^6 - médio; 10^8 - grande; 10^{10} - gigante e 10^{12} - monstruoso. Para a análise destes bancos, sugeriu-se a implantação de estágios para sua análise: planejamento da coleta de dados; coleta de dados atualizados; acesso aos dados; verificação inicial; análise dos dados; apresentação dos resultados e conclusão (HUBER, 1994, 1999).

O uso de *Big Data's* nas pesquisas em estudos clínicos vem se consolidando nas últimas décadas com diversos resultados apresentados. Como por exemplo, podemos citar os estudos de prevalência realizados por: Gebregziabher et al. (2012) sobre diabetes em 2002, nos Estados Unidos, com 890.394 indivíduos; Bermejo-Sánchez et al. (2011) em 23 milhões de recém nascidos sobre a doença Amélia e outras anomalias realizado em 4 continentes; e Elshazly et al. (2013) em 1,3 milhões de indivíduos adultos nos Estados Unidos sobre colesterol. Em estudos de fatores de risco podemos citar: Monnelly et al. (2008) em que foram pesquisados 14.433 pacientes em 155 centros médicos de veteranos nos Estados Unidos, de 50 estados sobre dependência de álcool; From et al. (2008) no qual se utilizaram 53.177 casos de uso de contraste em 35.922 pacientes em uma base de dados a fim de identificar a associação entre a incidência de nefropatia e o uso de bicarbonato de sódio; e uma pesquisa em registros clínicos realizada por Zhang e Hemod (2009) em 16.135 pacientes adultos sobre a relação entre a mortalidade na administração de insulina, glicemia e nível de glicose no sangue (apud WANG; KRISHNAN, 2014).

Ainda sobre estes estudos, os autores citam que o *Big Data* possui algumas limitações, nas quais estão incluídas a adequação dos dados, acurácia, dados completos, natureza dos registros e outras medidas de qualidade das bases utilizadas, como a sumarização dos textos. Um problema comum é a falta de registro nos atendimentos, o que pode causar incerteza nos resultados. Ainda assim, concluem que o uso de *Big Data* deve se tornar cada vez mais comum em estudos clínicos e que este crescimento deve acompanhar o aprimoramento e desenvolvimento de novas tecnologias pra sua análise (WANG; KRISHNAN, 2014).

O uso de *Big Data* vem sendo cada vez mais comum e várias inferências estatísticas de análise de causalidade podem ser utilizadas na sua validação. *Big Data's* podem, por exemplo, ser utilizados para validação de estudos observacionais, baseados no registro rotineiro de saúde de uma população inteira,

através de métodos de clusterização para analisar terapêuticas relevantes, em subgrupos de pacientes (BINDER; BLETTNER, 2015).

Em um estudo de busca sistemática sobre as oportunidades e desafios do uso de *Big Data* em cuidados na saúde, demonstrou-se como resultado temas de maior importância: a oportunidade de aumentar a qualidade da saúde de uma população (64%); a manutenção da saúde (61%); a descoberta precoce de doenças (61%); a acessibilidade, qualidade e estruturação dos dados (57%); a promoção da tomada de decisão (39%); a redução de custo na atenção à saúde (36%); o empoderamento dos pacientes (29%); a personalização do atendimento (24%); a globalização (24%); a identificação de fraudes (11%) e detecção de ameaças à saúde (4%). Os autores destacam que pelo menos 50% da literatura pesquisa mencionava o desafio pela estruturação e segurança dos dados e outros 50% a oportunidade na melhoria da qualidade em saúde (KRUSE et al., 2016).

As evidências do uso de *Big Data* em saúde ainda são incertas e escassas, devido tanto às dúvidas quanto às metodologias dos estudos observacionais a serem utilizadas para a extração e análise dos dados, como: qualidade do conteúdo; inconsistências e instabilidades; validação e, até aspectos legais no seu uso. Ainda segundo os autores, a técnica precisa ser integrada à prática para que possam surgir benefícios no seu uso, passando pela validação dos seus resultados (LEE; YOON, 2017).

O potencial no uso de *Big Data* para a melhoria de qualidade da assistência em saúde demonstra ser muito grande, enquanto as bases de dados sofrem um crescimento exponencial, devido à adesão de mais profissionais no uso de registros médicos eletrônicos. Anteriormente, restritos somente a registros administrativos, agora os dados abrangem informações de saúde, biométricos, sociais e toda uma gama de informações que compõe a base do *Big Data* em saúde. Porém, essa variedade não é premissa do sucesso na composição de um modelo de predição (KUZIEMSKY et al., 2014; RUMSFELD et al., 2016).

Buscando determinar problemas com o uso de *Big Data* em saúde, Househ et al. (2017) propôs uma busca de seis temas relevantes sobre o impacto negativo e ressalvas na sua utilização, separado-os em: falhas em *Big Data*; problemas de compreensão; erros em medicação; proteção da privacidade; interoperabilidade e qualidade dos dados; e pesquisa e estabelecimento de evidências. O estudo não exclui os benefícios do uso da técnica para a melhoria nos cuidados de saúde, mas

conclui não existir total preparação para o uso da mesma, citando algumas razões como: a falta de interesse dos profissionais de saúde; a falta de costume da área no uso de dados; as ameaças quanto à segurança dos pacientes; os problemas na qualidade dos dados; os relacionamentos quanto à privacidade dos dados; e a própria falha dos *Big Data's*.

Demonstrar resultados de textos livres não estruturados em *Big Data's* tem sido um dos desafios na sua utilização como base de dados. Uma das ferramentas utilizadas para representação da frequência é denominada *word cloud* ou *tag cloud* (nuvem de palavras). Muito utilizada na última década para representação da informação em redes sociais, fóruns e outros espaços comuns na internet, nos quais trafegam textos livres, a técnica é utilizada para uma interpretação gráfica da frequência com que uma palavra é representada nos textos, pelo tamanho da fonte de letra, além de espaços e cores (LEE et al., 2010; CHI et al., 2015).

Uma das vantagens da representação visual de uma *word cloud* é sua facilidade de compreensão. Ajustes visuais no formato da figura utilizada para sua apresentação também podem ser utilizados, de acordo com o tipo de estudo, do resultado esperado e do público alvo (HASSAN-MONTERO; HERRERO-SOLANA, 2006; VANSTONE et al., 2016).

A ferramenta *word cloud* pode ser utilizada em textos livres, assim como em termos já definidos, denominados também como *tags* ou marcadores. Uma vez que os textos livres contêm palavras que não pertencem ao domínio que se quer estudar é necessário executar a remoção de termos não relevantes, identificados como *stop-words*. Pode-se fazer uso de uma lista pré-definida, ou utilizar listas, geralmente de domínio público, facilitando assim sua implantação e resultados (WILBUR; SIROTKIN, 1992; BARION; LAGO, 2010) .

Em um estudo com 3.426 textos livres respondidos em um questionário aberto com 7.721 pacientes para aferição da qualidade de atendimento em uma unidade de pronto atendimento, o emprego de *word cloud* demonstrou ser viável na extração e representação dos resultados (MARAMBA et al., 2015).

3 MATERIAL E MÉTODOS

3.1 TIPO DE ESTUDO

O estudo configura-se como uma análise observacional, transversal, descritiva.

3.2 ANÁLISE DO COMITÊ DE ÉTICA

O projeto foi aprovado pelo Comitê de Ética em Pesquisa em Seres Humanos do Hospital IPO, sob número 79919417.4.0000.5529/2017, parecer 2.414.444 (ANEXO A).

3.3 LOCAL DO ESTUDO

O local de realização deste trabalho foi o Hospital Paranaense de Otorrinolaringologia – Hospital IPO em Curitiba, Paraná.

3.4 AMOSTRA

A base de dados da pesquisa foi composta por 375.095 registros de prontuários eletrônicos dos pacientes atendidos no Hospital IPO, entre o período de 01 de janeiro de 2016 e 31 de dezembro de 2016. A escolha do período foi por conveniência.

3.5 CRITÉRIOS E INCLUSÃO E EXCLUSÃO

Os critérios de inclusão foram:

1. Campo CID preenchido;
2. Campo histórico preenchido;
3. Campo histórico preenchido com no mínimo cinco caracteres.

Os critérios de exclusão foram:

1. Consultas de pós-operatório até 28 dias;
2. Múltiplos registros do mesmo paciente, em que a data da reconsulta foi inferior a 30 dias da data da primeira consulta.

3.6 MÉTODO

3.6.1 Planejamento da coleta de dados

Os campos do cadastro do paciente e do prontuário eletrônico foram escolhidos para escolha da seleção a ser exportada.

As variáveis selecionadas dos prontuários eletrônicos para exportação, no banco de dados do Sistema foram: histórico; idade, sexo e CID-10. Para o *corpus* documental foi utilizado o campo histórico.

Os nomes e números de ficha dos pacientes nos arquivos de prontuários foram suprimidos da exportação de dados, preservando assim a confidencialidade dos registros.

O banco de dados em uso pela instituição é o *software* Microsoft SQL, versão 2012 *Enterprise*, facilitando a configuração do comando de exportação.

3.6.2 Extração da base de dados

O arquivo contendo as variáveis foi exportado a partir do Sistema CLINIC®, sistema de gestão hospitalar e prontuário eletrônico do paciente, desenvolvido pela empresa Risc Sistemas em Saúde e registrado no INPI sob número 901281239, em uso pelo Hospital desde 1993 e utilizado pelos profissionais de saúde para o registro das informações dos pacientes.

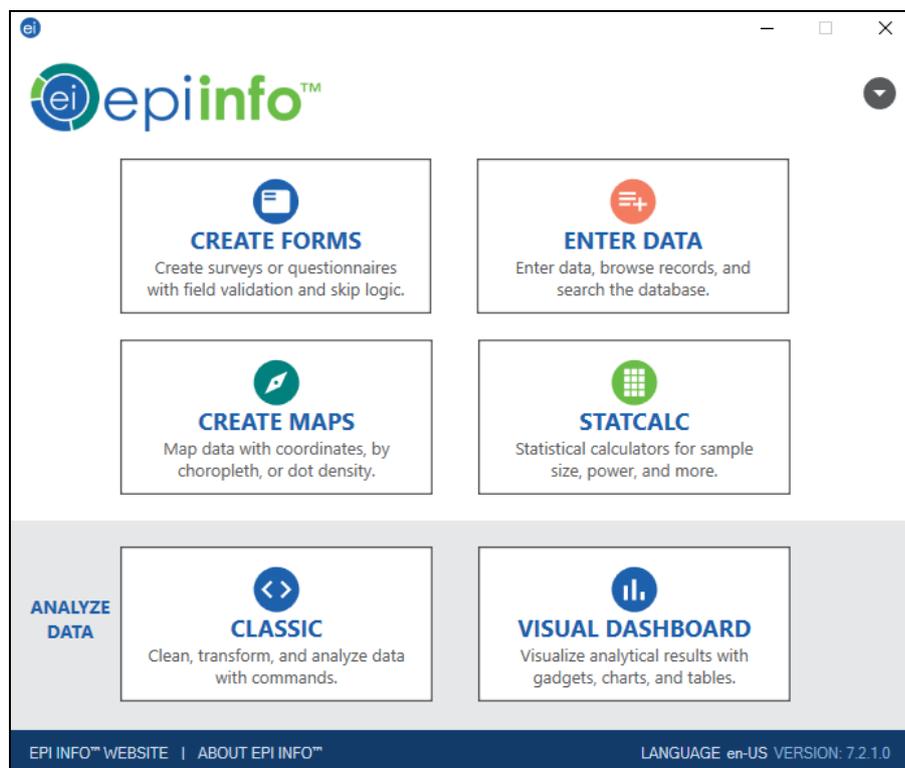
Uma vez selecionadas as variáveis, foi executada a extração do período desejado e a criação de um banco de dados, utilizando também o *software* Microsoft SQL, agrupando todas as variáveis em uma única tabela.

Em prontuários com múltiplos diagnósticos foi somente considerado o diagnóstico principal relatado pelo CID-10 do prontuário.

3.6.3 Elaboração de *word cloud's*

Para a criação das *word cloud's*, das palavras mais prevalentes para cada uma das doenças selecionadas, foi utilizado o aplicativo Epi Info for Windows (FIGURA 3), versão 7.2.1.0, desenvolvido pelo Centers of Disease Control and Prevention (CDC), módulo *Visual Dashboard* (CENTERS FOR DISEASE CONTROL AND PREVENTION (CDC), 2017).

FIGURA 3 - TELA DO APLICATIVO EPI INFO FOR WINDOWS VERSÃO 7.2.1.0



FONTE: CDC, (2017).

Visando padronizar as palavras e facilitar a busca, foi aplicado um filtro de substituição dos caracteres acentuados, assim como normalização somente de caracteres minúsculos, utilizando uma função do Microsoft SQL.

A remoção das palavras indesejadas, denominadas *stop word's*, que incluem pronomes, preposições, conjunções e outras, foi elaborada a partir de uma lista de termos em português, em domínio público (APÊNDICE B) (GITHUBGIST, 2017; SNOWBALL LANGUAGE, 2017).

3.6.4 Análise dos dados

Como protocolo de termos padronizados foi utilizado o Protocolo Mestre em Otorrinolaringologia, disponível no Sistema Integrado de Protocolos Eletrônicos (SINPE©), item sintomas. A lista de palavras foi importada do banco de dados desse sistema utilizando o aplicativo Microsoft SQL versão 12 – *Enterprise*. Para a criação da matriz de contagem de palavras, foram seguidos os seguintes passos:

- Criação de uma lista indexada de CID's dos prontuários;
- Criação de uma lista indexada dos sintomas do protocolo mestre;
- Criação de uma lista indexada cruzando a lista de CID's com a lista de sintomas (exemplo: id=1, CID=J30.4 - sintoma=dor; id=2, CID=J30.4 - sintoma=pressão facial; etc.);
- Contagem de palavras com base na lista cruzada (CID x sintoma) para cada registro de anamnese, correspondente ao CID da lista cruzada;
- Criação da tabela de sintomas por linha e CID por coluna, com a quantidade total mostrando os resultados;
- Exportação do resultado em formato "csv".

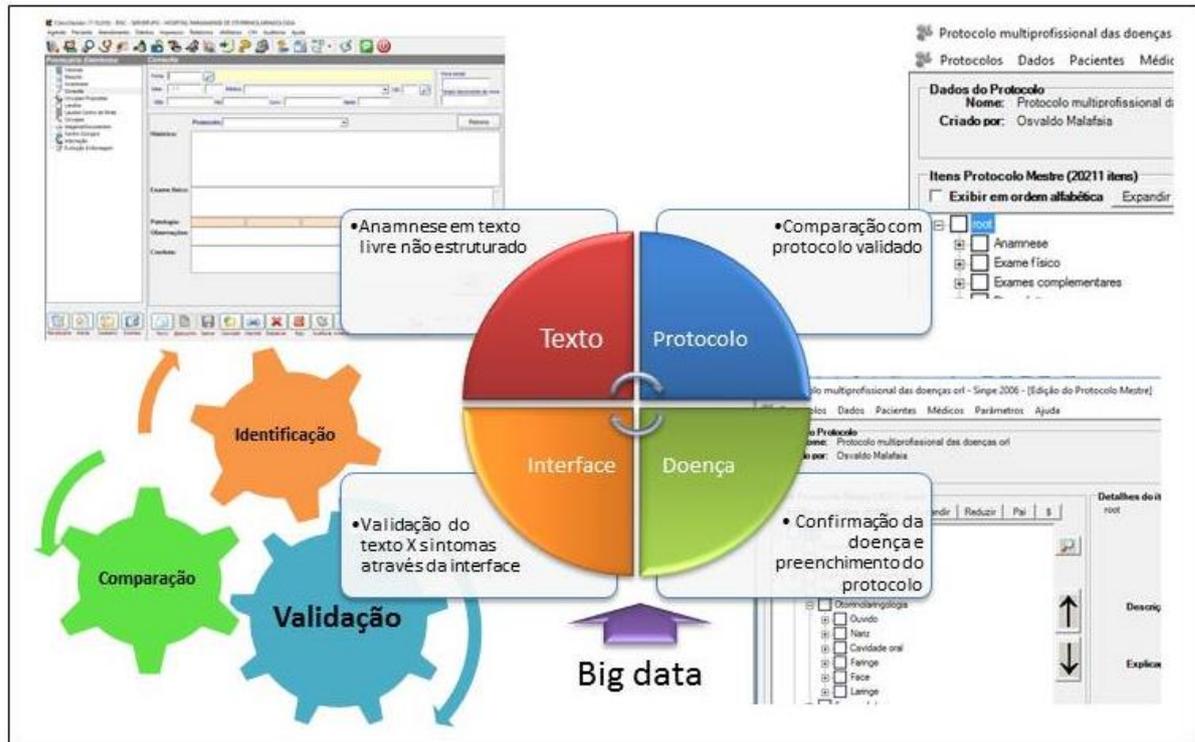
Foi utilizado o sistema SPSS, versão 20.0, desenvolvido pela empresa IBM, para elaboração dos cálculos dos intervalos de confiança, análise dos resultados e formulação das tabelas.

Para análise estatística foi utilizada a prevalência de cada sintoma, estratificada pelas doenças mais prevalentes e calculado o intervalo de confiança de 95%.

3.6.5 Desenvolvimento de interface

A partir da identificação das variáveis, foi desenvolvida uma interface de ligação entre os Sistemas CLINIC® e SINPE© (FIGURA 4) utilizando a linguagem de programação *Visual Basic*, versão 6.0, propondo que os textos livres não estruturados do prontuário eletrônico, campo denominado "histórico", fossem identificados no protocolo mestre em otorrinolaringologia e preenchidos automaticamente, assim como a doença, identificada pelo campo CID.

FIGURA 4 - DIAGRAMA PARA INTERFACE ENTRE OS SISTEMAS CLINIC® E SINPE®.

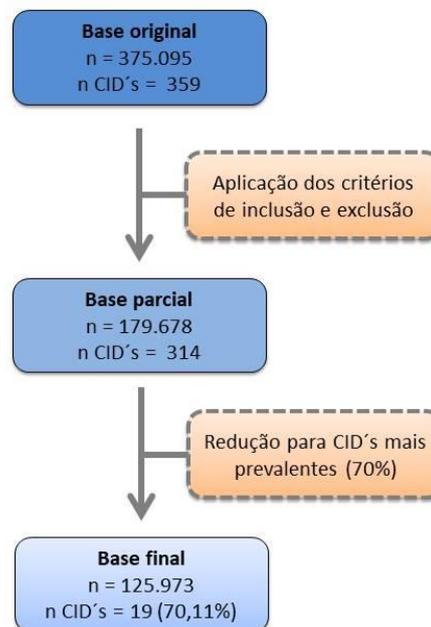


FONTE: O autor (2017).

4 RESULTADOS

A aplicação dos critérios de inclusão e exclusão na base exportada de 375.095 casos resultou num total de 179.678 prontuários. Desses registros, foi calculada a prevalência pela Classificação Internação de Doenças (CID-10), 10ª Edição, e identificadas as doenças de maior frequência, que somaram 125.973 (70,11%) prontuários (FIGURA 5).

FIGURA 5 – DEMONSTRAÇÃO DA REDUÇÃO DA BASE ORIGINAL PARA BASE FINAL



Fonte: O autor (2017).

A amostra foi composta de 125.973 pacientes, sendo 72.110 (57,24%) do sexo feminino e 53.863 (42,76%) do sexo masculino. A média de idade foi de 38,53 anos com desvio padrão de 17,22 variando entre 0 e 99 anos e a mediana foi de 36 anos.

Após a exclusão dos CID com prevalência acumulada menor que 70,11%, o resultado (TABELA 1) demonstra, que das doenças resultantes, as cinco mais prevalentes acumulam 51,40% da amostra.

TABELA 1 - DOENÇAS MAIS PREVALENTES NA AMOSTRA (N=125.973)

CID-10	Descrição	Prevalência	%	% acumulado
J30.4	Rinite alérgica não especificada	17179	13,64	13,64
J02	Faringite aguda devida a outros mic. especificados	13527	10,74	24,38
H61.2	Cerume impactado	11755	9,33	33,71
J06.9	Infecção aguda das vias aéreas sup. não especificada	11233	8,92	42,62
J03.9	Amigdalite aguda não especificada	11062	8,78	51,40
J01.0	Sinusite maxilar aguda	7616	6,05	57,45
J00	Nasofaringite aguda	7432	5,90	63,35
R05	Tosse	6615	5,25	68,60
J34.2	Desvio de septo nasal	4915	3,90	72,50
Z00.0	Exame médico geral	4502	3,57	76,08
J04.0	Laringite aguda	4321	3,43	79,51
H60.9	Otite externa não especificada	4292	3,41	82,91
R42	Tontura e instabilidade	3880	3,08	85,99
H65.1	Outras otites médias agudas não-supurativas	3595	2,85	88,85
G47.3	Apneia do sono	3181	2,53	91,37
H91	Perda de audição	3162	2,51	93,88
R07.0	Dor de garganta	2669	2,12	96,00
J30	Outras rinites alérgicas	2644	2,10	98,10
J02.8	Faringite aguda devida a outros mic. especificados	2393	1,90	100,00

FONTE: O autor (2017).

4.1 CONSTRUÇÃO DAS WORD CLOUDS POR DIAGNÓSTICOS MAIS PREVALENTES

A partir da descoberta dos CID's mais prevalentes, foi utilizado o sistema Epi Info versão 7.2.1.0, para criação das *word cloud's*.

Foram selecionadas as doenças, uma a uma, e aplicada a lista de *stop word's* para exclusão de termos não desejados.

Os resultados são apresentados (FIGURAS de 6 a 24) por ordem de doença mais prevalente.

FIGURA 8 – WORD CLOUD PARA CERUME IMPACTADO (CID H61.2) N=11.755



FONTE: O autor (2017).

FIGURA 9– WORD CLOUD PARA INFECÇÃO AGUDA DAS VIAS AÉREAS SUP. NÃO ESPECIFICADA (CID J06.9) N=11.233



FONTE: O autor (2017).

FIGURA 10 - WORD CLOUD PARA AMIGDALITE AGUDA NÃO ESPECIFICADA (CID J03.9)
N=11.062



FONTE: O autor (2017).

FIGURA 11 – WORD CLOUD PARA SINUSITE MAXILAR AGUDA (CID J01.0) N=7.616



FONTE: O autor (2017).

FIGURA 12 – WORD CLOUD PARA NASOFARINGITE AGUDA (CID J00) N=7.432



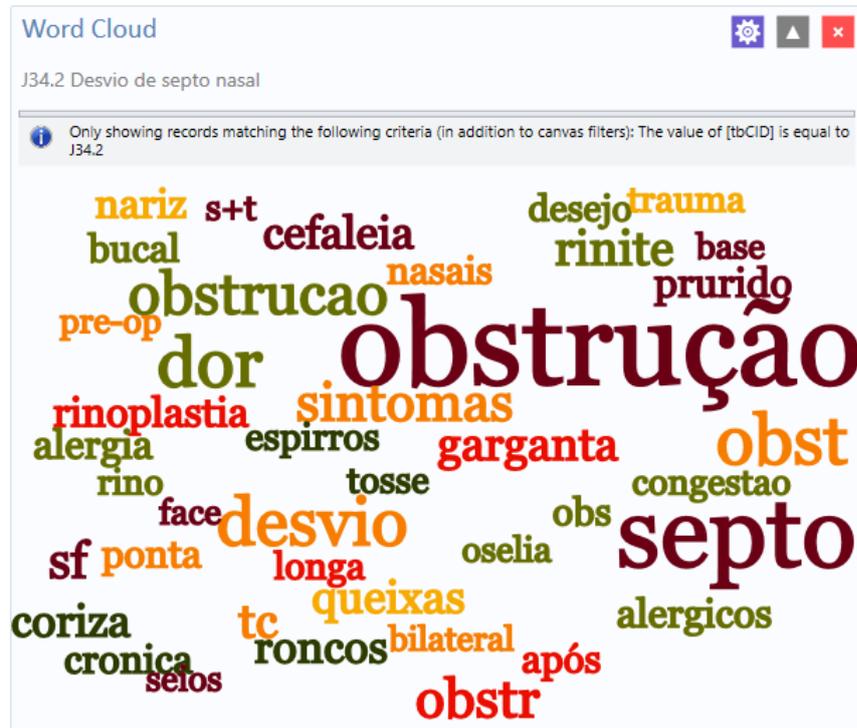
FONTE: O autor (2017).

FIGURA 13 – WORD CLOUD PARA TOSSE (CID R05) N=6.615



FONTE: O autor (2017).

FIGURA 14 – WORD CLOUD PARA DESVIO DE SEPTO NASAL (CID J34.2) N=4.915



FONTE: O autor (2017).

FIGURA 15 – WORD CLOUD PARA EXAME MÉDICO GERAL (CID Z00.0) N=4.502



FONTE: O autor (2017).

FIGURA 16 – WORD CLOUD PARA LARINGITE AGUDA (CID J04.0) N=4.321



FONTE: O autor (2017).

FIGURA 17 – WORD CLOUD PARA OTITE EXTERNA NÃO ESPECIFICADA (CID H60.9) N=4.292



FONTE: O autor (2017).

FIGURA 18 – WORD CLOUD PARA TONTURA (CID R42) N=3.880



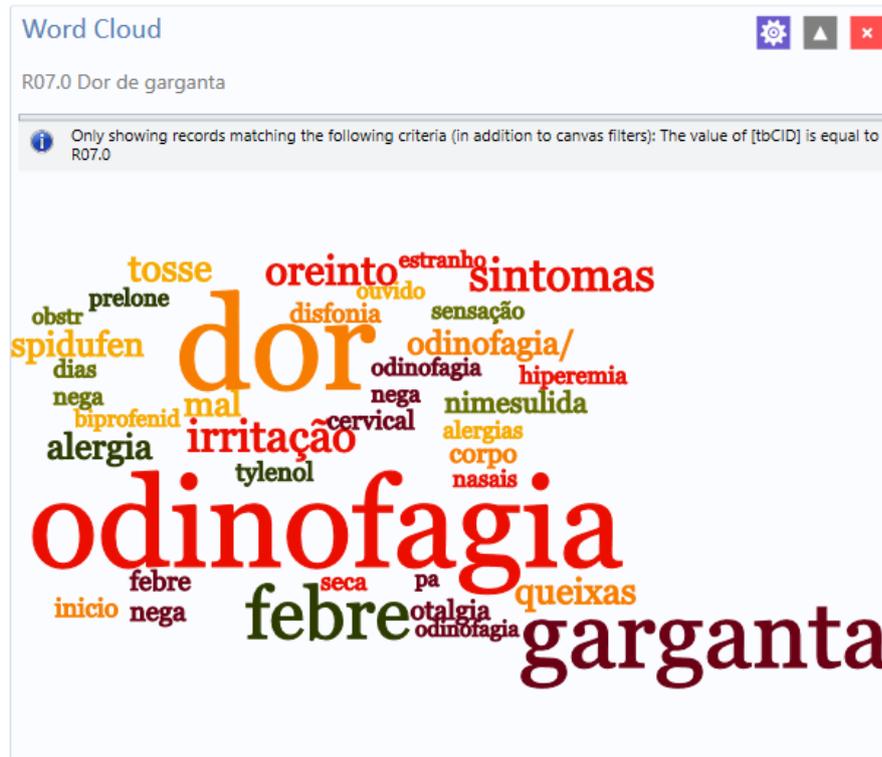
FONTE: O autor (2017).

FIGURA 19 – WORD CLOUD PARA OUTRAS OTITES MÉDIAS AGUDAS NÃO-SUPURATIVAS (CID H65.1) N=3.595



FONTE: O autor (2017).

FIGURA 22 – WORD CLOUD PARA DOR DE GARGANTA (CID R07.0) N=2.669



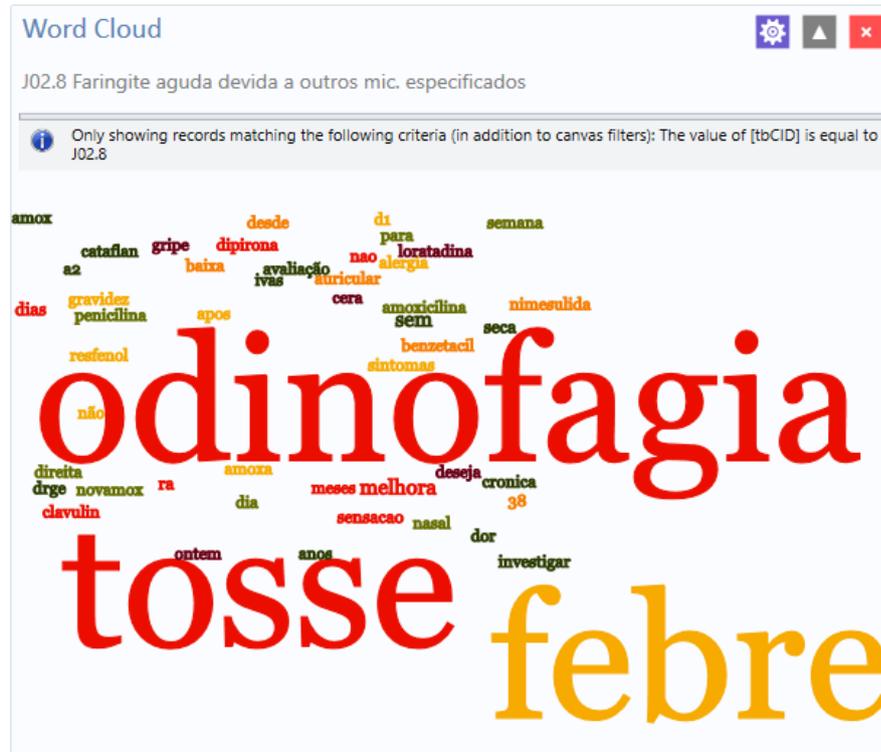
FONTE: O autor (2017).

FIGURA 23 – WORD CLOUD PARA OUTRAS RINITES ALÉRGICAS (CID J30) N=2.644



FONTE: O autor (2017).

FIGURA 24 – WORD CLOUD PARA FARINGITE AGUDA DEVIDA A OUTROS MICROORGANISMOS ESPECIFICADOS (CID J02.8) N=2.393



FONTE: O autor (2017).

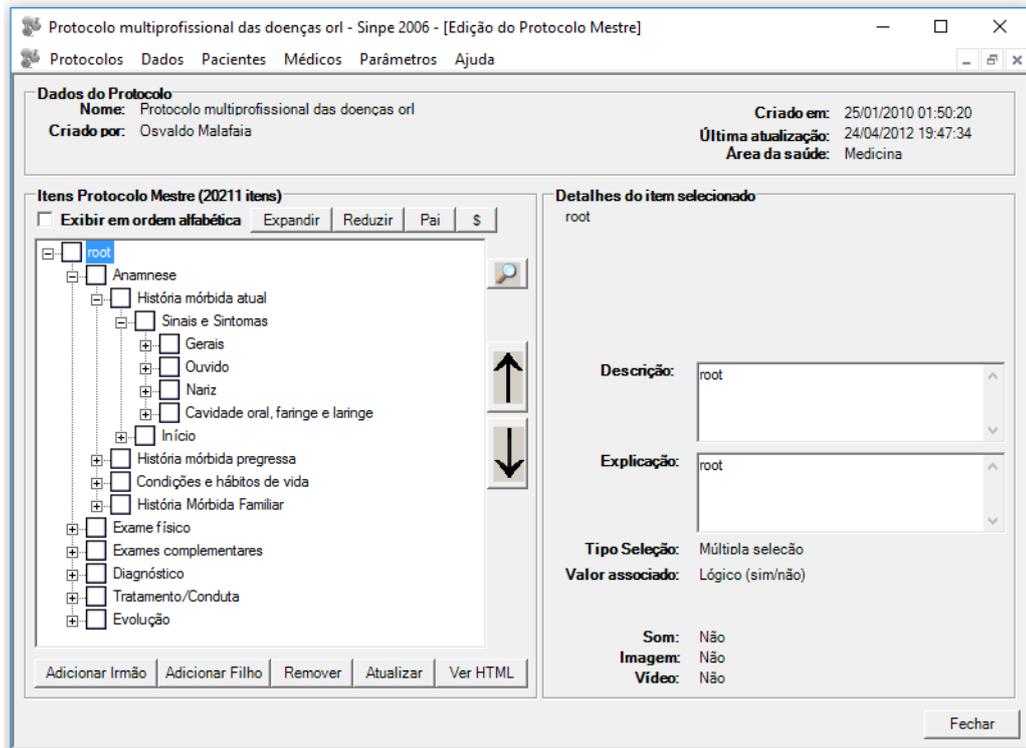
4.2 CRUZAMENTO DO PRONTUÁRIO COM OS TERMOS DO PROTOCOLO MESTRE DO SISTEMA SINPE©

O protocolo mestre de otorrinolaringologia do sistema SINPE© (FIGURA 25), foi utilizado para comparação dos dados de anamnese, nos textos não estruturados. Apresenta a estrutura hierarquizada: Anamnese – História mórbida atual – Sinais e Sintomas, sendo classificado nas subdivisões da otorrinolaringologia: ouvido; nariz; cavidade oral, faringe e laringe; e gerais e dentro de cada uma delas, os termos.

Os termos foram exportados para a mesma base de dados dos prontuários eletrônicos, a fim de ser comparados nos textos livres não estruturados.

Foram isoladas 353 palavras do protocolo representando os sintomas. Para identificar os termos em expressões negativas, foram incluídos na busca os termos “nega”; “ausência de” e “não apresenta”, sendo que esses foram os mais utilizados para expressar a ausência dos sintomas.

FIGURA 25 – TELA DO SINPE© COM O PROTOCOLO MESTRE DE OTORRINOLARINGOLOGIA



FONTE: Sistema SINPE©

A seguir, os CID's foram agrupados em conformidade com as mesmas subdivisões do Sistema SINPE©: Nariz; Ouvido; Cavidade oral, faringe e laringe; Outros (Gerais).

Os CID's com códigos J02; J06.9; J03.9; J01.0; R05; J04.0; G47.3; R07.0 e J02.8 foram agrupados na subdivisão Garganta (TABELA 2) com 55.033 (43,66%) dos diagnósticos e 143.804 (58,55%) dos termos.

Os CID's com códigos J30.4; J00; J34.2 e J30 foram agrupados na subdivisão Nariz (TABELA 2) com 39.785 (31,58%) dos diagnósticos e 61.704 (25,12%) dos termos.

Os CID's com códigos H61.2; H60.9; R42; H65.1 e H91 foram agrupados na subdivisão Ouvido (TABELA 2) com 26.683 (21,18) dos diagnósticos e 35.766 (14,56%) dos termos.

O CID Z00.0, subdivisão Outros, (TABELA 2) apresentou 4.502 (3,57%) dos diagnósticos e 4.331 (1,76%) dos termos.

TABELA 2 - PREVALÊNCIA DE CID'S POR SUBDIVISÃO GARGANTA; NARIZ; OUVIDO E OUTROS (N=125.973)

Subdivisão	Prevalência	%	Termos	%
Garganta	55.003	43,66%	143.804	58,55%
Nariz	39.785	31,58%	61.704	25,12%
Ouvido	26.683	21,18%	35.766	14,56%
Outros	4.502	3,57%	4.331	1,76%
Totais	125.973	100,00%	245.605	100,00%

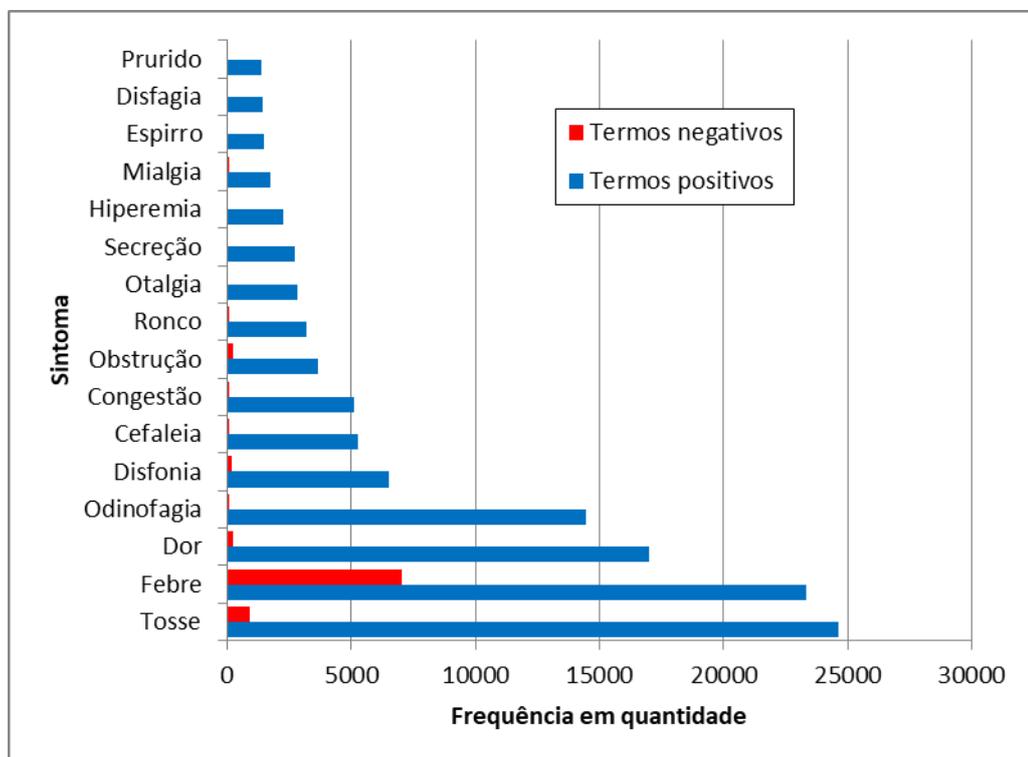
FONTE: O autor (2017).

4.3 DISTRIBUIÇÃO POR TERMOS ENCONTRADOS

Dos 353 sintomas do protocolo mestre em otorrinolaringologia 274 (78%) foram identificados nos textos livres e 79 (22%) não identificados.

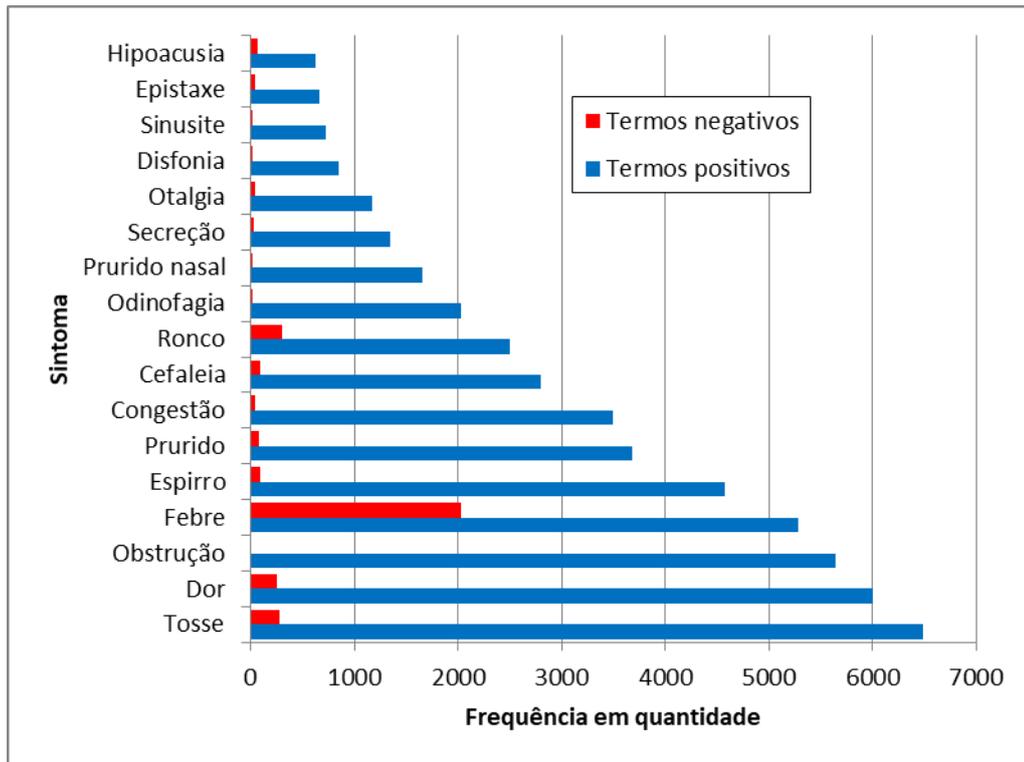
Foram aferidos os termos de maior frequência (GRÁFICOS 1 a 4) e selecionados aqueles que possuíam um resultado acima de um por cento (1%), em uma análise independente para cada uma das subdivisões do protocolo.

GRÁFICO 1 – DISTRIBUIÇÃO DOS TERMOS ENCONTRADOS NA ÁREA GARGANTA (N=143.804)



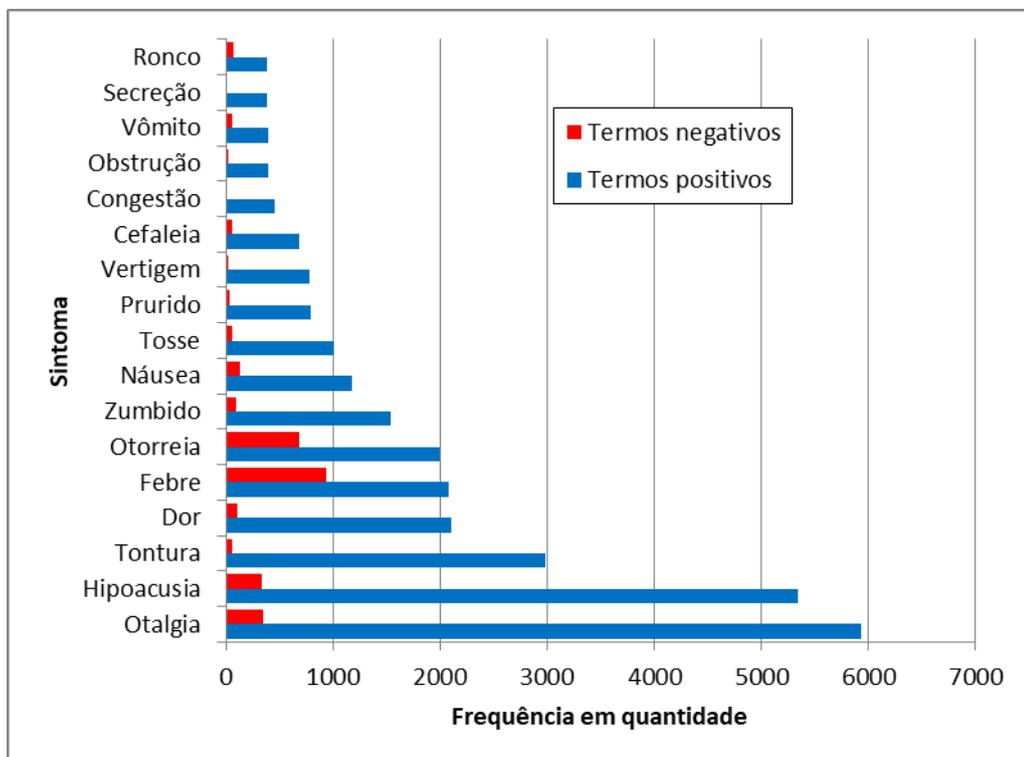
FONTE: O autor (2017).

GRÁFICO 2 – DISTRIBUIÇÃO DOS TERMOS ENCONTRADOS NA SUBDIVISÃO NARIZ (N=61.704)



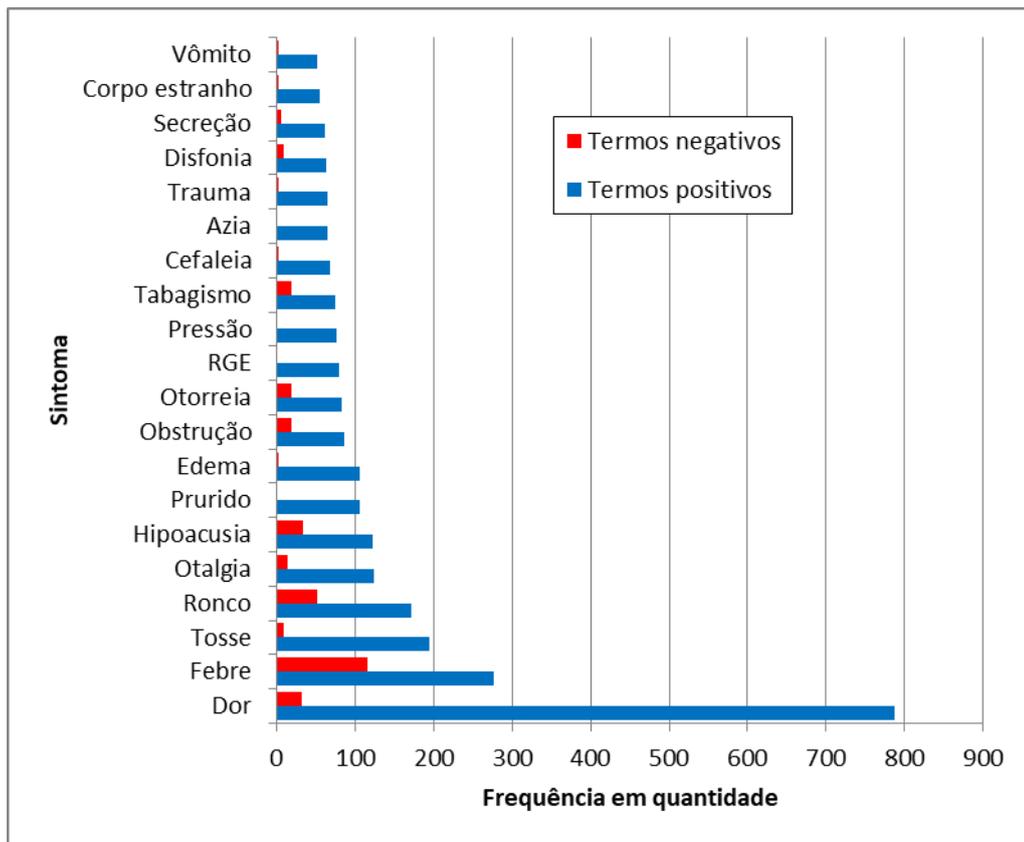
FONTE: O autor (2017).

GRÁFICO 3 – DISTRIBUIÇÃO DOS TERMOS ENCONTRADOS NA ÁREA OUVIDO (N=35.766)



Fonte: O autor (2017).

GRÁFICO 4 – DISTRIBUIÇÃO DOS TERMOS ENCONTRADOS EM OUTROS CID'S (N=4.331)



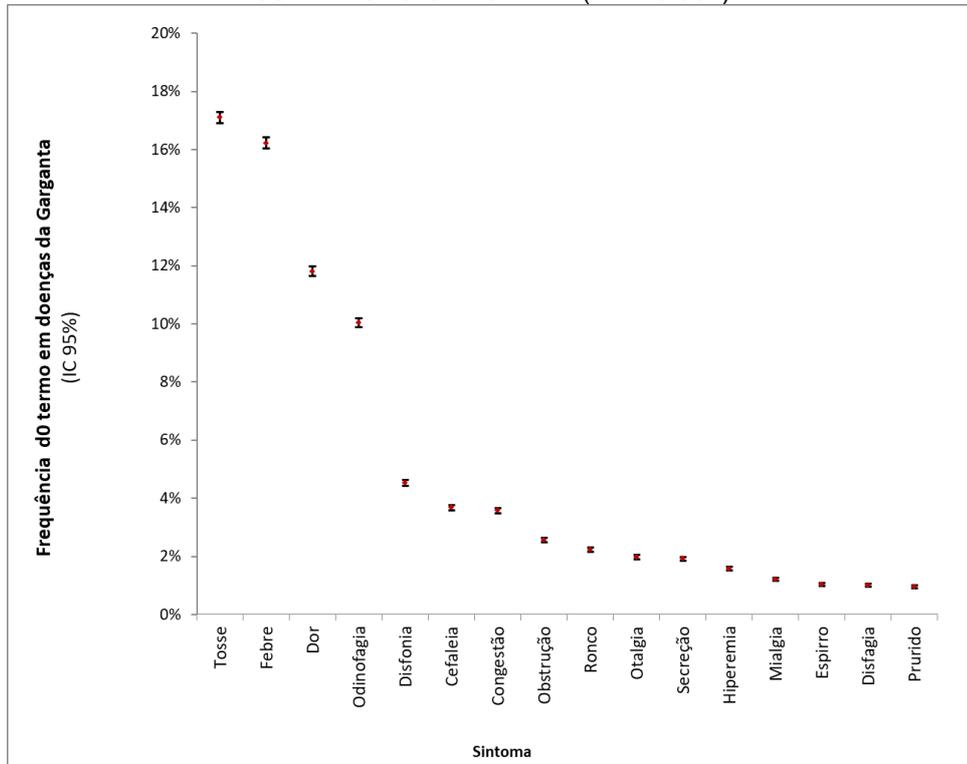
FONTE: O autor (2017).

Para se estimar a prevalência dos sintomas, foram construídos intervalos de confiança de 95%, obedecendo a mesma subdivisão do protocolo (GRÁFICOS 5, 6, 7 e 8).

Pode-se inferir que para 2 sintomas, a não intersecção de seus intervalos de confiança (95%) indica uma significância estatística entre as suas prevalências. Por exemplo, para a subdivisão Garganta, o sintoma tosse apresentou uma prevalência de 17,11% com IC entre 16,9% a 17,3%, já o sintoma febre na mesma subdivisão foi menos prevalente, 16,11%, com IC de 16,0% a 16,4%.

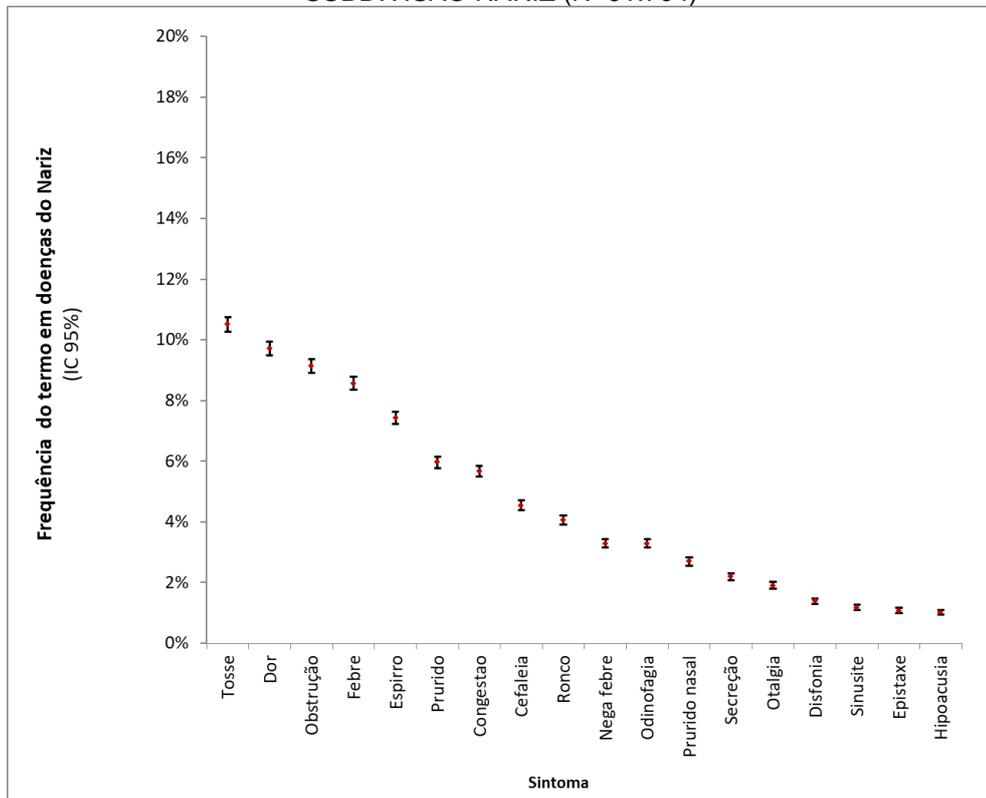
O limite superior do intervalo para o sintoma febre é ainda menor que o limite inferior para o sintoma tosse. A não intersecção entre os dois intervalos confirma que os sintomas tosse e febre têm prevalências significativamente diferentes, sendo tosse o de maior prevalência.

GRÁFICO 5 – FREQUÊNCIA E INTERVALO DE CONFIANÇA (95%) DE PALAVRAS NA SUBDIVISÃO GARGANTA (N=143.804)



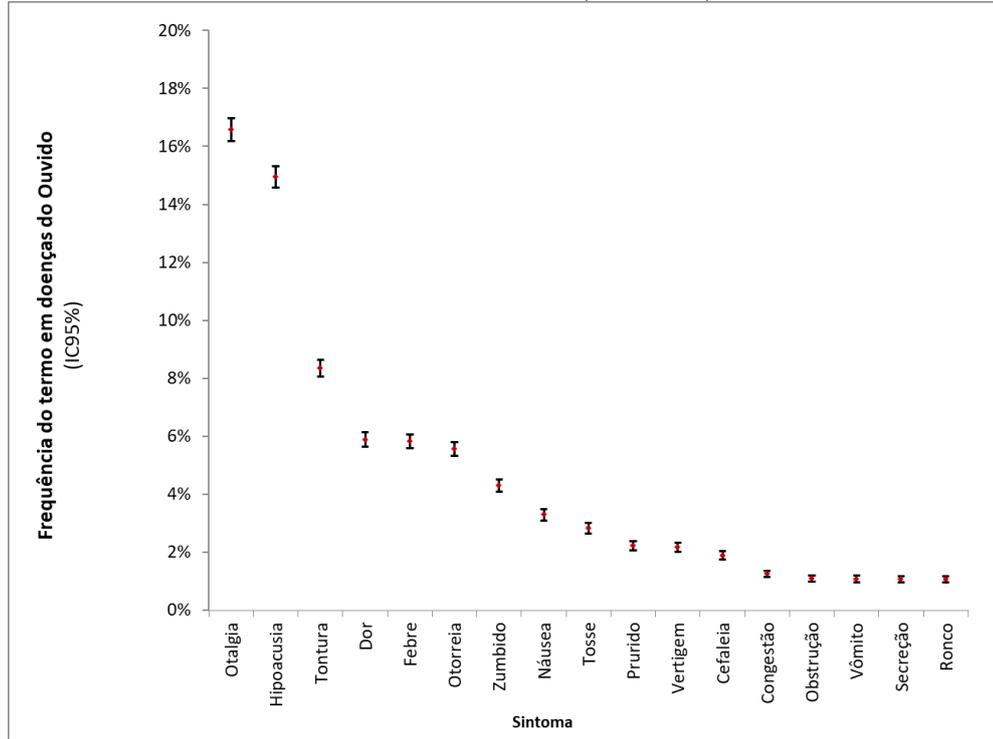
Fonte: O autor (2017).

GRÁFICO 6 - FREQUÊNCIA E INTERVALO DE CONFIANÇA (95%) DE PALAVRAS NA SUBDIVISÃO NARIZ (N=61.704)



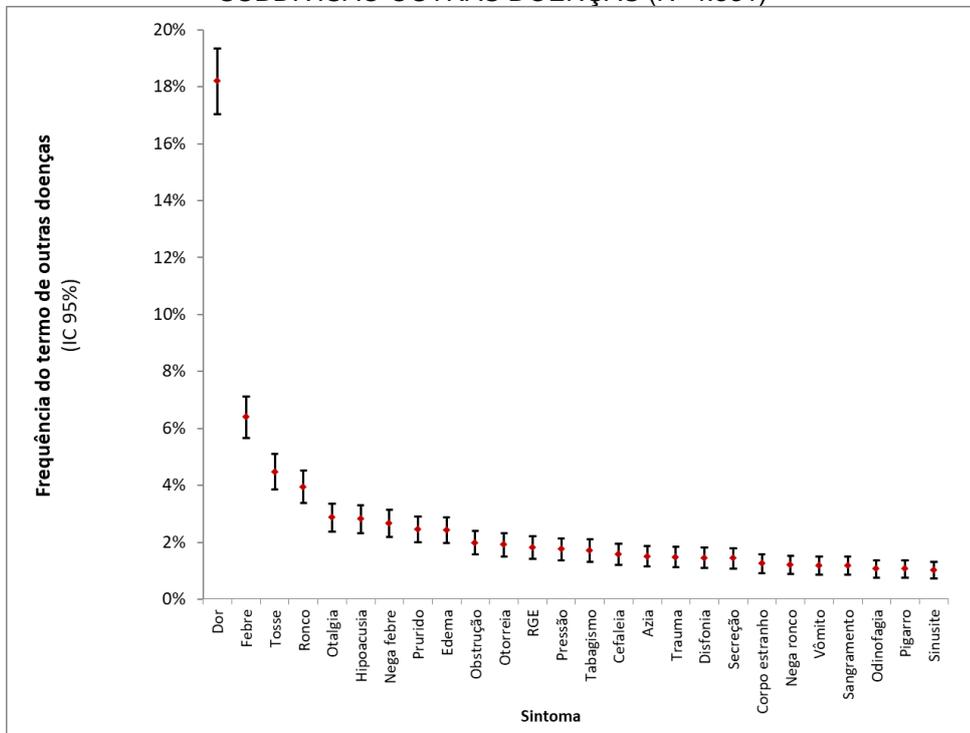
FONTE: O autor (2017).

GRÁFICO 7 - FREQUÊNCIA E INTERVALO DE CONFIANÇA (95%) DE PALAVRAS NA SUBDIVISÃO OUVIDO (N=35.766)



FONTE: O autor (2017).

GRÁFICO 8 - FREQUÊNCIA E INTERVALO DE CONFIANÇA (95%) DE PALAVRAS NA SUBDIVISÃO OUTRAS DOENÇAS (N=4.331)



FONTE: O autor (2017).

As dez palavras mais prevalentes para cada subdivisão foram calculadas e foi elaborada uma tabela demonstrando suas frequências, intervalos de confiança e margens de erro (TABELA 3).

Os três termos de maior frequência na subdivisão garganta: tosse; febre e dor também aparecem nas outras subdivisões, alternando de posição.

TABELA 3 – FREQUÊNCIA DE SINTOMAS POR SUBDIVISÃO (N=245.605)

Subdivisão	Sintoma	n	% de prevalência	% acumulado	Intervalo de confiança (95%)	Margem de erro
Garganta	*Tosse	24.598	17,11%	17,11%	17,1% (16,9% - 17,3%)	0,19%
	**Febre	23.327	16,22%	33,33%	16,2% (16% - 16,4%)	0,19%
	***Dor	16.976	11,80%	45,13%	11,8% (11,6% - 12%)	0,17%
	Odinofagia	14.440	10,04%	55,17%	10% (9,9% - 10,2%)	0,16%
	Disfonia	6.516	4,53%	59,70%	4,5% (4,4% - 4,6%)	0,11%
	Cefaleia	5.287	3,68%	63,38%	3,7% (3,6% - 3,8%)	0,10%
	Congestão	5.136	3,57%	66,95%	3,6% (3,5% - 3,7%)	0,10%
	Obstrução	3.682	2,56%	69,51%	2,6% (2,5% - 2,6%)	0,08%
	Ronco	3.209	2,23%	71,74%	2,2% (2,2% - 2,3%)	0,08%
	Otalgia	2.832	1,97%	73,71%	2% (1,9% - 2%)	0,07%
	Outros	37.801	26,29%	100,00%	26,3% (26,1% - 26,5%)	0,23%
	Total	143.804	100,00%			
Nariz	*Tosse	6.484	10,51%	10,51%	10,5% (10,3% - 10,8%)	0,24%
	***Dor	5.997	9,72%	20,23%	9,7% (9,5% - 10%)	0,23%
	Obstrução	5.644	9,15%	29,37%	9,1% (8,9% - 9,4%)	0,23%
	**Febre	5.288	8,57%	37,94%	8,6% (8,3% - 8,8%)	0,22%
	Espirro	4.579	7,42%	45,36%	7,4% (7,2% - 7,6%)	0,21%
	Prurido	3.679	5,96%	51,33%	6% (5,8% - 6,1%)	0,19%
	Congestão	3.494	5,66%	56,99%	5,7% (5,5% - 5,8%)	0,18%
	Cefaleia	2.805	4,55%	61,54%	4,5% (4,4% - 4,7%)	0,16%
	Ronco	2.502	4,05%	65,59%	4,1% (3,9% - 4,2%)	0,16%
	Nega febre	2.028	3,29%	68,88%	3,3% (3,1% - 3,4%)	0,14%
	Outros	19.204	31,12%	100,00%	31,1% (30,8% - 31,5%)	0,37%
	Total	61.704	100,00%			
Ouvido	Otalgia	5.929	16,58%	16,58%	16,6% (16,2% - 17%)	0,39%
	Hipoacusia	5.346	14,95%	31,52%	14,9% (14,6% - 15,3%)	0,37%
	Tontura	2.985	8,35%	39,87%	8,3% (8,1% - 8,6%)	0,29%
	***Dor	2.107	5,89%	45,76%	5,9% (5,6% - 6,1%)	0,24%
	**Febre	2.084	5,83%	51,59%	5,8% (5,6% - 6,1%)	0,24%
	Otorreia	1.992	5,57%	57,16%	5,6% (5,3% - 5,8%)	0,24%
	Zumbido	1.535	4,29%	61,45%	4,3% (4,1% - 4,5%)	0,21%
	Náusea	1.177	3,29%	64,74%	3,3% (3,1% - 3,5%)	0,18%
	*Tosse	1.012	2,83%	67,57%	2,8% (2,7% - 3%)	0,17%
	Prurido	796	2,23%	69,80%	2,2% (2,1% - 2,4%)	0,15%
	Outros	10.803	30,20%	100,00%	30,2% (29,7% - 30,7%)	0,48%
	Total	35.766	100,00%			
Outros	***Dor	788	18,19%	18,19%	18,2% (17% - 19,3%)	1,15%
	**Febre	277	6,40%	24,59%	6,4% (5,7% - 7,1%)	0,73%
	*Tosse	194	4,48%	29,07%	4,5% (3,9% - 5,1%)	0,62%
	Ronco	171	3,95%	33,02%	3,9% (3,4% - 4,5%)	0,58%
	Otalgia	124	2,86%	35,88%	2,9% (2,4% - 3,4%)	0,50%
	Hipoacusia	122	2,82%	38,70%	2,8% (2,3% - 3,3%)	0,49%
	Nega febre	115	2,66%	41,35%	2,7% (2,2% - 3,1%)	0,48%
	Prurido	106	2,45%	43,80%	2,4% (2% - 2,9%)	0,46%
	Edema	105	2,42%	46,22%	2,4% (2% - 2,9%)	0,46%
	Obstrução	86	1,99%	48,21%	2% (1,6% - 2,4%)	0,42%
	Outros	2.243	51,79%	100,00%	51,8% (50,3% - 53,3%)	1,49%
	Total	4.331	100,00%			

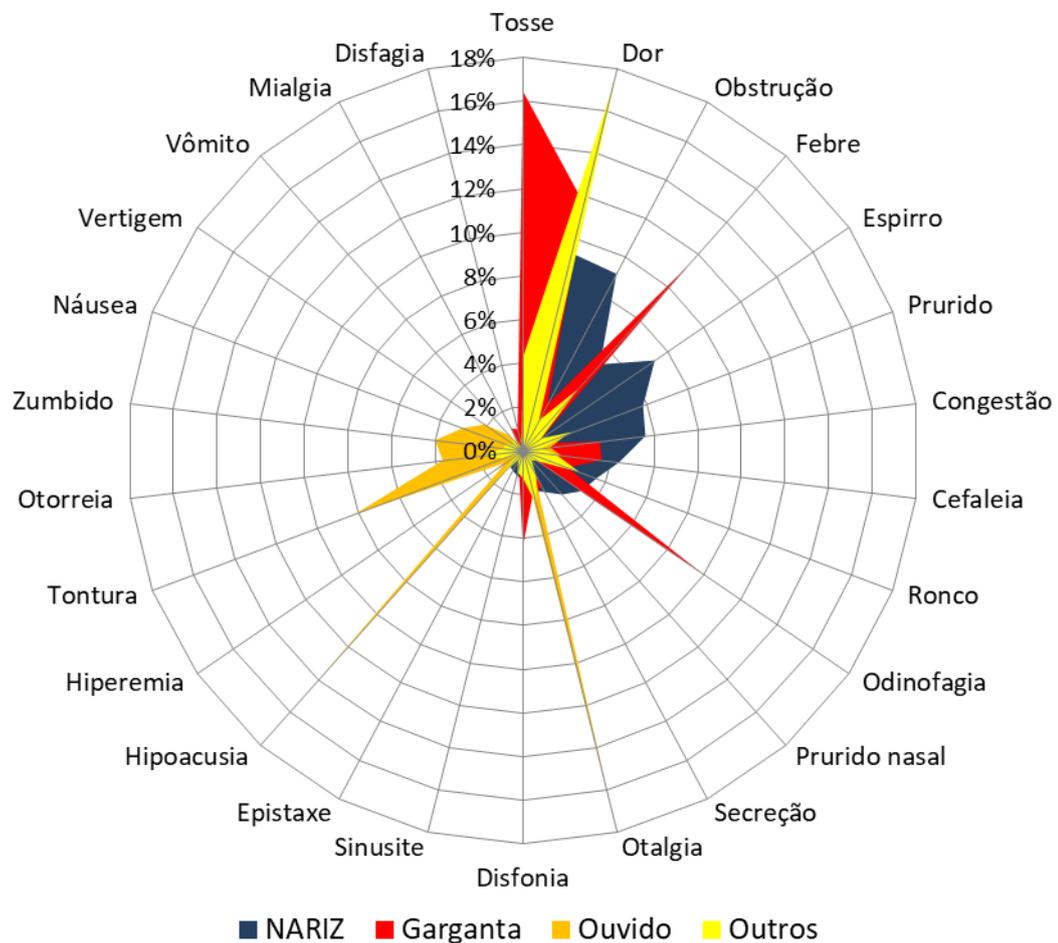
FONTE: O autor (2017).

LEGENDA: Tosse* (1ª-Garganta; 1º-Nariz; 9º-Ouvido; 3º-Outros); Febre** (2º-Garganta; 4º-Nariz; 5º-Ouvido; 2º-Outros); Dor*** (3º-Garganta; 2º-Nariz; 4º-Ouvido; 1º-Outros).

Foram também calculadas as frequências das palavras e os intervalos de confiança (IC=95%) para todas as doenças mais prevalentes (APÊNDICE B, TABELAS 4 a 22) e suas margens de erro.

Uma vez que as palavras poderão aparecer em mais de uma subdivisão: nariz; garganta; ouvido; e outras, para demonstrar a sobreposição desses termos com relação aos CID's, foi elaborado um gráfico do tipo radar (GRÁFICO 9) com a frequência das palavras em cada área.

GRÁFICO 9 – DEMONSTRAÇÃO RADIAL DA DISTRIBUIÇÃO DOS TERMOS POR SUBDIVISÃO



FONTE: O autor (2017).

4.4 DESENVOLVIMENTO DA INTERFACE ENTRE O SISTEMA CLINIC® E O SISTEMA SINPE©

Foi desenvolvida uma interface para interligação entre o sistema de prontuário eletrônico do Hospital IPO - Sistema CLINIC® e o Sistema SINPE©, permitindo que as palavras do campo histórico, identificadas no protocolo mestre, fossem preenchidas diretamente no item “Sinais e Sintomas”. O campo CID confirmado no prontuário, também foi transcrito para o protocolo, no item “Diagnóstico”. Esta rotina não é visível ao usuário dos sistemas.

Em uma simulação, inseriram-se os textos livres no prontuário e reconheceram-se dois termos: otalgia e prurido; e o diagnóstico selecionado de otite externa (CID = H60.9). Os termos foram marcados no Sistema SINPE©, assim como o diagnóstico. Os dados do cadastro do paciente igualmente foram inseridos, de forma automática, pela interface de integração. O comando do Sistema SINPE© para finalizar a coleta também foi executado (FIGURAS 26, 27 e 28).

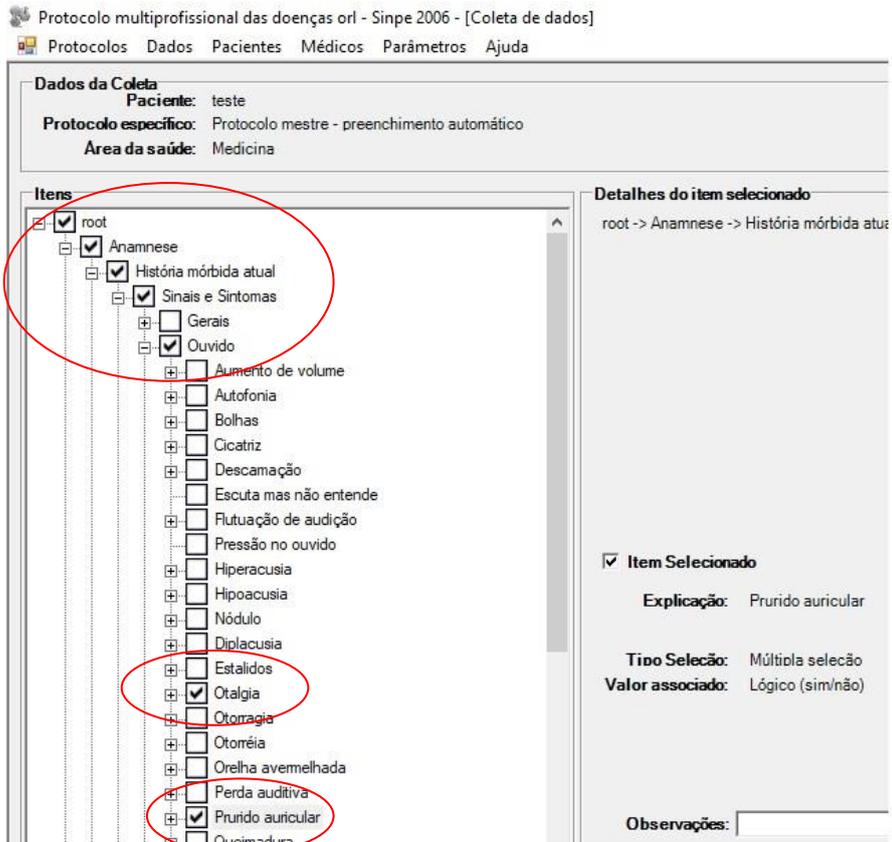
FIGURA 26 – SIMULAÇÃO COM TERMOS E DIAGNÓSTICO

The screenshot shows the CLINIC software interface for a consultation. The top menu bar includes 'Agenda', 'Paciente', 'Atendimento', 'Débitos', 'Impressos', 'Relatórios', 'Utilitários', 'CIH', 'Auditoria', and 'Ajuda'. The main window is titled 'Consulta' and displays the following information:

- Ficha: 869973
- teste
- Data: 19/11/2017 21:44
- Médico: TESTE
- CID: H60.9
- Mãe: Pat: Conv.: AML Idade: 056.10.18
- Protocolo: [dropdown menu]
- Histórico: há 2 dias reclama de otalgia e relatou apresentar prurido

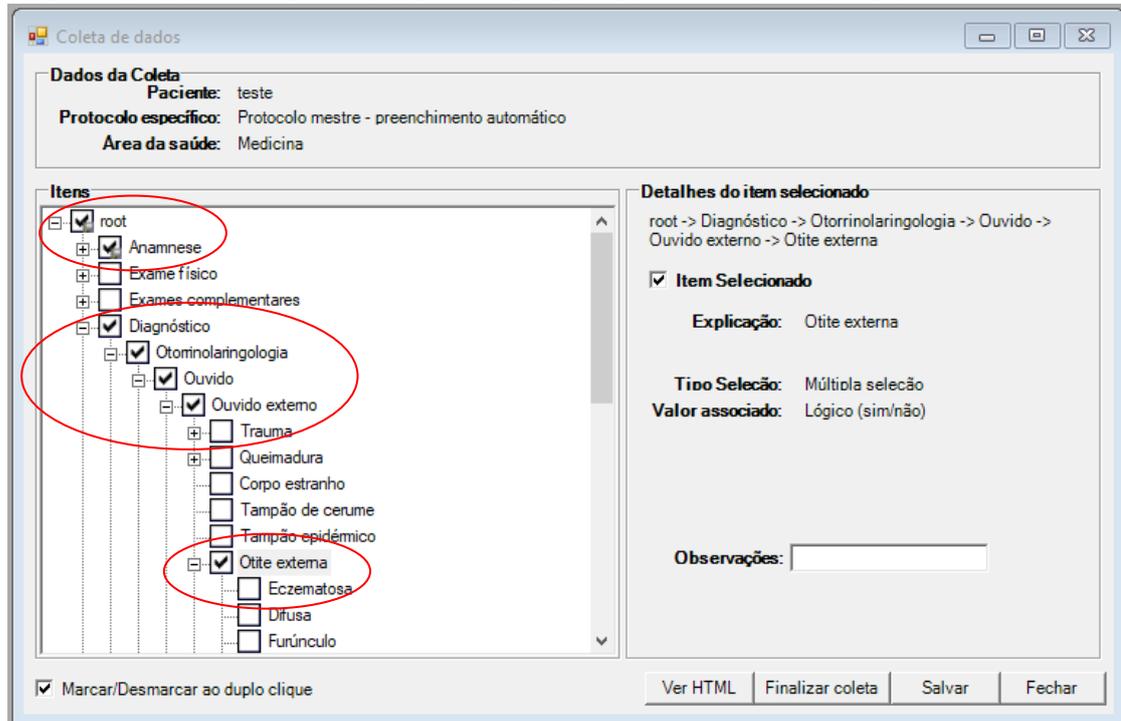
FONTE: O autor (2017).

FIGURA 27 - SISTEMA SINPE© COM SINTOMAS PREENCHIDOS



FONTE: O autor (2017).

FIGURA 28 - SISTEMA SINPE© COM DIAGNÓSTICO PREENCHIDO E COLETA FINALIZADA



FONTE: O autor (2017).

5 DISCUSSÃO

Os resultados dos CID's mais prevalentes demonstram que, apesar da grande quantidade de doenças diagnosticadas no Hospital IPO, existe a concentração em uma quantidade restrita. Apenas 19 CID's representam 70,11% da amostra e desses, cinco representam 51,40%. Essa concentração em uma menor quantidade pode facilitar o monitoramento e a implantação de ferramentas para seu diagnóstico e acompanhamento (NEWMAN, 2005).

A metodologia de busca dos termos utilizada nos textos não estruturados, seguiu uma estratégia definida e as etapas seguidas corroboram com Huber (1999) e Lopes (2002), os quais propõem que o planejamento da estratégia de busca siga aos padrões mínimos para a obtenção dos objetivos propostos.

Quanto à utilização dos CID's como padrão-ouro na determinação dos diagnósticos, não foi empregado outro método de comprovação além do registrado nos prontuários. Isso pode levar a falsos positivos, uma vez que não houve cruzamento. O mesmo foi relatado por Mikalsen et al. (2017), que alertou sobre o risco de vieses em estudos com base nos registros de prontuários eletrônicos.

Outro ponto a ser considerado refere-se aos múltiplos diagnósticos, em que os termos podem indicar uma ou outra doença, ou até, serem comuns a múltiplas. Nesse caso, a identificação será de difícil conclusão automática, devendo passar por um processo de revisão, por parte dos pesquisadores envolvidos na coleta de dados (DIAS-DA-SILVA et al., 2002).

O uso da ferramenta de representação pictórica *word cloud* é ainda pouco aplicado à saúde, na busca de palavras em prontuários eletrônicos. De fácil utilização, pode ser integrada aos sistemas de saúde e vir a servir como um monitor, em tempo real, de termos mais prevalentes, assim como ajudar na detecção de surtos de doenças e no controle epidemiologia de grandes populações em sistemas integrados (LEPENDU et al., 2011; DRESSEL; SCHÜLLE, 2014)

O desfecho das *word cloud's*, criadas para cada especialidade, demonstrou ser possível de implantação no banco de dados de prontuários eletrônicos, considerado um *big data* de tamanho intermediário entre pequeno e médio, mesmo após aplicados os critérios de inclusão e exclusão. As palavras apresentadas em cada um dos CID's de maior prevalência podem servir para composição de novas versões de protocolos mestres e/ou específicos. A ferramenta possui uma facilidade

de leitura e interpretação dos resultados e seu uso pode ser difundido para novas pesquisas na especialidade (SLOBOGEAN et al., 2015; LEE; YOON, 2017).

O resultado apresentou destaque em palavras-chave em todas as doenças, identificando sintomas que podem ser representativos e definirem o diagnóstico. Os textos livres, vistos de forma global, expressam de forma coletiva como os médicos e outros profissionais redigem os prontuários eletrônicos. Comparando os resultados das *word cloud's* de cada diagnóstico com a lista de termos mais frequentes no sistema de protocolos, pode-se identificar uma similaridade de palavras mais frequentes.

Na busca bibliográfica, não foi encontrado nenhum trabalho publicado sobre o uso de *word cloud* na especialidade de otorrinolaringologia, indicando uma oportunidade para próximos estudos. Nos trabalhos publicados em outras especialidades, a ferramenta demonstrou ótimos resultados na recuperação da informação, principalmente quando aplicada em textos livres não estruturados (DRESSEL; SCHÜLLE, 2014; CHI et al., 2015; VANSTONE et al., 2016).

Os resultados do agrupamento de doenças por subdivisões da otorrinolaringologia apresentaram um efeito proporcional ao número de termos encontrados, quando do cruzamento dos textos livres não estruturados com os termos de sintomas, do protocolo mestre de otorrinolaringologia do Sistema SINPE©.

A quantidade de termos encontrados (78%) demonstrou que o protocolo possui representatividade quando comparado à liberdade dos profissionais relatarem o histórico de seus pacientes, em texto livre não estruturado, justificando a manutenção desse tipo de variável nos prontuários eletrônicos. Quanto aos termos do protocolo que não foram identificados no *big data*, faz-se necessária uma reflexão do seu emprego no vocabulário cotidiano da equipe e indica-se uma necessidade de revisão. Sua exclusão não deve ser considerada somente por esse motivo, pois os termos podem fazer referência a doenças de menor prevalência, não representadas na seleção de datas, do banco de dados da pesquisa.

A prevalência de palavras por subdivisão (TABELA 3), valida a ocorrência da distribuição pelo Princípio de Pareto, em que as dez palavras de maior frequência representaram de 68,88% a 73,71% nas três subdivisões. A menor distribuição, 48,21%, na subdivisão de outras doenças, pode ser justificada pela menor

amostragem entre o conjunto, assim como pela menor ocorrência de palavras (NEWMAN, 2005).

O uso de *stemming* (redução ao radical da palavra) não foi empregado na busca por palavras mais prevalentes, uma vez que as mesmas foram comparadas com um protocolo já desenvolvido e defendido por autores citados. Porém, o uso da técnica pode ser uma estratégia a ser empregada no Sistema SINPE© e, com isso, identificar palavras com a mesma raiz morfológica como, por exemplo: pálido; palidez; pálida. Isso poderia aumentar o grau de recuperação e, por sua vez, de confiança e validação dos termos (MARTHA et al., 2010).

O uso de termos negativos foi mais prevalente na palavra “febre”, nas quatro subdivisões, seguido de “tosse” e “dor”. O protocolo de preenchimento não contempla a negação de palavras e o resultado constata a necessidade de um estudo sobre o tema, justificado pela alta frequência com que se apresentaram.

A relação entre sintomas, dentro de uma mesma subdivisão, apresentou significância estatística, justificada pela não intersecção entre os intervalos de confiança de 95%. Isso pode ser observado nos cinco principais sintomas da subdivisão garganta, nos dez principais sintomas de nariz, nos quatro principais sintomas de ouvido e nos dois principais sintomas de doenças de outras subdivisões.

O gráfico radial, indicado para apresentação de dados multivariáveis demonstrou, quando aplicado nos sintomas mais frequentes nas subdivisões, a concentração de alguns termos como: tosse; dor; obstrução; e febre nas subdivisões de garganta, nariz e outras, enquanto a subdivisão ouvido mostrou apresentar termos mais prevalentes de forma isolada como: tontura; hipoacusia e otalgia. Essa demonstração também ficou aparente na aplicação das *word cloud's* relativas a cada uma das subdivisões.

A interface de ligação desenvolvida entre os Sistemas CLINIC® e SINPE©, possibilitou o preenchimento automático dos protocolos, a partir dos textos livres, no protocolo mestre. O texto utilizado para a simulação preencheu os sintomas encontrados no *corpus* não estruturado do prontuário eletrônico bem como o campo diagnóstico.

Uma vez que o protocolo não possui a identificação do campo CID, o preenchimento do diagnóstico foi realizado por equiparação dos termos. Este

procedimento poderia ser otimizado com a incorporação do campo CID no item diagnóstico do Sistema SINPE©, além do uso de uma codificação padronizada.

Da mesma forma, o Sistema CLINIC® não possui corretor de textos nos campos não estruturados, permitindo o erro de escrita ou falha de digitação na entrada do dado. A implantação de um corretor da língua portuguesa poderia reduzir o número de palavras digitadas erradas e que não contribuem para a recuperação da informação.

O protocolo mestre em otorrinolaringologia não possui o recurso de sinônimo e fonetização das palavras. Esse é um problema comum em todas as línguas e uma das maiores dificuldades na interpretação de textos livres. A incorporação de um dicionário de sinônimos, baseado nos Descritores em Ciências da Saúde – DeCS, além da vantagem da padronização, poderia facilitar a busca de artigos científicos, através de casos clínicos, corroborando com Pellizzon (2004).

Com a disseminação e disponibilização de novas bases de prontuários eletrônicos, é importante que o foco da documentação seja centralizado no acesso da informação ao paciente e na utilização desta informação para o processo de tomada de decisão médica.

A automatização de outros sistemas com o sistema SINPE© pode sugerir o uso de bases de dados para o acréscimo de indivíduos de pesquisas em prontuários eletrônicos.

A valorização dos textos livres não estruturados deve ser mantida, para que os profissionais possam traduzir a história clínica do paciente, sem que se tenha a obrigação de utilizar uma forma rígida de entrada de dados.

Estudos que apliquem a mesma técnica em outras bases de dados, de igual ou diferentes especialidades devem ser incentivados para que se obtenham resultados que possam ser comparados e analisados a presente pesquisa.

O uso de protocolos informatizados como o Sistema SINPE© deve continuar contribuindo para a coleta de dados de pesquisa, oportunizando aos pesquisadores uma fonte segura e eficiente para a criação de bases de dados.

Novos estudos devem avançar na busca pela criação de uma padronização da comunicação em saúde, utilizando como base o prontuário eletrônico, incorporando a esse, novos aplicativos que facilitem a troca de informação e oferecendo ao paciente o acesso a seus dados de forma estruturada.

6 CONCLUSÕES

Foram extraídas de textos livres não estruturados as palavras de maior frequência e comparadas com um protocolo de termos já existente, nas doenças de maior prevalência.

Foi desenvolvida uma interface do Sistema CLINIC® com o Sistema SINPE© capaz de preencher, de forma automática, os itens do protocolo mestre, tanto em sintomas como em diagnósticos.

7 REFERÊNCIAS

- AMARAL NETO, O. F. DO. **Utilização do protocolo eletrônico SINPE para a avaliação de pacientes submetidos a septoplastia através da Técnica Eclética Universal de Maniglia**, 2011. Dissertação (Mestrado em Clínica Cirúrgica) - Setor de Ciências da Saúde, Universidade Federal do Paraná, Curitiba. Disponível em: <[http://dspace.c3sl.ufpr.br/dspace/bitstream/handle/1884/26932/Dissertacao Dr. Odin Ferreira do Amaral Neto.pdf?sequence=1](http://dspace.c3sl.ufpr.br/dspace/bitstream/handle/1884/26932/Dissertacao%20Dr.%20Odin%20Ferreira%20do%20Amaral%20Neto.pdf?sequence=1)>. .
- ARANHA JUNIOR, A. A.; CAMPOS, A. C. L.; PINTO, J. S. D. P.; et al. Protocolo eletrônico para coleta estruturada de dados clínicos para pacientes pediátricos em terapia nutricional utilizando o SINPE© (Sistema Integrado de Protocolos Eletrônicos). **Revista do Colégio Brasileiro de Cirurgias**, v. 36, n. 1, p. 73–77, 2009. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-69912009000100013&lng=pt&nrm=iso&tlng=pt>. Acesso em: 9/9/2017.
- ATKINSON, G. DP. Big Data - What is it and whats use is it? **Journal of Ambulatory Care Management**, v. 37, n. 3, p. 196–198, 2014.
- AUERBUCH, M.; KARSON, T. H.; BEN-AMI, B.; MAIMON, O.; ROKACH, L. Context-sensitive medical information retrieval. **Studies in health technology and informatics**, v. 107, n. Pt 1, p. 282–6, 2004. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/15360819>>. Acesso em: 1/11/2017.
- BARION, E. C. N.; LAGO, D. Mineração de textos. **Revista de Ciências Exatas e Tecnologia**, v. 3, n. 3, p. 123–140, 2010. Disponível em: <<http://www.pgsskroton.com.br/seer/index.php/rcext/article/view/2372/2276>>. .
- BATES, M. J. **Understanding information retrieval systems: management, types, and standards**. 1^a ed. Boca Raton: CRC Press, 2012.
- BERGER, C.; FREITAS, R.; MALAFAIA, O.; et al. Electronic data collection for the analysis of surgical maneuvers on patients submitted to rhinoplasty. **International Archives of Otorhinolaryngology**, v. 16, n. 4, p. 497–501, 2012. Disponível em: <<http://www.thieme-connect.de/DOI/DOI?10.7162/S1809-97772012000400011>>. Acesso em: 16/12/2013.
- BERMEJO-SÁNCHEZ, E.; CUEVAS, L.; AMAR, E.; et al. Amelia: a multi-center descriptive epidemiologic study in a large dataset from the international clearinghouse for birth defects surveillance and research, and overview of the literature. **American Journal of Medical Genetics**, v. 157C, n. 4, p. 288–304, 2011.
- BINDER, H.; BLETTNER, M. Big data in medical science: a biostatistical view. **Dtsch. Arztebl. Int.**, v. 112, n. 1866–0452 (Electronic), p. 137–142, 2015.
- BIREME - CENTRO LATINO-AMERICANO E DO CARIBE DE INFORMAÇÃO EM CIÊNCIAS DA SAÚDE. Descritores em Ciências da Saúde. Disponível em: <<http://decs.bvs.br/cgi-bin/wxis1660.exe/decsserver/>>. Acesso em: 3/11/2017.
- BULEGON, H.; MORO, C. M. C. Text mining and natural language processing in discharge summaries. **Journal of Health Informatics**, v. 2, n. 2, p. 51–56, 2010.

CANO, D. V. B.; MALAFAIA, O.; ALVES, V. L. D. S.; AVANZI, O.; PINTO, J. S. D. P. Electronic protocol of respiratory physical therapy in patients with idiopathic adolescent scoliosis. **Revista do Colégio Brasileiro de Cirurgiões**, v. 38, n. 3, p. 177–80, 2011. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/21789456>>. Acesso em: 1/11/2017.

CATANI, G.; CARVALHO, B.; FILHO, J.; et al. Proposed computerized protocol for epidemiological study of patients undergoing microsurgery of the larynx. **International Archives of Otorhinolaryngology**, v. 16, n. 3, p. 346–352, 2013. Disponível em: <<http://www.thieme-connect.de/DOI/DOI?10.7162/S1809-9772012000300008>>. Acesso em: 16/12/2013.

CENTERS FOR DISEASE CONTROL AND PREVENTION (CDC). Epi Info for Windows. , 2017. Atlanta. Disponível em: <<https://www.cdc.gov/epiinfo/pc.html>>. Acesso em: 3/11/2017.

CHI, M.-T.; LIN, S.-S.; CHEN, S.-Y.; LIN, C.-H.; LEE, T.-Y. Morphable word clouds for time-varying text data visualization. **IEEE Transactions on Visualization and Computer Graphics**, v. 21, n. 12, p. 1415–1426, 2015. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/26529462>>. Acesso em: 12/11/2017.

CHOO, C. W. **A organização do conhecimento**. 1ª ed. São Paulo: Senac São Paulo, 2003.

CHUTE, C. G. Clinical classification and terminology: some history and current observations. **Journal of the American Medical Informatics Association: JAMIA**, v. 7, n. 3, p. 298–303, 2000. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=61433&tool=pmcentrez&endertype=abstract>>. Acesso em: 10/10/2017.

COLLEN, M. F.; BALL, M. J. **The history of medical informatics in the United States**. 2nd ed. London: Springer-Verlag, 2015.

COLLEN, M. F.; SHORTLIFFE, E. H. The creation of a new discipline. In: E. H. Shortliffe; M. J. Ball (Eds.); **The history of medical informatics in the United States**. 2nd ed., p.75–120, 2015. London: Springer-Verlag.

COLLEN, M. F.; SLACK, W. V.; BLEICH, H. L. Medical databases and patient record systems. In: E. H. Shortliffe; M. J. Ball (Eds.); **The history of medical informatics in the United States**. 2nd ed., p.207–288, 2015. London: Springer-Verlag.

CONSELHO FEDERAL DE MEDICINA (CFM). **Resolução 1638/2002**. Brasília-DF: Define prontuário médico e torna obrigatória a criação da Comissão de Revisão de Prontuários nas instituições de saúde., 2002.

DEGOULET, P.; FIESCHI, M. Medical language and classification systems. **Computers in health care**. 1st ed., p.242, 1996. Nova York: Springer.

DIAS-DA-SILVA, B. C.; OLIVEIRA, M. F. DE; MORAES, H. R. DE. Groundwork for the development og the brazilian portuguese wordnet. In: E. Ranchhod; N. J. Mamede (Eds.); **Advances in Natural Language Processing**. 1ª ed., p.189–196, 2002. Lisboa: Springer.

DRESSEL, K.; SCHÜLLE, S. Using word clouds for risk perception in the field of public health - the case of Vector-Borne diseases. **GRF Davos Planet@Risk**, v. 2, n. 2, p. 85–88, 2014. Disponível em: <<http://planet-risk.org/index.php/pr/article/view/49/176>>. Acesso em: 12/10/2017.

ELSHAZLY, M. B.; MARTIN, S. S.; BLAHA, M. J.; et al. Non-high-density lipoprotein cholesterol, guideline targets, and population percentiles for secondary prevention in 1.3 million adults: the VLDL-2 study (very large database of lipids). **Journal of the American College of Cardiology**, v. 62, n. 21, p. 1960–1965, 2013.

FAGUNDES, M. S. C. **Avaliação do uso das suturas intercrura mediais e de rotação da ponta nasal no pós-operatório de rinoplastia**, 2014. Dissertação (Mestrado em Clínica Cirúrgica) - Setor de Ciências da Saúde, Universidade Federal do Paraná, Curitiba.

FARIAS, J. S.; GUIMARAES, T. DE A.; VARGAS, E. R. DE; ALBUQUERQUE, P. H. M. Adoção de prontuário eletrônico do paciente em hospitais universitários de Brasil e Espanha: a percepção de profissionais de saúde. **Revista de Administração Pública**, v. 45, n. 5, p. 1303–1326, 2011. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-76122011000500004&lng=pt&tlng=pt>. Acesso em: 2/11/2017.

FERREIRA, A. B. DE H. **Novo Aurélio Século XXI: o dicionário da língua portuguesa**. 3 ed ed. Rio de Janeiro: Nova Fronteira, 1999.

FROM, A. M.; BARTHOLMAI, B. J.; WILLIAMS, A. W.; et al. Sodium bicarbonate is associated with an increased incidence of contrast nephropathy: a retrospective cohort study of 7977 patients at mayo clinic. **Clinical journal of the American Society of Nephrology : CJASN**, v. 3, n. 1, p. 10–8, 2008. American Society of Nephrology. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/18057306>>. Acesso em: 21/10/2017.

GEBREGZIABHER, M.; EGEDE, L.; GILBERT, G. E.; et al. Fitting parametric random effects models in very large data sets with application to VHA national data. **BMC Medical Research Methodology**, v. 12, n. 1, p. 163, 2012. Disponível em: <<http://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-12-163>>. Acesso em: 24/1/2016.

GITHUBGIST. Instantly share code, notes, and snippets. Disponível em: <<https://gist.github.com/alopes/5358189>>. Acesso em: 1/9/2017.

GOSS, F. R.; ZHOU, L.; PLASEK, J. M.; et al. Evaluating standard terminologies for encoding allergy information. **Journal of the American Medical Informatics Association: JAMIA**, v. 20, n. 5, p. 969–79, 2013. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3756252&tool=pmcentrez&rendertype=abstract>>. Acesso em: 1/11/2017.

HASSAN-MONTERO, Y.; HERRERO-SOLANA, V. Improving tag-clouds as visual information retrieval interfaces. **Information Sciences**, v. 1, n. 2, p. 25–28, 2006. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.85.9998&rep=rep1&>

amp?type=pdf>. Acesso em: 10/10/2017.

HOGAN, W. R.; WAGNER, M. M. Accuracy of data in computer-based patient records. **Journal of the American Medical Informatics Association**, v. 4, n. 5, p. 342–355, 1997. Disponível em: <<https://academic.oup.com/jamia/article-lookup/doi/10.1136/jamia.1997.0040342>>. Acesso em: 20/9/2017.

HOUSEH, M. S.; ALDOSARI, B.; ALANAZI, A.; KUSHNIRUK, A. W.; BORYCKI, E. M. Big Data, big problems: a healthcare perspective. **Informatics Empowers Healthcare Transformation**. 1st ed., p.36–39, 2017. Amsterdam: Press, IOS. Disponível em: <<https://books.google.com.br/books?id=xHc2DwAAQBAJ&pg=PA79&lpg=PA79&dq=Informatics+Empowers+Healthcare+Transformation+pdf&source=bl&ots=zan73GJkGa&sig=1asvw32nznuSxY2bHSQCAYZjK1A&hl=pt-BR&sa=X&ved=0ahUKEwjvrYnroILXAhXDkJAKHZVwDY8Q6AEILzAB#v=onepage&q=Infor>>. .

HOWE, D.; COSTANZO, M.; FEY, P.; et al. Big data: the future of biocuration. **Nature**, v. 455, n. 7209, p. 47–50, 2010.

HUBER, P. J. Huge datasets. In: Verlag (Ed.); Compstat. **Anais...** . p.7, 1994. Heidelberg: Physica.

HUBER, P. J. Massive datasets workshop: four years after. **J Comput Graph Stat**, v. 83, p. 635–652, 1999.

KALIL FILHO, F. A. **Protocolo eletrônico multiprofissional de cirurgia do aparelho digestivo**, 2012. Tese (Doutorado em Clínica Cirúrgica) - Setor de Ciências da Saúde, Universidade Federal do Paraná, Curitiba.

KANG, N.; SINGH, B.; AFZAL, Z.; MULLIGEN, E. M. VAN; KORS, J. A. Using rule-based natural language processing to improve disease normalization in biomedical text. **Journal of the American Medical Informatics Association: JAMIA**, v. 20, n. 5, p. 876–81, 2013. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/23043124>>. Acesso em: 12/12/2013.

KRUSE, C. S.; GOSWAMY, R.; RAVAL, Y.; MARAWI, S. Challenges and opportunities of big data in health care: a systematic review. **JMIR Medical Informatics**, v. 4, n. 4, p. e38, 2016. Disponível em: <<http://medinform.jmir.org/2016/4/e38/>>. Acesso em: 13/11/2016.

KURETZKI, C. H. **Implementação de testes estatísticos para o sistema integrado de protocolos eletrônicos**, 2009. Tese (Doutorado em Clínica Cirúrgica) - Setor de Ciências da Saúde, Universidade Federal do Paraná, Curitiba.

KURETZKI, C. H.; CAMPOS, A. C. L.; MALAFAIA, O.; et al. Implementação e validação de testes estatísticos em software de pesquisa pra coleta e análise de dados em protocolos de cirurgia. **ABCD Arq Bras Cir Dig**, v. 29, n. 1, p. 14–16, 2016.

KUZIEMSKY, C. E.; MONKMAN, H.; PETERSEN, C.; et al. Big data in healthcare - defining the digital persona through user contexts from the micro to the macro. **IMIA**

Yearbook of Medical Informatics, v. 9, n. 1, p. 82–89, 2014. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/25123726>>. Acesso em: 1/3/2017.

LARSON, R. R. Information retrieval systems. In: M. J. Bates (Ed.); **Understanding information retrieval systems: management, types, and standards**. 1st ed., p.15–30, 2012. Boca Raton: CRC Press.

LEE, B.; RICHE, N. H.; KARLSON, A. K.; CARPENDALE, S. SparkClouds: visualizing trends in tag clouds. **IEEE Transactions on Visualization and Computer Graphics**, v. 16, n. 6, p. 1182–1189, 2010.

LEE, C. H.; YOON, H.-J. Medical big data: promise and challenges. **Kidney research and clinical practice**, v. 36, n. 1, p. 3–11, 2017. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/28392994>%0A<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5331970>>. Acesso em: 20/9/2017.

LEPENDU, P.; MUSEN, M. A.; SHAH, N. H. Enabling enrichment analysis with the human disease ontology. **J Biomed Inform.**, v. 44, n. 1, p. 118–125, 2011.

LOPES, I. L. Estratégia de busca na recuperação da informação: revisão da literatura. **Ciência da Informação**, v. 31, n. 2, p. 60–71, 2002. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652002000200007&lng=pt&tlng=pt>. Acesso em: 12/10/2017.

MACEDO, C. A. C. DE. **O uso do sistema integrado de protocolos eletrônicos Sinpe no estudo comparativo do tratamento fonoterápico após microcirurgia de laringe**, 2011. Dissertação (Mestrado em Clínica Cirúrgica) - Setor de Ciências da Saúde, Universidade Federal do Paraná, Curitiba. Disponível em: <<http://hdl.handle.net/1884/27121>>. .

MALAFAIA, O.; BORSATO, E.; PINTO, J. S. DE P. Manual do usuário do SINPE©. , 2003. Curitiba: UFPR.

MANIGLIA, F. F. **Avaliação de pacientes submetidos à dacriocistorrinostomia usando o Protocolo SINPE**, 2011. Dissertação (Mestrado em Clínica Cirúrgica) - Setor de Ciências da Saúde, Universidade Federal do Paraná, Curitiba.

MARAMBA, I. D.; DAVEY, A.; ELLIOTT, M. N.; et al. Web-based textual analysis of free-text patient experience comments from a survey in primary care. **JMIR Medical Informatics**, v. 3, n. 2, p. e20, 2015. Disponível em: <<http://medinform.jmir.org/2015/2/e20/>>. .

MARIN, H. D. F. Sistemas de informação em saúde: considerações gerais. **Journal of Health Informatics**, v. 2, n. 1, p. 20–24, 2010. Disponível em: <<http://www.jhi-sbis.saude.ws/ojs-jhi/index.php/jhi-sbis/article/view/4>>. Acesso em: 8/9/2017.

MARTHA, A. S.; JOSÉ, C.; CAMPOS, R. DE; SIGULEM, D. Recuperação de informações em campos de texto livres de prontuários eletrônicos do paciente baseada em semelhança semântica e ortográfica. **Journal of Health Informatics**, v. 2, n. 3, p. 63–71, 2010.

MCDONALD, C. J. The practice of informatics the barriers to electronic medical

record systems and how to overcome them. **Journal of the American Medical Informatics Association**, v. 4, n. 3, 1997.

MIKALSEN, K. Ø.; SOGUERO-RUIZ, C.; JENSEN, K.; et al. Using anchors from free text in electronic health records to diagnose postoperative delirium. **Computer Methods and Programs in Biomedicine**, v. 152, p. 105–114, 2017.

MILLER, R. A.; GEISSBUHLER, A. Clinical diagnostic decision support systems: an overview. In: E. S. Berner (Ed.); **Clinical decision support systems**. 1^a ed., p.3–34, 1999. New York: Springer.

MONNELLY, E. P.; LOCASTRO, J. S.; GAGNON, D.; YOUNG, M.; FIORE, L. D. Quetiapine versus trazodone in reducing rehospitalization for alcohol dependence: a large data-base study. **Journal of addiction medicine**, v. 2, n. 3, p. 128–34, 2008. Disponível em: <<http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=01271255-200809000-00003>>. Acesso em: 21/10/2017.

MOOERS, C. N. Information retrieval viewed as temporal signalling. Proceedings of International Conference of Mathematicians. **Anais...** . p.572–573, 1950. Cambridge: AMS: Providence.

MÜLLER, F.; DORMANN, H.; PFISTERMEISTER, B.; et al. Application of the Pareto principle to identify and address drug-therapy safety issues. **European Journal of Clinical Pharmacology**, v. 70, n. 6, p. 727–736, 2014. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/24652477>>. Acesso em: 15/11/2017.

NASSIF, D. S. B.; MALAFAIA, O.; NASSIF, P. A. N.; et al. Protocolo eletrônico multiprofissional em cirurgia bariátrica. **Arquivos brasileiros de cirurgia digestiva: ABCD Brazilian archives of digestive surgery**, v. 24, n. 3, p. 215–218, 2011.

NEWMAN, M. E. J. Power laws, Pareto distributions and Zipf's law. **Contemporary Physics**, v. 46, n. 5, p. 323–351, 2005.

NICOLAS, F. P.; POLLETTINI, J. T.; PANICO, S. G.; C, J.; MACEDO, A. A. O uso de UMLS para aprimorar a recomendação de graus de vigilância para pacientes do setor primário. Anais do X Workshop de Informática Médica. **Anais...** . p.1512–1521, 2010. Ribeiro Preto.

NLM - NATIONAL LIBRARY OF MEDICINE. The MeSH history. Disponível em: <https://www.nlm.nih.gov/mesh/intro_preface.html>. Acesso em: 23/9/2017.

ONG, K. Introduction to medical informatics. **Medical informatics an executive primer**. 3rd ed., p.xvii, 2015. Boca Raton: CRC Press. Disponível em: <<https://books.google.com.br/books?id=IRtjDAAAQBAJ&pg=PA97&lpg=PA97&dq=Informatics+Empowers+Healthcare+Transformation+pdf&source=bl&ots=3tQDIS9czY&sig=sSZRwOBV5hU9HVSAxiqFfIR0120&hl=pt-BR&sa=X&ved=0ahUKEwjvYnroILXAhXDkJAKHZVwDY8Q6AEIXTAI#v=onepage&q&f=fal>>. .

PATRÍCIO, C. M.; MAIA, M. M.; MACHIAVELLI, J. L.; NAVAES, M. DE A. O prontuário eletrônico do paciente no sistema de saúde brasileiro: uma realidade para

os médicos? **Scientia Medica**, v. 21, n. 3, p. 121–131, 2011.

PELLIZZON, R. D. F. Base de dados DeCS (Descritores em Ciências da Saúde). **Acta Cirurgica Brasileira**, v. 19, n. 2, p. 153–163, 2004.

PERES VANTI, N. A. Da bibliometria à webometria: uma exploração conceitual dos mecanismos utilizados para medir o registro da informação e a difusão do conhecimento. **Ciência da Informação**, v. 31, n. Bibliometria e webometria, p. 152–162, 2002. Disponível em: <<https://www.mendeley.com/viewer/?fileId=03055145-b4e7-db61-6fce-66727f754efa&documentId=95bffdc7-693e-3402-954c-573ba2bef>>. Acesso em: 1/10/2017.

PIANTADOSE, S. T. Zipf's word frequency law in natural language: a critical review and future directions. **Psychon Bull Rev.**, v. 21, n. 5, p. 1112–1130, 2014.

PINTO, J. S. DE P. **Interface de visibilização de informações para o Sistema Integrado de Protocolos Eletrônicos**, 2005. Tese (Doutorado em Clínica Cirúrgica) - Setor de Ciências da Saúde, Universidade Federal do Paraná, Curitiba.

REZENDE, D. A.; ABREU, A. F. DE. **Tecnologia da informação aplicada a sistemas de informação empresariais**. 9ª ed. São Paulo: Atlas, 2013.

RIBEIRO, C. G.; MOREIRA, A. T. R.; PINTO, J. S. D. P.; MALAFAIA, O. Multiprofissional electronic protocol in ophthalmology with emphasis in strabismus. **Revista do Colegio Brasileiro de Cirurgioes**, v. 43, n. 3, p. 171–7, 2016. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-69912016000300171&lng=en&nrm=iso&tlng=en%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/27556541>. Acesso em: 1/11/2017.

ROSIER, A.; BURGUN, A.; MABO, P. Using regular expressions to extract information on pacemaker implantation procedures from clinical reports. AMIA Annual Symposium proceedings reports. **Anais...** . p.81–85, 2008. AMIA Symposium.

RUMSFELD, J. S.; JOYNT, K. E.; MADDOX, T. M. Big data analytics to improve cardiovascular care: promise and challenges. **Nature Reviews Cardiology**, v. 13, n. 6, p. 350–359, 2016. Nature Publishing Group. Disponível em: <<http://www.nature.com/doi/10.1038/nrcardio.2016.42>>. Acesso em: 10/10/2017.

SCHIEFERDECKER, M. E. M.; KURETZKI, C. H.; CAMPOS, A. C. L.; et al. New electronic protocol for home enteral nutrition therapy. **Arquivos brasileiros de cirurgia digestiva: ABCD Brazilian archives of digestive surgery**, v. 26, n. 3, p. 195–9, 2013. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/24190377>>. Acesso em: 1/11/2017.

SHORTLIFFE, E. H. **The science of biomedical computing**. London: Med Inform, 1984.

SHORTLIFFE, E. H.; BARNETT, G. O. Medical Data: their acquisition, storage, and use. **Medical informatics: computer applications in health care and biomedicine**. 2ed. p.45–82, 2001.

SIGWALT, M. F. **Aplicação multicêntrica informatizada da coleta de dados na doença do refluxo gastroesofágico**, 2004. Tese (Doutorado em Clínica Cirúrgica) - Setor de Ciências da Saúde, Universidade Federal do Paraná, Curitiba.

SINHA, A.; HRIPCSAK, G.; MARKATOU, M. Large datasets in biomedicine: a discussion of salient analytic issues. **Journal of the American Medical Informatics Association**, v. 16, n. 6, p. 759–767, 2009. Elsevier Inc. Disponível em: <<http://dx.doi.org/10.1197/jamia.M2780>>. Acesso em: 1/9/2017.

SLOBOGEAN, G. P.; GIANNOUDIS, P. V.; FRIHAGEN, F.; et al. Bigger data, bigger problems. **Journal of orthopaedic trauma**, v. 29 Suppl 1, n. December, p. S43-6, 2015. Disponível em: <<http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00005131-201512001-00010%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/26584266>>. Acesso em: 15/10/2017.

SMITH, R. What clinical information do doctors need? **Bmj**, v. 313, n. 7064, p. 1062–1068, 1996. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/8898602%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2352351%5Cnhttp://www.bmj.com/cgi/content/abstract/313/7064/1062>>. Acesso em: 1/11/2017.

SNOWBALL LANGUAGE. Snowball algorithms. Disponível em: <<http://snowball.tartarus.org/algorithms/portuguese/stop.txt>>. Acesso em: 1/10/2017.

SOARES, J. M. L.; BORGES, R. DA R.; PINTO, V. B. Mapeamento e análise dos conceitos de prontuários do paciente nefropata visando a categorização e representação da informação. **Biblionline**, , n. Xxxiii, p. 81–91, 2010.

TIPLE, B.; DHANDE, S. Knowledge retrieval and data mining. **Journal of Eletronics, Communication & Soft Computing Science & Engineering**, , n. Special Issue, p. 54–58, 2012.

VANDENBUSSCHE, P.-Y.; CORMONT, S.; ANDRÉ, C.; et al. Implementation and management of a biomedical observation dictionary in a large healthcare information system. **Journal of the American Medical Informatics Association: JAMIA**, v. 20, n. 5, p. 940–6, 2013. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/23635601>>. Acesso em: 22/9/2017.

VANSTONE, M.; TOLEDO, F.; CLARKE, F.; et al. Narrative medicine and death in the ICU: word clouds as a visual legacy. **BMJ Supportive & Palliative Care**, , n. 0, p. 1–8, 2016. Disponível em: <<http://spcare.bmj.com/lookup/doi/10.1136/bmjspcare-2016-001179>>. Acesso em: 8/11/2016.

VARASCHIN, C.; LÚCIA, V.; LIMA, S. DE. **Fundamentos do processamento estatístico da linguagem natural**. Porto Alegre, 2001.

WANG, W.; KRISHNAN, E. Big data and clinicians: A review on the state of the science. **Journal of Medical Internet Research**, v. 16, n. 1, p. 1–11, 2014.

WILBUR, W. J.; SIROTKIN, K. The automatic identification of stop words. **Journal of Information Science**, v. 18, n. 1, p. 45–55, 1992.

WORLD HEALTH ORGANIZATION. **Health information systems development and strengthening: guidance on needs assessment for national health information development**. 2000.

WRIGHT, A.; BATES, D. W. Distribution of problems, medications and lab results in electronic health records: the Pareto Principle at work. **Applied Clinical Informatics**, v. 1, n. 1, p. 32–37, 2010. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3189502&tool=pmcentrez&rendertype=abstract>>. .

XIE, I. Information searching and search models. In: M. J. Bates (Ed.); **Understanding information retrieval systems: management, types, and standards**. 1st ed., p.31–46, 2012. Boca Raton: CRC Press.

YAO, Y.; ZENG, Y.; ZHONG, N.; HUANG, X. Knowledge retrieval (KR). Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, WI 2007. **Anais...** . p.729–735, 2007.

ZAHRA, F. M.; CARVALHO, D. R.; MALUCELLI, A. Poronto: ferramenta para construção semiautomática de ontologias em português. **Journal of Health Informatics**, v. 5, n. 2, p. 52–59, 2013.

ZHANG, Y.; HEMOND, M. S. C. Uncovering the predictive value of minimum blood glucose through statistical analysis of a large clinical dataset. **Symposium a Quarterly Journal in Modern Foreign Literatures**, p. 725–729, 2009.

GLOSSÁRIO

Clusterização – É uma técnica de mineração de dados, destinada ao agrupamento automático de dados, segundo seu grau de semelhança.

Corpus documental – Conjunto de dados linguísticos coletados como objeto de uma pesquisa linguística.

Fonetização – Área da linguística que estuda os sons (fones) das palavras.

Stemming – Processo de reduzir palavras flexionadas ao seu tronco (*stem*), base ou raiz.

Stop words – Palavras que podem ser consideradas irrelevantes para o conjunto de resultados a ser exibido, em uma busca realizada em textos livres.

Word cloud – Nuvem de palavras, também denominada *tag cloud*; é a representação visual, hierarquizada, para apresentar a frequência de palavras em um determinado *corpus* documental.

APÊNDICE A – TABELAS DE DISTRIBUIÇÃO POR DOENÇA

TABELA 4 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM RINITE ALÉRGICA (CID J30.4) N=17.179

Sintoma	n	% de prevalência estimada	Intervalo de confiança (95%)	Margem de erro
Espirro	3605	20,98%	21% (20,4% - 21,6%)	0,61%
Obstrução	3162	18,41%	18,4% (17,8% - 19%)	0,58%
Prurido	2917	16,98%	17% (16,4% - 17,5%)	0,56%
Tosse	2280	13,27%	13,3% (12,8% - 13,8%)	0,51%
Dor	1976	11,50%	11,5% (11% - 12%)	0,48%
Ronco	1693	9,86%	9,9% (9,4% - 10,3%)	0,45%
Prurido nasal	1452	8,45%	8,5% (8% - 8,9%)	0,42%
Febre	1419	8,26%	8,3% (7,8% - 8,7%)	0,41%
Congestão	1386	8,07%	8,1% (7,7% - 8,5%)	0,41%
Cefaleia	1143	6,65%	6,7% (6,3% - 7%)	0,37%
Nega febre	750	4,37%	4,4% (4,1% - 4,7%)	0,31%
Otalgia	478	2,78%	2,8% (2,5% - 3%)	0,25%
Sinusite	469	2,73%	2,7% (2,5% - 3%)	0,24%
Epistaxe	467	2,72%	2,7% (2,5% - 3%)	0,24%
Odinofagia	465	2,71%	2,7% (2,5% - 2,9%)	0,24%
Secreção	428	2,49%	2,5% (2,3% - 2,7%)	0,23%
Hipoacusia	392	2,28%	2,3% (2,1% - 2,5%)	0,22%
Disfonia	273	1,59%	1,6% (1,4% - 1,8%)	0,19%
Nega ronco	261	1,52%	1,5% (1,3% - 1,7%)	0,18%
Hialina	257	1,50%	1,5% (1,3% - 1,7%)	0,18%
RGE	219	1,27%	1,3% (1,1% - 1,4%)	0,17%
Palidez	218	1,27%	1,3% (1,1% - 1,4%)	0,17%
Pigarro	191	1,11%	1,1% (1% - 1,3%)	0,16%
Edema	175	1,02%	1% (0,9% - 1,2%)	0,15%
Nega dor	167	0,97%	1% (0,8% - 1,1%)	0,15%
Hiposmia	158	0,92%	0,9% (0,8% - 1,1%)	0,14%
Cornetos	150	0,87%	0,9% (0,7% - 1%)	0,14%
Otorreia	147	0,86%	0,9% (0,7% - 1%)	0,14%
Zumbido	135	0,79%	0,8% (0,7% - 0,9%)	0,13%
Pressão	130	0,76%	0,8% (0,6% - 0,9%)	0,13%
Nega tosse	126	0,73%	0,7% (0,6% - 0,9%)	0,13%
Hiperemia	111	0,65%	0,6% (0,5% - 0,8%)	0,12%
Trauma	100	0,58%	0,6% (0,5% - 0,7%)	0,11%
Sangramento	93	0,54%	0,5% (0,4% - 0,7%)	0,11%
Amarelada	92	0,54%	0,5% (0,4% - 0,6%)	0,11%
Dispneia	90	0,52%	0,5% (0,4% - 0,6%)	0,11%
Tontura	80	0,47%	0,5% (0,4% - 0,6%)	0,10%
Tabagismo	77	0,45%	0,4% (0,3% - 0,5%)	0,10%
Mialgia	76	0,44%	0,4% (0,3% - 0,5%)	0,10%
Crosta	66	0,38%	0,4% (0,3% - 0,5%)	0,09%
Cacosmia	55	0,32%	0,3% (0,2% - 0,4%)	0,08%
Gripe	54	0,31%	0,3% (0,2% - 0,4%)	0,08%
Azia	53	0,31%	0,3% (0,2% - 0,4%)	0,08%

FONTE: O autor (2017).

TABELA 5- PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM FARINGITE AGUDA (CID J02) N=13.527

Sintoma	n	% de prevalência estimada	Intervalo de confiança (95%)	Margem de erro
Febre	5314	39,28%	39,3% (38,5% - 40,1%)	0,82%
Dor	5197	38,42%	38,4% (37,6% - 39,2%)	0,82%
Tosse	4341	32,09%	32,1% (31,3% - 32,9%)	0,79%
Odinofagia	3419	25,28%	25,3% (24,5% - 26%)	0,73%
Nega febre	1627	12,03%	12% (11,5% - 12,6%)	0,55%
Disfonia	1494	11,04%	11% (10,5% - 11,6%)	0,53%
Cefaleia	1149	8,49%	8,5% (8% - 9%)	0,47%
Congestão	955	7,06%	7,1% (6,6% - 7,5%)	0,43%
Hiperemia	845	6,25%	6,2% (5,8% - 6,7%)	0,41%
Otalgia	814	6,02%	6% (5,6% - 6,4%)	0,40%
Mialgia	770	5,69%	5,7% (5,3% - 6,1%)	0,39%
Disfagia	729	5,39%	5,4% (5% - 5,8%)	0,38%
Secreção	572	4,23%	4,2% (3,9% - 4,6%)	0,34%
Obstrução	540	3,99%	4% (3,7% - 4,3%)	0,33%
Nega tosse	321	2,37%	2,4% (2,1% - 2,6%)	0,26%
Mal estar	313	2,31%	2,3% (2,1% - 2,6%)	0,25%
Prurido	306	2,26%	2,3% (2% - 2,5%)	0,25%
Espirro	262	1,94%	1,9% (1,7% - 2,2%)	0,23%
Náusea	258	1,91%	1,9% (1,7% - 2,1%)	0,23%
Pigarro	247	1,83%	1,8% (1,6% - 2,1%)	0,23%
Gripe	180	1,33%	1,3% (1,1% - 1,5%)	0,19%
Calafrios	150	1,11%	1,1% (0,9% - 1,3%)	0,18%
Outros	117	0,86%	0,9% (0,7% - 1%)	0,16%
Hipoacusia	116	0,86%	0,9% (0,7% - 1%)	0,16%
Dispneia	109	0,81%	0,8% (0,7% - 1%)	0,15%
Otorreia	102	0,75%	0,8% (0,6% - 0,9%)	0,15%
Edema	102	0,75%	0,8% (0,6% - 0,9%)	0,15%
Vômito	102	0,75%	0,8% (0,6% - 0,9%)	0,15%
Hiposmia	99	0,73%	0,7% (0,6% - 0,9%)	0,14%
Pirose	80	0,59%	0,6% (0,5% - 0,7%)	0,13%
RGE	69	0,51%	0,5% (0,4% - 0,6%)	0,12%
Ronco	67	0,50%	0,5% (0,4% - 0,6%)	0,12%
Placa	65	0,48%	0,5% (0,4% - 0,6%)	0,12%
Hialina	58	0,43%	0,4% (0,3% - 0,5%)	0,11%
Prurido nasal	48	0,35%	0,4% (0,3% - 0,5%)	0,10%
Nega disfonia	46	0,34%	0,3% (0,2% - 0,4%)	0,10%
Sudorese	42	0,31%	0,3% (0,2% - 0,4%)	0,09%
Purulenta	41	0,30%	0,3% (0,2% - 0,4%)	0,09%
nega Mialgia	38	0,28%	0,3% (0,2% - 0,4%)	0,09%
Tabagismo	36	0,27%	0,3% (0,2% - 0,4%)	0,09%
Tontura	34	0,25%	0,3% (0,2% - 0,3%)	0,08%
Amarelada	34	0,25%	0,3% (0,2% - 0,3%)	0,08%
Halitose	34	0,25%	0,3% (0,2% - 0,3%)	0,08%
Queimação	34	0,25%	0,3% (0,2% - 0,3%)	0,08%

FONTE: O autor (2017).

TABELA 6 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM CERUME IMPACTADO (CID H61.2) N=11.755

Sintoma	n	% de prevalência estimada	Intervalo de confiança (95%)	Margem de erro
Hipoacusia	2381	20,26%	20,3% (19,5% - 21%)	0,73%
Otalgia	1364	11,60%	11,6% (11% - 12,2%)	0,58%
Otorreia	548	4,66%	4,7% (4,3% - 5%)	0,38%
Dor	421	3,58%	3,6% (3,2% - 3,9%)	0,34%
Zumbido	398	3,39%	3,4% (3,1% - 3,7%)	0,33%
Febre	326	2,77%	2,8% (2,5% - 3,1%)	0,30%
Prurido	251	2,14%	2,1% (1,9% - 2,4%)	0,26%
Nega otalgia	231	1,97%	2% (1,7% - 2,2%)	0,25%
Nega otorreia	201	1,71%	1,7% (1,5% - 1,9%)	0,23%
Tosse	184	1,57%	1,6% (1,3% - 1,8%)	0,22%
Nega febre	183	1,56%	1,6% (1,3% - 1,8%)	0,22%
Tontura	139	1,18%	1,2% (1% - 1,4%)	0,20%
Obstrução	130	1,11%	1,1% (0,9% - 1,3%)	0,19%
Ronco	102	0,87%	0,9% (0,7% - 1%)	0,17%
Odinofagia	77	0,66%	0,7% (0,5% - 0,8%)	0,15%
Cefaleia	67	0,57%	0,6% (0,4% - 0,7%)	0,14%
Nega hipoacusia	66	0,56%	0,6% (0,4% - 0,7%)	0,14%
Congestão	57	0,48%	0,5% (0,4% - 0,6%)	0,13%
Vertigem	55	0,47%	0,5% (0,3% - 0,6%)	0,12%
Hiperemia	51	0,43%	0,4% (0,3% - 0,6%)	0,12%
Pressão	49	0,42%	0,4% (0,3% - 0,5%)	0,12%
Espirro	47	0,40%	0,4% (0,3% - 0,5%)	0,11%
Nega dor	47	0,40%	0,4% (0,3% - 0,5%)	0,11%
Secreção	42	0,36%	0,4% (0,2% - 0,5%)	0,11%
Pigarro	37	0,31%	0,3% (0,2% - 0,4%)	0,10%
Disfonia	36	0,31%	0,3% (0,2% - 0,4%)	0,10%
Sensação de plenitude auricular	35	0,30%	0,3% (0,2% - 0,4%)	0,10%
Edema	33	0,28%	0,3% (0,2% - 0,4%)	0,10%
Perda auditiva	31	0,26%	0,3% (0,2% - 0,4%)	0,09%
Trauma	28	0,24%	0,2% (0,2% - 0,3%)	0,09%
Corpo estranho	24	0,20%	0,2% (0,1% - 0,3%)	0,08%
Aspiração	22	0,19%	0,2% (0,1% - 0,3%)	0,08%
Náusea	21	0,18%	0,2% (0,1% - 0,3%)	0,08%
Sinusite	21	0,18%	0,2% (0,1% - 0,3%)	0,08%
Prurido nasal	20	0,17%	0,2% (0,1% - 0,2%)	0,07%
Descamação	20	0,17%	0,2% (0,1% - 0,2%)	0,07%
Nega tontura	19	0,16%	0,2% (0,1% - 0,2%)	0,07%
Gripe	18	0,15%	0,2% (0,1% - 0,2%)	0,07%
Otorragia	18	0,15%	0,2% (0,1% - 0,2%)	0,07%
RGE	17	0,14%	0,1% (0,1% - 0,2%)	0,07%
Nega tosse	16	0,14%	0,1% (0,1% - 0,2%)	0,07%
Nega ronco	15	0,13%	0,1% (0,1% - 0,2%)	0,06%
Crosta	14	0,12%	0,1% (0,1% - 0,2%)	0,06%
Epistaxe	13	0,11%	0,1% (0,1% - 0,2%)	0,06%

FONTE: O autor (2017).

TABELA 7 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM INFECÇÃO AGUDA DAS VIAS AÉREAS SUPERIORES NÃO ESPECIFICADA (CID J06.9) N=11.233

Sintoma	n	% de prevalência estimada	Intervalo de confiança (95%)	Margem de erro
Tosse	4849	43,17%	43,2% (42,3% - 44,1%)	0,92%
Febre	4109	36,58%	36,6% (35,7% - 37,5%)	0,89%
Dor	2398	21,35%	21,3% (20,6% - 22,1%)	0,76%
Congestão	1999	17,80%	17,8% (17,1% - 18,5%)	0,71%
Nega febre	1840	16,38%	16,4% (15,7% - 17,1%)	0,68%
Odinofagia	1648	14,67%	14,7% (14% - 15,3%)	0,65%
Disfonia	1187	10,57%	10,6% (10% - 11,1%)	0,57%
Cefaleia	1149	10,23%	10,2% (9,7% - 10,8%)	0,56%
Obstrução	810	7,21%	7,2% (6,7% - 7,7%)	0,48%
Otalgia	636	5,66%	5,7% (5,2% - 6,1%)	0,43%
Hiperemia	558	4,97%	5% (4,6% - 5,4%)	0,40%
Secreção	524	4,66%	4,7% (4,3% - 5,1%)	0,39%
Espirro	488	4,34%	4,3% (4% - 4,7%)	0,38%
Mialgia	276	2,46%	2,5% (2,2% - 2,7%)	0,29%
Mal estar	256	2,28%	2,3% (2% - 2,6%)	0,28%
Prurido	246	2,19%	2,2% (1,9% - 2,5%)	0,27%
Gripe	226	2,01%	2% (1,8% - 2,3%)	0,26%
Hialina	196	1,74%	1,7% (1,5% - 2%)	0,24%
Nega tosse	192	1,71%	1,7% (1,5% - 1,9%)	0,24%
Dispneia	178	1,58%	1,6% (1,4% - 1,8%)	0,23%
Disfagia	175	1,56%	1,6% (1,3% - 1,8%)	0,23%
Vômito	129	1,15%	1,1% (1% - 1,3%)	0,20%
Ronco	120	1,07%	1,1% (0,9% - 1,3%)	0,19%
Amarelada	108	0,96%	1% (0,8% - 1,1%)	0,18%
Hipoacusia	88	0,78%	0,8% (0,6% - 0,9%)	0,16%
Otorreia	88	0,78%	0,8% (0,6% - 0,9%)	0,16%
Pressão	72	0,64%	0,6% (0,5% - 0,8%)	0,15%
Calafrios	61	0,54%	0,5% (0,4% - 0,7%)	0,14%
Sinusite	59	0,53%	0,5% (0,4% - 0,7%)	0,13%
Pigarro	58	0,52%	0,5% (0,4% - 0,6%)	0,13%
Edema	57	0,51%	0,5% (0,4% - 0,6%)	0,13%
Purulenta	51	0,45%	0,5% (0,3% - 0,6%)	0,12%
Nega dor	49	0,44%	0,4% (0,3% - 0,6%)	0,12%
Nega disfonia	46	0,41%	0,4% (0,3% - 0,5%)	0,12%
Epistaxe	45	0,40%	0,4% (0,3% - 0,5%)	0,12%
Palidez	41	0,36%	0,4% (0,3% - 0,5%)	0,11%
Náusea	39	0,35%	0,3% (0,2% - 0,5%)	0,11%
Tontura	38	0,34%	0,3% (0,2% - 0,4%)	0,11%
Mucoide	38	0,34%	0,3% (0,2% - 0,4%)	0,11%
Zumbido	33	0,29%	0,3% (0,2% - 0,4%)	0,10%
Prurido nasal	32	0,28%	0,3% (0,2% - 0,4%)	0,10%
Outros	30	0,27%	0,3% (0,2% - 0,4%)	0,10%
Tabagismo	28	0,25%	0,2% (0,2% - 0,3%)	0,09%
Dor torácica	27	0,24%	0,2% (0,1% - 0,3%)	0,09%

FONTE: O autor (2017).

TABELA 8 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM AMIGDALITE AGUDA NÃO ESPECIFICADA (CID J03.9) N=11.062

Sintoma	n	% de prevalência estimada	Intervalo de confiança (95%)	Margem de erro
Febre	5696	51,49%	51,5% (50,6% - 52,4%)	0,93%
Odinofagia	4025	36,39%	36,4% (35,5% - 37,3%)	0,90%
Dor	3664	33,12%	33,1% (32,2% - 34%)	0,88%
Tosse	1545	13,97%	14% (13,3% - 14,6%)	0,65%
Nega febre	772	6,98%	7% (6,5% - 7,5%)	0,47%
Placa	702	6,35%	6,3% (5,9% - 6,8%)	0,45%
Cefaleia	649	5,87%	5,9% (5,4% - 6,3%)	0,44%
Otalgia	647	5,85%	5,8% (5,4% - 6,3%)	0,44%
Disfonia	507	4,58%	4,6% (4,2% - 5%)	0,39%
Mialgia	395	3,57%	3,6% (3,2% - 3,9%)	0,35%
Congestão	366	3,31%	3,3% (3% - 3,6%)	0,33%
Disfagia	339	3,06%	3,1% (2,7% - 3,4%)	0,32%
Obstrução	323	2,92%	2,9% (2,6% - 3,2%)	0,31%
Nega tosse	269	2,43%	2,4% (2,1% - 2,7%)	0,29%
Ronco	219	1,98%	2% (1,7% - 2,2%)	0,26%
Mal estar	205	1,85%	1,9% (1,6% - 2,1%)	0,25%
Hiperemia	165	1,49%	1,5% (1,3% - 1,7%)	0,23%
Secreção	147	1,33%	1,3% (1,1% - 1,5%)	0,21%
Vômito	126	1,14%	1,1% (0,9% - 1,3%)	0,20%
Espirro	108	0,98%	1% (0,8% - 1,2%)	0,18%
Náusea	97	0,88%	0,9% (0,7% - 1,1%)	0,17%
Otorreia	92	0,83%	0,8% (0,7% - 1%)	0,17%
Calafrios	88	0,80%	0,8% (0,6% - 1%)	0,17%
Exsudato	87	0,79%	0,8% (0,6% - 1%)	0,16%
Prurido	84	0,76%	0,8% (0,6% - 0,9%)	0,16%
Edema	69	0,62%	0,6% (0,5% - 0,8%)	0,15%
Caseum	66	0,60%	0,6% (0,5% - 0,7%)	0,14%
Gripe	61	0,55%	0,6% (0,4% - 0,7%)	0,14%
Hipoacusia	57	0,52%	0,5% (0,4% - 0,6%)	0,13%
Outros	56	0,51%	0,5% (0,4% - 0,6%)	0,13%
Halitose	55	0,50%	0,5% (0,4% - 0,6%)	0,13%
Purulenta	54	0,49%	0,5% (0,4% - 0,6%)	0,13%
Dispneia	41	0,37%	0,4% (0,3% - 0,5%)	0,11%
Infecção	37	0,33%	0,3% (0,2% - 0,4%)	0,11%
Hialina	36	0,33%	0,3% (0,2% - 0,4%)	0,11%
Amarelada	34	0,31%	0,3% (0,2% - 0,4%)	0,10%
Caseosa	33	0,30%	0,3% (0,2% - 0,4%)	0,10%
Nega disfonia	31	0,28%	0,3% (0,2% - 0,4%)	0,10%
Sinusite	28	0,25%	0,3% (0,2% - 0,3%)	0,09%
Nega ronco	25	0,23%	0,2% (0,1% - 0,3%)	0,09%
Pigarro	23	0,21%	0,2% (0,1% - 0,3%)	0,08%
Aftas	22	0,20%	0,2% (0,1% - 0,3%)	0,08%
Nega mialgia	20	0,18%	0,2% (0,1% - 0,3%)	0,08%
Tabagismo	19	0,17%	0,2% (0,1% - 0,2%)	0,08%

FONTE: O autor (2017).

TABELA 9 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM SINUSITE MAXILAR AGUDA (CID J01.0) N=7.616

Sintoma	n	% de prevalência estimada	Intervalo de confiança (95%)	Margem de erro
Dor	2372	31,14%	31,1% (30,1% - 32,2%)	1,04%
Tosse	2277	29,90%	29,9% (28,9% - 30,9%)	1,03%
Febre	1879	24,67%	24,7% (23,7% - 25,6%)	0,97%
Cefaleia	1694	22,24%	22,2% (21,3% - 23,2%)	0,93%
Congestão	1142	14,99%	15% (14,2% - 15,8%)	0,80%
Obstrução	993	13,04%	13% (12,3% - 13,8%)	0,76%
Sinusite	910	11,95%	11,9% (11,2% - 12,7%)	0,73%
Secreção	855	11,23%	11,2% (10,5% - 11,9%)	0,71%
Nega febre	844	11,08%	11,1% (10,4% - 11,8%)	0,71%
Odinofagia	511	6,71%	6,7% (6,1% - 7,3%)	0,56%
Otalgia	340	4,46%	4,5% (4% - 4,9%)	0,46%
Amarelada	328	4,31%	4,3% (3,9% - 4,8%)	0,46%
Purulenta	255	3,35%	3,3% (2,9% - 3,8%)	0,40%
Espirro	247	3,24%	3,2% (2,8% - 3,6%)	0,40%
Disfonia	235	3,09%	3,1% (2,7% - 3,5%)	0,39%
Prurido	192	2,52%	2,5% (2,2% - 2,9%)	0,35%
Gripe	140	1,84%	1,8% (1,5% - 2,1%)	0,30%
Hiperemia	139	1,83%	1,8% (1,5% - 2,1%)	0,30%
Hiposmia	121	1,59%	1,6% (1,3% - 1,9%)	0,28%
Ronco	117	1,54%	1,5% (1,3% - 1,8%)	0,28%
Pressão	110	1,44%	1,4% (1,2% - 1,7%)	0,27%
Disfagia	94	1,23%	1,2% (1% - 1,5%)	0,25%
Nega tosse	87	1,14%	1,1% (0,9% - 1,4%)	0,24%
Mialgia	73	0,96%	1% (0,7% - 1,2%)	0,22%
Hipoacusia	69	0,91%	0,9% (0,7% - 1,1%)	0,21%
Cacosmia	64	0,84%	0,8% (0,6% - 1%)	0,21%
Mal estar	61	0,80%	0,8% (0,6% - 1%)	0,20%
Edema	58	0,76%	0,8% (0,6% - 1%)	0,20%
Epistaxe	53	0,70%	0,7% (0,5% - 0,9%)	0,19%
Prurido nasal	51	0,67%	0,7% (0,5% - 0,9%)	0,18%
Otorreia	46	0,60%	0,6% (0,4% - 0,8%)	0,17%
Hialina	43	0,56%	0,6% (0,4% - 0,7%)	0,17%
Pigarro	43	0,56%	0,6% (0,4% - 0,7%)	0,17%
Nega dor	43	0,56%	0,6% (0,4% - 0,7%)	0,17%
Fadiga	37	0,49%	0,5% (0,3% - 0,6%)	0,16%
Dispneia	32	0,42%	0,4% (0,3% - 0,6%)	0,15%
Tabagismo	32	0,42%	0,4% (0,3% - 0,6%)	0,15%
Tontura	31	0,41%	0,4% (0,3% - 0,6%)	0,14%
Vômito	29	0,38%	0,4% (0,2% - 0,5%)	0,14%
Anosmia	27	0,35%	0,4% (0,2% - 0,5%)	0,13%
Nega obstrução	26	0,34%	0,3% (0,2% - 0,5%)	0,13%
Nega tabagismo	23	0,30%	0,3% (0,2% - 0,4%)	0,12%
Náusea	22	0,29%	0,3% (0,2% - 0,4%)	0,12%
Nega cefaleia	22	0,29%	0,3% (0,2% - 0,4%)	0,12%

FONTE: O autor (2017).

TABELA 10 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM NASOFARINGITE AGUDA (CID J00) N=7.432

Sintoma	n	% de prevalência estimada	Intervalo de confiança (95%)	Margem de erro
Tosse	3606	48,52%	48,5% (47,4% - 49,7%)	1,14%
Febre	3434	46,21%	46,2% (45,1% - 47,3%)	1,13%
Dor	2912	39,18%	39,2% (38,1% - 40,3%)	1,11%
Congestão	1695	22,81%	22,8% (21,9% - 23,8%)	0,95%
Odinofagia	1400	18,84%	18,8% (17,9% - 19,7%)	0,89%
Nega febre	1119	15,06%	15,1% (14,2% - 15,9%)	0,81%
Cefaleia	1032	13,89%	13,9% (13,1% - 14,7%)	0,79%
Secreção	627	8,44%	8,4% (7,8% - 9,1%)	0,63%
Obstrução	528	7,10%	7,1% (6,5% - 7,7%)	0,58%
Otalgia	526	7,08%	7,1% (6,5% - 7,7%)	0,58%
Disfonia	500	6,73%	6,7% (6,2% - 7,3%)	0,57%
Espirro	431	5,80%	5,8% (5,3% - 6,3%)	0,53%
Mialgia	295	3,97%	4% (3,5% - 4,4%)	0,44%
Mal estar	256	3,44%	3,4% (3% - 3,9%)	0,41%
Prurido	249	3,35%	3,4% (2,9% - 3,8%)	0,41%
Gripe	194	2,61%	2,6% (2,2% - 3%)	0,36%
Nega tosse	129	1,74%	1,7% (1,4% - 2%)	0,30%
Hialina	127	1,71%	1,7% (1,4% - 2%)	0,29%
Dispneia	114	1,53%	1,5% (1,3% - 1,8%)	0,28%
Hiperemia	98	1,32%	1,3% (1,1% - 1,6%)	0,26%
Ronco	86	1,16%	1,2% (0,9% - 1,4%)	0,24%
Otorreia	77	1,04%	1% (0,8% - 1,3%)	0,23%
Hipoacusia	74	1,00%	1% (0,8% - 1,2%)	0,23%
Vômito	68	0,91%	0,9% (0,7% - 1,1%)	0,22%
Prurido nasal	54	0,73%	0,7% (0,5% - 0,9%)	0,19%
Náusea	54	0,73%	0,7% (0,5% - 0,9%)	0,19%
Sinusite	53	0,71%	0,7% (0,5% - 0,9%)	0,19%
Purulenta	47	0,63%	0,6% (0,5% - 0,8%)	0,18%
Hiposmia	46	0,62%	0,6% (0,4% - 0,8%)	0,18%
Pigarro	45	0,61%	0,6% (0,4% - 0,8%)	0,18%
Amarelada	44	0,59%	0,6% (0,4% - 0,8%)	0,17%
Pressão	42	0,57%	0,6% (0,4% - 0,7%)	0,17%
Epistaxe	36	0,48%	0,5% (0,3% - 0,6%)	0,16%
Disfagia	34	0,46%	0,5% (0,3% - 0,6%)	0,15%
Nega dor	32	0,43%	0,4% (0,3% - 0,6%)	0,15%
Calafrios	32	0,43%	0,4% (0,3% - 0,6%)	0,15%
Nega dispneia	28	0,38%	0,4% (0,2% - 0,5%)	0,14%
Edema	24	0,32%	0,3% (0,2% - 0,5%)	0,13%
Zumbido	22	0,30%	0,3% (0,2% - 0,4%)	0,12%
Tontura	19	0,26%	0,3% (0,1% - 0,4%)	0,11%
RGE	19	0,26%	0,3% (0,1% - 0,4%)	0,11%
Nega otorreia	17	0,23%	0,2% (0,1% - 0,3%)	0,11%
Vertigem	16	0,22%	0,2% (0,1% - 0,3%)	0,11%
Dor torácica	16	0,22%	0,2% (0,1% - 0,3%)	0,11%

FONTE: O autor (2017).

TABELA 11 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM TOSSE (CID R05) N=6.615

Sintoma	n	% de prevalência estimada	Intervalo de confiança (95%)	Margem de erro
Tosse	6369	96,28%	96,3% (95,8% - 96,7%)	0,46%
Febre	1907	28,83%	28,8% (27,7% - 29,9%)	1,09%
Dor	1053	15,92%	15,9% (15% - 16,8%)	0,88%
Nega febre	924	13,97%	14% (13,1% - 14,8%)	0,84%
Disfonia	685	10,36%	10,4% (9,6% - 11,1%)	0,73%
Odinofagia	463	7,00%	7% (6,4% - 7,6%)	0,61%
Secreção	401	6,06%	6,1% (5,5% - 6,6%)	0,58%
Dispneia	359	5,43%	5,4% (4,9% - 6%)	0,55%
Congestão	349	5,28%	5,3% (4,7% - 5,8%)	0,54%
Cefaleia	301	4,55%	4,6% (4% - 5,1%)	0,50%
Hiperemia	298	4,50%	4,5% (4% - 5%)	0,50%
Obstrução	290	4,38%	4,4% (3,9% - 4,9%)	0,49%
Prurido	267	4,04%	4% (3,6% - 4,5%)	0,47%
Pigarro	247	3,73%	3,7% (3,3% - 4,2%)	0,46%
Espirro	163	2,46%	2,5% (2,1% - 2,8%)	0,37%
Ronco	148	2,24%	2,2% (1,9% - 2,6%)	0,36%
Tabagismo	135	2,04%	2% (1,7% - 2,4%)	0,34%
Otalgia	112	1,69%	1,7% (1,4% - 2%)	0,31%
Pirose	101	1,53%	1,5% (1,2% - 1,8%)	0,30%
Nega tabagismo	99	1,50%	1,5% (1,2% - 1,8%)	0,29%
Mal estar	92	1,39%	1,4% (1,1% - 1,7%)	0,28%
Mialgia	83	1,25%	1,3% (1% - 1,5%)	0,27%
Nega dor	79	1,19%	1,2% (0,9% - 1,5%)	0,26%
RGE	78	1,18%	1,2% (0,9% - 1,4%)	0,26%
Azia	74	1,12%	1,1% (0,9% - 1,4%)	0,25%
Gripe	73	1,10%	1,1% (0,9% - 1,4%)	0,25%
Hialina	73	1,10%	1,1% (0,9% - 1,4%)	0,25%
Sinusite	70	1,06%	1,1% (0,8% - 1,3%)	0,25%
Amarelada	68	1,03%	1% (0,8% - 1,3%)	0,24%
Dor torácica	60	0,91%	0,9% (0,7% - 1,1%)	0,23%
Nega disfonia	54	0,82%	0,8% (0,6% - 1%)	0,22%
Outros	50	0,76%	0,8% (0,5% - 1%)	0,21%
Vômito	48	0,73%	0,7% (0,5% - 0,9%)	0,20%
Purulenta	47	0,71%	0,7% (0,5% - 0,9%)	0,20%
Nega dispneia	42	0,63%	0,6% (0,4% - 0,8%)	0,19%
Hipoacusia	40	0,60%	0,6% (0,4% - 0,8%)	0,19%
Prurido nasal	39	0,59%	0,6% (0,4% - 0,8%)	0,18%
nega odinofagia	39	0,59%	0,6% (0,4% - 0,8%)	0,18%
Náusea	34	0,51%	0,5% (0,3% - 0,7%)	0,17%
Nega Obstrução	28	0,42%	0,4% (0,3% - 0,6%)	0,16%
Edema	27	0,41%	0,4% (0,3% - 0,6%)	0,15%
Palidez	26	0,39%	0,4% (0,2% - 0,5%)	0,15%
Nega ronco	25	0,38%	0,4% (0,2% - 0,5%)	0,15%
Nega cefaleia	22	0,33%	0,3% (0,2% - 0,5%)	0,14%

FONTE: O autor (2017).

TABELA 12 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM DESVIO DE SEPTO NASAL (CID J34.2) N=4.915

Sintoma	n	% de prevalência estimada	Intervalo de confiança (95%)	Margem de erro
Obstrução	1469	29,89%	29,9% (28,6% - 31,2%)	1,28%
Dor	739	15,04%	15% (14% - 16%)	1,00%
Ronco	523	10,64%	10,6% (9,8% - 11,5%)	0,86%
Cefaleia	382	7,77%	7,8% (7% - 8,5%)	0,75%
Prurido	226	4,60%	4,6% (4% - 5,2%)	0,59%
Espirro	205	4,17%	4,2% (3,6% - 4,7%)	0,56%
Tosse	195	3,97%	4% (3,4% - 4,5%)	0,55%
Congestão	192	3,91%	3,9% (3,4% - 4,4%)	0,54%
Trauma	163	3,32%	3,3% (2,8% - 3,8%)	0,50%
Sinusite	156	3,17%	3,2% (2,7% - 3,7%)	0,49%
Febre	150	3,05%	3,1% (2,6% - 3,5%)	0,48%
Cornetos	122	2,48%	2,5% (2% - 2,9%)	0,43%
Secreção	103	2,10%	2,1% (1,7% - 2,5%)	0,40%
Epistaxe	91	1,85%	1,9% (1,5% - 2,2%)	0,38%
Hipoacusia	90	1,83%	1,8% (1,5% - 2,2%)	0,37%
Prurido nasal	83	1,69%	1,7% (1,3% - 2%)	0,36%
Edema	81	1,65%	1,6% (1,3% - 2%)	0,36%
Otalgia	80	1,63%	1,6% (1,3% - 2%)	0,35%
Odinofagia	74	1,51%	1,5% (1,2% - 1,8%)	0,34%
Nega febre	69	1,40%	1,4% (1,1% - 1,7%)	0,33%
Palidez	66	1,34%	1,3% (1% - 1,7%)	0,32%
Pigarro	56	1,14%	1,1% (0,8% - 1,4%)	0,30%
Tabagismo	56	1,14%	1,1% (0,8% - 1,4%)	0,30%
Hiposmia	54	1,10%	1,1% (0,8% - 1,4%)	0,29%
RGE	52	1,06%	1,1% (0,8% - 1,3%)	0,29%
Hiperemia	51	1,04%	1% (0,8% - 1,3%)	0,28%
Nega dor	42	0,85%	0,9% (0,6% - 1,1%)	0,26%
Disfonia	38	0,77%	0,8% (0,5% - 1%)	0,24%
Pressão	38	0,77%	0,8% (0,5% - 1%)	0,24%
Desvio septal	33	0,67%	0,7% (0,4% - 0,9%)	0,23%
Crosta	32	0,65%	0,7% (0,4% - 0,9%)	0,22%
Sangramento	32	0,65%	0,7% (0,4% - 0,9%)	0,22%
Sinequia	32	0,65%	0,7% (0,4% - 0,9%)	0,22%
Nega tabagismo	29	0,59%	0,6% (0,4% - 0,8%)	0,21%
Otorreia	28	0,57%	0,6% (0,4% - 0,8%)	0,21%
Zumbido	28	0,57%	0,6% (0,4% - 0,8%)	0,21%
Crostas	27	0,55%	0,5% (0,3% - 0,8%)	0,21%
Fadiga	24	0,49%	0,5% (0,3% - 0,7%)	0,19%
Válvula	24	0,49%	0,5% (0,3% - 0,7%)	0,19%
Caseum	22	0,45%	0,4% (0,3% - 0,6%)	0,19%
Nega cefaleia	21	0,43%	0,4% (0,2% - 0,6%)	0,18%
Tontura	21	0,43%	0,4% (0,2% - 0,6%)	0,18%
Gripe	19	0,39%	0,4% (0,2% - 0,6%)	0,17%
Ansiedade	17	0,35%	0,3% (0,2% - 0,5%)	0,16%

FONTE: O autor (2017).

TABELA 13 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM EXAME MÉDICO GERAL (CID Z00.0) N=4.502

Sintoma	n	% de prevalência estimada	Intervalo de confiança (95%)	Margem de erro
Dor	788	17,50%	17,5% (16,4% - 18,6%)	1,11%
Febre	277	6,15%	6,2% (5,5% - 6,9%)	0,70%
Tosse	194	4,31%	4,3% (3,7% - 4,9%)	0,59%
Ronco	171	3,80%	3,8% (3,2% - 4,4%)	0,56%
Otalgia	124	2,75%	2,8% (2,3% - 3,2%)	0,48%
Hipoacusia	122	2,71%	2,7% (2,2% - 3,2%)	0,47%
Nega febre	115	2,55%	2,6% (2,1% - 3%)	0,46%
Prurido	106	2,35%	2,4% (1,9% - 2,8%)	0,44%
Edema	105	2,33%	2,3% (1,9% - 2,8%)	0,44%
Obstrução	86	1,91%	1,9% (1,5% - 2,3%)	0,40%
Otorreia	83	1,84%	1,8% (1,5% - 2,2%)	0,39%
RGE	79	1,75%	1,8% (1,4% - 2,1%)	0,38%
Pressão	76	1,69%	1,7% (1,3% - 2,1%)	0,38%
Tabagismo	74	1,64%	1,6% (1,3% - 2%)	0,37%
Cefaleia	68	1,51%	1,5% (1,2% - 1,9%)	0,36%
Azia	65	1,44%	1,4% (1,1% - 1,8%)	0,35%
Trauma	64	1,42%	1,4% (1,1% - 1,8%)	0,35%
Disfonia	63	1,40%	1,4% (1,1% - 1,7%)	0,34%
Secreção	62	1,38%	1,4% (1% - 1,7%)	0,34%
Corpo estranho	54	1,20%	1,2% (0,9% - 1,5%)	0,32%
Nega ronco	52	1,16%	1,2% (0,8% - 1,5%)	0,31%
Vômito	51	1,13%	1,1% (0,8% - 1,4%)	0,31%
Sangramento	51	1,13%	1,1% (0,8% - 1,4%)	0,31%
Odinofagia	46	1,02%	1% (0,7% - 1,3%)	0,29%
Pigarro	46	1,02%	1% (0,7% - 1,3%)	0,29%
Sinusite	44	0,98%	1% (0,7% - 1,3%)	0,29%
Espirro	43	0,96%	1% (0,7% - 1,2%)	0,28%
Tontura	43	0,96%	1% (0,7% - 1,2%)	0,28%
Zumbido	38	0,84%	0,8% (0,6% - 1,1%)	0,27%
Congestão	35	0,78%	0,8% (0,5% - 1%)	0,26%
Gestação	34	0,76%	0,8% (0,5% - 1%)	0,25%
Nega hipoacusia	34	0,76%	0,8% (0,5% - 1%)	0,25%
Infecção	33	0,73%	0,7% (0,5% - 1%)	0,25%
Nega dor	32	0,71%	0,7% (0,5% - 1%)	0,25%
Gripe	30	0,67%	0,7% (0,4% - 0,9%)	0,24%
Nódulo	28	0,62%	0,6% (0,4% - 0,9%)	0,23%
Queimação	26	0,58%	0,6% (0,4% - 0,8%)	0,22%
Prurido nasal	25	0,56%	0,6% (0,3% - 0,8%)	0,22%
Caseum	25	0,56%	0,6% (0,3% - 0,8%)	0,22%
Dispneia	24	0,53%	0,5% (0,3% - 0,7%)	0,21%
Náusea	20	0,44%	0,4% (0,2% - 0,6%)	0,19%
Ansiedade	20	0,44%	0,4% (0,2% - 0,6%)	0,19%
Nega tabagismo	19	0,42%	0,4% (0,2% - 0,6%)	0,19%
Epistaxe	19	0,42%	0,4% (0,2% - 0,6%)	0,19%

FONTE: O autor (2017).

TABELA 14 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO LARINGITE AGUDA (CID J04.0) N=4.321

Sintoma	n	% de prevalência estimada	Intervalo de confiança (95%)	Margem de erro
Tosse	2279	52,74%	52,7% (51,3% - 54,2%)	1,49%
Disfonia	2201	50,94%	50,9% (49,4% - 52,4%)	1,49%
Febre	1130	26,15%	26,2% (24,8% - 27,5%)	1,31%
Dor	936	21,66%	21,7% (20,4% - 22,9%)	1,23%
Odinofagia	661	15,30%	15,3% (14,2% - 16,4%)	1,07%
Nega febre	602	13,93%	13,9% (12,9% - 15%)	1,03%
Cefaleia	179	4,14%	4,1% (3,5% - 4,7%)	0,59%
Secreção	166	3,84%	3,8% (3,3% - 4,4%)	0,57%
Hiperemia	149	3,45%	3,4% (2,9% - 4%)	0,54%
Congestão	145	3,36%	3,4% (2,8% - 3,9%)	0,54%
Obstrução	100	2,31%	2,3% (1,9% - 2,8%)	0,45%
Prurido	94	2,18%	2,2% (1,7% - 2,6%)	0,43%
Mialgia	86	1,99%	2% (1,6% - 2,4%)	0,42%
Otalgia	85	1,97%	2% (1,6% - 2,4%)	0,41%
Disfagia	80	1,85%	1,9% (1,4% - 2,3%)	0,40%
Pigarro	70	1,62%	1,6% (1,2% - 2%)	0,38%
Espirro	66	1,53%	1,5% (1,2% - 1,9%)	0,37%
Mal estar	66	1,53%	1,5% (1,2% - 1,9%)	0,37%
Afonia	64	1,48%	1,5% (1,1% - 1,8%)	0,36%
Tabagismo	57	1,32%	1,3% (1% - 1,7%)	0,34%
RGE	54	1,25%	1,2% (0,9% - 1,6%)	0,33%
Dispneia	53	1,23%	1,2% (0,9% - 1,6%)	0,33%
Nega tabagismo	43	1,00%	1% (0,7% - 1,3%)	0,30%
Nega dor	39	0,90%	0,9% (0,6% - 1,2%)	0,28%
Nega tosse	38	0,88%	0,9% (0,6% - 1,2%)	0,28%
Purulenta	31	0,72%	0,7% (0,5% - 1%)	0,25%
Edema	28	0,65%	0,6% (0,4% - 0,9%)	0,24%
Gripe	28	0,65%	0,6% (0,4% - 0,9%)	0,24%
Hialina	26	0,60%	0,6% (0,4% - 0,8%)	0,23%
Nódulo	23	0,53%	0,5% (0,3% - 0,7%)	0,22%
Esforço	23	0,53%	0,5% (0,3% - 0,7%)	0,22%
Calafrios	22	0,51%	0,5% (0,3% - 0,7%)	0,21%
Pirose	21	0,49%	0,5% (0,3% - 0,7%)	0,21%
Outros	18	0,42%	0,4% (0,2% - 0,6%)	0,19%
Hipoacusia	17	0,39%	0,4% (0,2% - 0,6%)	0,19%
Pressão	17	0,39%	0,4% (0,2% - 0,6%)	0,19%
Azia	16	0,37%	0,4% (0,2% - 0,6%)	0,18%
Amarelada	16	0,37%	0,4% (0,2% - 0,6%)	0,18%
Nega odinofagia	16	0,37%	0,4% (0,2% - 0,6%)	0,18%
Nega dispneia	15	0,35%	0,3% (0,2% - 0,5%)	0,18%
Fadiga	15	0,35%	0,3% (0,2% - 0,5%)	0,18%
Tontura	13	0,30%	0,3% (0,1% - 0,5%)	0,16%
Nega disfonia	12	0,28%	0,3% (0,1% - 0,4%)	0,16%
Nega mialgia	11	0,25%	0,3% (0,1% - 0,4%)	0,15%

FONTE: O autor (2017).

TABELA 15 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM OTITE EXTERNA NÃO ESPECIFICADA (CID H60.9) N=4.292

Sintoma	n	% de prevalência estimada	Intervalo de confiança (95%)	Margem de erro
Otalgia	2037	47,46%	47,5% (46% - 49%)	1,49%
Otorreia	796	18,55%	18,5% (17,4% - 19,7%)	1,16%
Dor	581	13,54%	13,5% (12,5% - 14,6%)	1,02%
Febre	545	12,70%	12,7% (11,7% - 13,7%)	1,00%
Hipoacusia	411	9,58%	9,6% (8,7% - 10,5%)	0,88%
Prurido	362	8,43%	8,4% (7,6% - 9,3%)	0,83%
Nega febre	301	7,01%	7% (6,2% - 7,8%)	0,76%
Edema	201	4,68%	4,7% (4,1% - 5,3%)	0,63%
Nega otorreia	195	4,54%	4,5% (3,9% - 5,2%)	0,62%
Zumbido	152	3,54%	3,5% (3% - 4,1%)	0,55%
Hiperemia	150	3,49%	3,5% (2,9% - 4%)	0,55%
Tosse	127	2,96%	3% (2,5% - 3,5%)	0,51%
Secreção	105	2,45%	2,4% (2% - 2,9%)	0,46%
Odinofagia	51	1,19%	1,2% (0,9% - 1,5%)	0,32%
Nega hipoacusia	51	1,19%	1,2% (0,9% - 1,5%)	0,32%
Trauma	47	1,10%	1,1% (0,8% - 1,4%)	0,31%
Otorragia	42	0,98%	1% (0,7% - 1,3%)	0,29%
Descamação	42	0,98%	1% (0,7% - 1,3%)	0,29%
Obstrução	41	0,96%	1% (0,7% - 1,2%)	0,29%
Cefaleia	39	0,91%	0,9% (0,6% - 1,2%)	0,28%
Congestão	31	0,72%	0,7% (0,5% - 1%)	0,25%
Ronco	27	0,63%	0,6% (0,4% - 0,9%)	0,24%
Nega otalgia	25	0,58%	0,6% (0,4% - 0,8%)	0,23%
Tontura	24	0,56%	0,6% (0,3% - 0,8%)	0,22%
Corpo estranho	19	0,44%	0,4% (0,2% - 0,6%)	0,20%
Disfonia	18	0,42%	0,4% (0,2% - 0,6%)	0,19%
Amarelada	18	0,42%	0,4% (0,2% - 0,6%)	0,19%
Espirro	17	0,40%	0,4% (0,2% - 0,6%)	0,19%
Nega dor	16	0,37%	0,4% (0,2% - 0,6%)	0,18%
Nega tosse	15	0,35%	0,3% (0,2% - 0,5%)	0,18%
Pressão	14	0,33%	0,3% (0,2% - 0,5%)	0,17%
Infecção	14	0,33%	0,3% (0,2% - 0,5%)	0,17%
Sangramento	14	0,33%	0,3% (0,2% - 0,5%)	0,17%
Crosta	12	0,28%	0,3% (0,1% - 0,4%)	0,16%
Gripe	10	0,23%	0,2% (0,1% - 0,4%)	0,14%
Perfuração	10	0,23%	0,2% (0,1% - 0,4%)	0,14%
Vertigem	10	0,23%	0,2% (0,1% - 0,4%)	0,14%
Purulenta	9	0,21%	0,2% (0,1% - 0,3%)	0,14%
Depressão	9	0,21%	0,2% (0,1% - 0,3%)	0,14%
Crostas	8	0,19%	0,2% (0,1% - 0,3%)	0,13%
Crônico	8	0,19%	0,2% (0,1% - 0,3%)	0,13%
Nega prurido	8	0,19%	0,2% (0,1% - 0,3%)	0,13%
Aspiração	8	0,19%	0,2% (0,1% - 0,3%)	0,13%
Hialina	7	0,16%	0,2% (0% - 0,3%)	0,12%

FONTE: O autor (2017).

TABELA 16 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM TONTURA E INSTABILIDADE (CID R42) N=3.880

Sintoma	n	% de prevalência estimada	Intervalo de confiança (95%)	Margem de erro
Tontura	2709	69,82%	69,8% (68,4% - 71,3%)	1,44%
Náusea	1137	29,30%	29,3% (27,9% - 30,7%)	1,43%
Vertigem	673	17,35%	17,3% (16,2% - 18,5%)	1,19%
Zumbido	646	16,65%	16,6% (15,5% - 17,8%)	1,17%
Hipoacusia	494	12,73%	12,7% (11,7% - 13,8%)	1,05%
Cefaleia	436	11,24%	11,2% (10,2% - 12,2%)	0,99%
Vômito	330	8,51%	8,5% (7,6% - 9,4%)	0,88%
Dor	259	6,68%	6,7% (5,9% - 7,5%)	0,79%
Nega hipoacusia	173	4,46%	4,5% (3,8% - 5,1%)	0,65%
Pressão	140	3,61%	3,6% (3% - 4,2%)	0,59%
Nega náusea	124	3,20%	3,2% (2,6% - 3,7%)	0,55%
Otalgia	100	2,58%	2,6% (2,1% - 3,1%)	0,50%
Ansiedade	96	2,47%	2,5% (2% - 3%)	0,49%
Mal estar	81	2,09%	2,1% (1,6% - 2,5%)	0,45%
Depressão	57	1,47%	1,5% (1,1% - 1,8%)	0,38%
Tosse	52	1,34%	1,3% (1% - 1,7%)	0,36%
Estresse	51	1,31%	1,3% (1% - 1,7%)	0,36%
Nega vômito	50	1,29%	1,3% (0,9% - 1,6%)	0,35%
Nega zumbido	49	1,26%	1,3% (0,9% - 1,6%)	0,35%
Nega cefaleia	49	1,26%	1,3% (0,9% - 1,6%)	0,35%
Sudorese	41	1,06%	1,1% (0,7% - 1,4%)	0,32%
Prurido	32	0,82%	0,8% (0,5% - 1,1%)	0,28%
Flutuação	32	0,82%	0,8% (0,5% - 1,1%)	0,28%
Febre	31	0,80%	0,8% (0,5% - 1,1%)	0,28%
Otorreia	30	0,77%	0,8% (0,5% - 1%)	0,28%
Obstrução	30	0,77%	0,8% (0,5% - 1%)	0,28%
RGE	30	0,77%	0,8% (0,5% - 1%)	0,28%
Trauma	29	0,75%	0,7% (0,5% - 1%)	0,27%
Congestão	24	0,62%	0,6% (0,4% - 0,9%)	0,25%
Sinusite	24	0,62%	0,6% (0,4% - 0,9%)	0,25%
AVC	23	0,59%	0,6% (0,4% - 0,8%)	0,24%
Ronco	22	0,57%	0,6% (0,3% - 0,8%)	0,24%
Nistagmo	22	0,57%	0,6% (0,3% - 0,8%)	0,24%
Elação	20	0,52%	0,5% (0,3% - 0,7%)	0,23%
Mialgia	19	0,49%	0,5% (0,3% - 0,7%)	0,22%
Outros	17	0,44%	0,4% (0,2% - 0,6%)	0,21%
Perda auditiva	17	0,44%	0,4% (0,2% - 0,6%)	0,21%
Nega otorreia	16	0,41%	0,4% (0,2% - 0,6%)	0,20%
Pigarro	16	0,41%	0,4% (0,2% - 0,6%)	0,20%
Nega febre	15	0,39%	0,4% (0,2% - 0,6%)	0,20%
Agudo	15	0,39%	0,4% (0,2% - 0,6%)	0,20%
Secreção	13	0,34%	0,3% (0,2% - 0,5%)	0,18%
Odinofagia	13	0,34%	0,3% (0,2% - 0,5%)	0,18%
Tabagismo	13	0,34%	0,3% (0,2% - 0,5%)	0,18%

FONTE: O autor (2017).

TABELA 17 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM OUTRAS OTITES MÉDIAS GUDAS NÃO-SUPURATIVAS (CID H65.1) N=3.595

Sintoma	n	% de prevalência estimada	Intervalo de confiança (95%)	Margem de erro
Otalgia	2049	57,00%	57% (55,4% - 58,6%)	1,62%
Febre	1141	31,74%	31,7% (30,2% - 33,3%)	1,52%
Dor	728	20,25%	20,3% (18,9% - 21,6%)	1,31%
Tosse	607	16,88%	16,9% (15,7% - 18,1%)	1,22%
Nega febre	410	11,40%	11,4% (10,4% - 12,4%)	1,04%
Otorreia	349	9,71%	9,7% (8,7% - 10,7%)	0,97%
Congestão	311	8,65%	8,7% (7,7% - 9,6%)	0,92%
Hipoacusia	269	7,48%	7,5% (6,6% - 8,3%)	0,86%
Secreção	206	5,73%	5,7% (5% - 6,5%)	0,76%
Odinofagia	129	3,59%	3,6% (3% - 4,2%)	0,61%
Hiperemia	125	3,48%	3,5% (2,9% - 4,1%)	0,60%
Obstrução	115	3,20%	3,2% (2,6% - 3,8%)	0,58%
Nega otorreia	105	2,92%	2,9% (2,4% - 3,5%)	0,55%
Cefaleia	99	2,75%	2,8% (2,2% - 3,3%)	0,53%
Ronco	65	1,81%	1,8% (1,4% - 2,2%)	0,44%
Zumbido	59	1,64%	1,6% (1,2% - 2,1%)	0,42%
Abaulamento	46	1,28%	1,3% (0,9% - 1,6%)	0,37%
Prurido	45	1,25%	1,3% (0,9% - 1,6%)	0,36%
Vômito	43	1,20%	1,2% (0,8% - 1,6%)	0,36%
Espirro	43	1,20%	1,2% (0,8% - 1,6%)	0,36%
Disfonia	43	1,20%	1,2% (0,8% - 1,6%)	0,36%
Gripe	31	0,86%	0,9% (0,6% - 1,2%)	0,30%
Nega dor	28	0,78%	0,8% (0,5% - 1,1%)	0,29%
Amarelada	28	0,78%	0,8% (0,5% - 1,1%)	0,29%
Purulenta	24	0,67%	0,7% (0,4% - 0,9%)	0,27%
Nega tosse	23	0,64%	0,6% (0,4% - 0,9%)	0,26%
Hialina	22	0,61%	0,6% (0,4% - 0,9%)	0,25%
Pressão	21	0,58%	0,6% (0,3% - 0,8%)	0,25%
Sinusite	21	0,58%	0,6% (0,3% - 0,8%)	0,25%
Mialgia	19	0,53%	0,5% (0,3% - 0,8%)	0,24%
Irritabilidade	19	0,53%	0,5% (0,3% - 0,8%)	0,24%
Mal estar	18	0,50%	0,5% (0,3% - 0,7%)	0,23%
Nega hipoacusia	16	0,45%	0,4% (0,2% - 0,7%)	0,22%
Otorragia	15	0,42%	0,4% (0,2% - 0,6%)	0,21%
Disfagia	14	0,39%	0,4% (0,2% - 0,6%)	0,20%
Edema	11	0,31%	0,3% (0,1% - 0,5%)	0,18%
Prurido nasal	10	0,28%	0,3% (0,1% - 0,5%)	0,17%
Epistaxe	9	0,25%	0,3% (0,1% - 0,4%)	0,16%
Tontura	8	0,22%	0,2% (0,1% - 0,4%)	0,15%
Nega ronco	8	0,22%	0,2% (0,1% - 0,4%)	0,15%
Dispneia	7	0,19%	0,2% (0,1% - 0,3%)	0,14%
Halitose	7	0,19%	0,2% (0,1% - 0,3%)	0,14%
Depressão	6	0,17%	0,2% (0% - 0,3%)	0,13%
Nega otalgia	6	0,17%	0,2% (0% - 0,3%)	0,13%

FONTE: O autor (2017).

TABELA 18 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM APNEIA DO SONO (CID G47.3) N=3.181

Sintoma	n	% de prevalência estimada	Intervalo de confiança (95%)	Margem de erro
Ronco	2496	78,47%	78,5% (77% - 79,9%)	1,43%
Obstrução	540	16,98%	17% (15,7% - 18,3%)	1,30%
Dor	412	12,95%	13% (11,8% - 14,1%)	1,17%
Tabagismo	190	5,97%	6% (5,1% - 6,8%)	0,82%
Nega tabagismo	155	4,87%	4,9% (4,1% - 5,6%)	0,75%
Prurido	143	4,50%	4,5% (3,8% - 5,2%)	0,72%
Espirro	137	4,31%	4,3% (3,6% - 5%)	0,71%
Nega Obstrução	125	3,93%	3,9% (3,3% - 4,6%)	0,68%
Tosse	85	2,67%	2,7% (2,1% - 3,2%)	0,56%
Pressão	85	2,67%	2,7% (2,1% - 3,2%)	0,56%
Prurido nasal	79	2,48%	2,5% (1,9% - 3%)	0,54%
Hipoacusia	77	2,42%	2,4% (1,9% - 3%)	0,53%
Febre	70	2,20%	2,2% (1,7% - 2,7%)	0,51%
Cefaleia	61	1,92%	1,9% (1,4% - 2,4%)	0,48%
Congestão	50	1,57%	1,6% (1,1% - 2%)	0,43%
Fadiga	45	1,41%	1,4% (1% - 1,8%)	0,41%
Otalgia	37	1,16%	1,2% (0,8% - 1,5%)	0,37%
Pigarro	36	1,13%	1,1% (0,8% - 1,5%)	0,37%
Palidez	33	1,04%	1% (0,7% - 1,4%)	0,35%
Secreção	32	1,01%	1% (0,7% - 1,4%)	0,35%
Nega febre	27	0,85%	0,8% (0,5% - 1,2%)	0,32%
RGE	27	0,85%	0,8% (0,5% - 1,2%)	0,32%
Desvio septal	27	0,85%	0,8% (0,5% - 1,2%)	0,32%
Otorreia	23	0,72%	0,7% (0,4% - 1%)	0,29%
Edema	22	0,69%	0,7% (0,4% - 1%)	0,29%
Zumbido	21	0,66%	0,7% (0,4% - 0,9%)	0,28%
Sinusite	21	0,66%	0,7% (0,4% - 0,9%)	0,28%
Trauma	21	0,66%	0,7% (0,4% - 0,9%)	0,28%
Engasgos	21	0,66%	0,7% (0,4% - 0,9%)	0,28%
Caseum	21	0,66%	0,7% (0,4% - 0,9%)	0,28%
Epistaxe	20	0,63%	0,6% (0,4% - 0,9%)	0,27%
Disfonia	18	0,57%	0,6% (0,3% - 0,8%)	0,26%
Matinal	18	0,57%	0,6% (0,3% - 0,8%)	0,26%
Perda de peso	18	0,57%	0,6% (0,3% - 0,8%)	0,26%
Depressão	17	0,53%	0,5% (0,3% - 0,8%)	0,25%
Odinofagia	16	0,50%	0,5% (0,3% - 0,7%)	0,25%
Halitose	16	0,50%	0,5% (0,3% - 0,7%)	0,25%
Gestação	15	0,47%	0,5% (0,2% - 0,7%)	0,24%
Ansiedade	15	0,47%	0,5% (0,2% - 0,7%)	0,24%
Hiperemia	14	0,44%	0,4% (0,2% - 0,7%)	0,23%
Tontura	14	0,44%	0,4% (0,2% - 0,7%)	0,23%
Pirose	14	0,44%	0,4% (0,2% - 0,7%)	0,23%
Nega ronco	13	0,41%	0,4% (0,2% - 0,6%)	0,22%
Nega espirro	13	0,41%	0,4% (0,2% - 0,6%)	0,22%

FONTE: O autor (2017).

TABELA 19 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM PERDA DE AUDIÇÃO (CID H91) N=3.162

Sintoma	n	% de prevalência estimada	Intervalo de confiança (95%)	Margem de erro
Hipoacusia	1791	56,64%	56,6% (54,9% - 58,4%)	1,73%
Otalgia	379	11,99%	12% (10,9% - 13,1%)	1,13%
Zumbido	280	8,86%	8,9% (7,9% - 9,8%)	0,99%
Otorreia	269	8,51%	8,5% (7,5% - 9,5%)	0,97%
Ronco	162	5,12%	5,1% (4,4% - 5,9%)	0,77%
Nega otorreia	161	5,09%	5,1% (4,3% - 5,9%)	0,77%
Dor	118	3,73%	3,7% (3,1% - 4,4%)	0,66%
Prurido	106	3,35%	3,4% (2,7% - 4%)	0,63%
Tontura	105	3,32%	3,3% (2,7% - 3,9%)	0,62%
Perda auditiva	82	2,59%	2,6% (2% - 3,1%)	0,55%
Obstrução	77	2,44%	2,4% (1,9% - 3%)	0,54%
Nega otalgia	76	2,40%	2,4% (1,9% - 2,9%)	0,53%
Tosse	42	1,33%	1,3% (0,9% - 1,7%)	0,40%
Febre	41	1,30%	1,3% (0,9% - 1,7%)	0,39%
Trauma	41	1,30%	1,3% (0,9% - 1,7%)	0,39%
Pressão	38	1,20%	1,2% (0,8% - 1,6%)	0,38%
Cefaleia	35	1,11%	1,1% (0,7% - 1,5%)	0,36%
Nega ronco	35	1,11%	1,1% (0,7% - 1,5%)	0,36%
Vertigem	35	1,11%	1,1% (0,7% - 1,5%)	0,36%
Gestação	34	1,08%	1,1% (0,7% - 1,4%)	0,36%
Nega zumbido	32	1,01%	1% (0,7% - 1,4%)	0,35%
Espirro	25	0,79%	0,8% (0,5% - 1,1%)	0,31%
Congestão	25	0,79%	0,8% (0,5% - 1,1%)	0,31%
Nega hipoacusia	24	0,76%	0,8% (0,5% - 1,1%)	0,30%
Pigarro	21	0,66%	0,7% (0,4% - 0,9%)	0,28%
Nega febre	21	0,66%	0,7% (0,4% - 0,9%)	0,28%
Sinusite	21	0,66%	0,7% (0,4% - 0,9%)	0,28%
Nega tontura	21	0,66%	0,7% (0,4% - 0,9%)	0,28%
Disfonia	19	0,60%	0,6% (0,3% - 0,9%)	0,27%
Agudo	19	0,60%	0,6% (0,3% - 0,9%)	0,27%
Epistaxe	18	0,57%	0,6% (0,3% - 0,8%)	0,26%
AVC	18	0,57%	0,6% (0,3% - 0,8%)	0,26%
Retração	18	0,57%	0,6% (0,3% - 0,8%)	0,26%
Sensação de plenitude auricular	17	0,54%	0,5% (0,3% - 0,8%)	0,25%
Secreção	16	0,51%	0,5% (0,3% - 0,8%)	0,25%
Odinofagia	16	0,51%	0,5% (0,3% - 0,8%)	0,25%
Autofonia	16	0,51%	0,5% (0,3% - 0,8%)	0,25%
Tabagismo	14	0,44%	0,4% (0,2% - 0,7%)	0,23%
Nega obstrução	14	0,44%	0,4% (0,2% - 0,7%)	0,23%
Prurido nasal	14	0,44%	0,4% (0,2% - 0,7%)	0,23%
Infecção	14	0,44%	0,4% (0,2% - 0,7%)	0,23%
Náusea	12	0,38%	0,4% (0,2% - 0,6%)	0,21%
Nega tabagismo	10	0,32%	0,3% (0,1% - 0,5%)	0,20%
Coma	10	0,32%	0,3% (0,1% - 0,5%)	0,20%

FONTE: O autor (2017).

TABELA 20 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM DOR DE GARGANTA (CID R07.0) N=2.669

Sintoma	n	% de prevalência estimada	Intervalo de confiança (95%)	Margem de erro
Odinofagia	1306	48,93%	48,9% (47% - 50,8%)	1,90%
Dor	940	35,22%	35,2% (33,4% - 37%)	1,81%
Febre	850	31,85%	31,8% (30,1% - 33,6%)	1,77%
Tosse	469	17,57%	17,6% (16,1% - 19%)	1,44%
Nega febre	423	15,85%	15,8% (14,5% - 17,2%)	1,39%
Disfonia	185	6,93%	6,9% (6% - 7,9%)	0,96%
Otalgia	159	5,96%	6% (5,1% - 6,9%)	0,90%
Mal estar	149	5,58%	5,6% (4,7% - 6,5%)	0,87%
Congestão	130	4,87%	4,9% (4,1% - 5,7%)	0,82%
Cefaleia	103	3,86%	3,9% (3,1% - 4,6%)	0,73%
Hiperemia	95	3,56%	3,6% (2,9% - 4,3%)	0,70%
Pigarro	88	3,30%	3,3% (2,6% - 4%)	0,68%
Obstrução	85	3,18%	3,2% (2,5% - 3,9%)	0,67%
Secreção	56	2,10%	2,1% (1,6% - 2,6%)	0,54%
Corpo estranho	46	1,72%	1,7% (1,2% - 2,2%)	0,49%
Edema	44	1,65%	1,6% (1,2% - 2,1%)	0,48%
Mialgia	41	1,54%	1,5% (1,1% - 2%)	0,47%
Caseum	41	1,54%	1,5% (1,1% - 2%)	0,47%
Prurido	39	1,46%	1,5% (1% - 1,9%)	0,46%
Tabagismo	39	1,46%	1,5% (1% - 1,9%)	0,46%
Nega tabagismo	33	1,24%	1,2% (0,8% - 1,7%)	0,42%
Sensação de corpo estranho	33	1,24%	1,2% (0,8% - 1,7%)	0,42%
Pirose	32	1,20%	1,2% (0,8% - 1,6%)	0,41%
Ronco	31	1,16%	1,2% (0,8% - 1,6%)	0,41%
RGE	30	1,12%	1,1% (0,7% - 1,5%)	0,40%
Espirro	27	1,01%	1% (0,6% - 1,4%)	0,38%
Disfagia	25	0,94%	0,9% (0,6% - 1,3%)	0,37%
Cervicalgia	22	0,82%	0,8% (0,5% - 1,2%)	0,34%
Hipoacusia	21	0,79%	0,8% (0,5% - 1,1%)	0,34%
Azia	21	0,79%	0,8% (0,5% - 1,1%)	0,34%
Palidez	20	0,75%	0,7% (0,4% - 1,1%)	0,33%
Aumento de volume	19	0,71%	0,7% (0,4% - 1%)	0,32%
Queimação	19	0,71%	0,7% (0,4% - 1%)	0,32%
Purulenta	17	0,64%	0,6% (0,3% - 0,9%)	0,30%
Nega tosse	17	0,64%	0,6% (0,3% - 0,9%)	0,30%
Gripe	16	0,60%	0,6% (0,3% - 0,9%)	0,29%
Nega disfonia	16	0,60%	0,6% (0,3% - 0,9%)	0,29%
Pressão	14	0,52%	0,5% (0,3% - 0,8%)	0,27%
Placa	14	0,52%	0,5% (0,3% - 0,8%)	0,27%
Vômito	14	0,52%	0,5% (0,3% - 0,8%)	0,27%
Dispneia	14	0,52%	0,5% (0,3% - 0,8%)	0,27%
Trauma	13	0,49%	0,5% (0,2% - 0,8%)	0,26%
Halitose	13	0,49%	0,5% (0,2% - 0,8%)	0,26%
Hialina	13	0,49%	0,5% (0,2% - 0,8%)	0,26%

FONTE: O autor (2017).

TABELA 21 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM OUTRAS RINITES ALÉRGICAS (CID J30) N=2.644

Sintoma	n	% de prevalência estimada	Intervalo de confiança (95%)	Margem de erro
Obstrução	485	18,34%	18,3% (16,9% - 19,8%)	1,48%
Tosse	403	15,24%	15,2% (13,9% - 16,6%)	1,37%
Dor	370	13,99%	14% (12,7% - 15,3%)	1,32%
Espirro	338	12,78%	12,8% (11,5% - 14,1%)	1,27%
Prurido	287	10,85%	10,9% (9,7% - 12%)	1,19%
Febre	285	10,78%	10,8% (9,6% - 12%)	1,18%
Cefaleia	248	9,38%	9,4% (8,3% - 10,5%)	1,11%
Congestão	221	8,36%	8,4% (7,3% - 9,4%)	1,05%
Ronco	200	7,56%	7,6% (6,6% - 8,6%)	1,01%
Secreção	194	7,34%	7,3% (6,3% - 8,3%)	0,99%
Otalgia	92	3,48%	3,5% (2,8% - 4,2%)	0,70%
Nega febre	90	3,40%	3,4% (2,7% - 4,1%)	0,69%
Odinofagia	87	3,29%	3,3% (2,6% - 4%)	0,68%
Epistaxe	72	2,72%	2,7% (2,1% - 3,3%)	0,62%
Hipoacusia	71	2,69%	2,7% (2,1% - 3,3%)	0,62%
Prurido nasal	71	2,69%	2,7% (2,1% - 3,3%)	0,62%
Purulenta	49	1,85%	1,9% (1,3% - 2,4%)	0,51%
Hialina	49	1,85%	1,9% (1,3% - 2,4%)	0,51%
Sinusite	49	1,85%	1,9% (1,3% - 2,4%)	0,51%
Disfonia	43	1,63%	1,6% (1,1% - 2,1%)	0,48%
Edema	39	1,48%	1,5% (1% - 1,9%)	0,46%
Palidez	39	1,48%	1,5% (1% - 1,9%)	0,46%
Hiperemia	31	1,17%	1,2% (0,8% - 1,6%)	0,41%
Pressão	30	1,13%	1,1% (0,7% - 1,5%)	0,40%
Hiposmia	30	1,13%	1,1% (0,7% - 1,5%)	0,40%
Pigarro	26	0,98%	1% (0,6% - 1,4%)	0,38%
Mal estar	23	0,87%	0,9% (0,5% - 1,2%)	0,35%
Amarelada	20	0,76%	0,8% (0,4% - 1,1%)	0,33%
Otorreia	20	0,76%	0,8% (0,4% - 1,1%)	0,33%
Crosta	20	0,76%	0,8% (0,4% - 1,1%)	0,33%
Nega ronco	20	0,76%	0,8% (0,4% - 1,1%)	0,33%
Nega espirro	20	0,76%	0,8% (0,4% - 1,1%)	0,33%
Tontura	19	0,72%	0,7% (0,4% - 1%)	0,32%
Crônico	19	0,72%	0,7% (0,4% - 1%)	0,32%
Crostas	18	0,68%	0,7% (0,4% - 1%)	0,31%
Zumbido	16	0,61%	0,6% (0,3% - 0,9%)	0,30%
Nega prurido	16	0,61%	0,6% (0,3% - 0,9%)	0,30%
Gripe	15	0,57%	0,6% (0,3% - 0,9%)	0,29%
Sangramento	15	0,57%	0,6% (0,3% - 0,9%)	0,29%
Mialgia	11	0,42%	0,4% (0,2% - 0,7%)	0,25%
Caseum	11	0,42%	0,4% (0,2% - 0,7%)	0,25%
RGE	10	0,38%	0,4% (0,1% - 0,6%)	0,23%
Nega tosse	10	0,38%	0,4% (0,1% - 0,6%)	0,23%
Dispneia	10	0,38%	0,4% (0,1% - 0,6%)	0,23%

FONTE: O autor (2017).

TABELA 22 - PREVALÊNCIA DE SINTOMAS CONFORME TERMOS DO PROTOCOLO EM FARINGITE AGUDA DEVIDA A OUTROS MICROORGANISMOS ESPECIFICADOS (CID J02.8)
N=2.393

Sintoma	n	% de prevalência estimada	Intervalo de confiança (95%)	Margem de erro
Odinofagia	2391	99,92%	99,9% (99,8% - 100%)	0,12%
Tosse	2384	99,62%	99,6% (99,4% - 99,9%)	0,25%
Febre	2372	99,12%	99,1% (98,7% - 99,5%)	0,37%
RGE	11	0,46%	0,5% (0,2% - 0,7%)	0,27%
Gripe	6	0,25%	0,3% (0,1% - 0,5%)	0,20%
Dor	4	0,17%	0,2% (0% - 0,3%)	0,16%
Disfonia	4	0,17%	0,2% (0% - 0,3%)	0,16%
Aftas	4	0,17%	0,2% (0% - 0,3%)	0,16%
Cefaleia	2	0,08%	0,1% (0% - 0,2%)	0,12%
Otalgia	2	0,08%	0,1% (0% - 0,2%)	0,12%
Hiposmia	2	0,08%	0,1% (0% - 0,2%)	0,12%
Obstrução	1	0,04%	0% (0% - 0,1%)	0,08%
Ronco	1	0,04%	0% (0% - 0,1%)	0,08%
Hipoacusia	1	0,04%	0% (0% - 0,1%)	0,08%
Sinusite	1	0,04%	0% (0% - 0,1%)	0,08%
Edema	1	0,04%	0% (0% - 0,1%)	0,08%
Pigarro	1	0,04%	0% (0% - 0,1%)	0,08%
Crônico	1	0,04%	0% (0% - 0,1%)	0,08%
Trauma	1	0,04%	0% (0% - 0,1%)	0,08%
Vômito	1	0,04%	0% (0% - 0,1%)	0,08%
Tabagismo	1	0,04%	0% (0% - 0,1%)	0,08%
Calor	1	0,04%	0% (0% - 0,1%)	0,08%
Nega tabagismo	1	0,04%	0% (0% - 0,1%)	0,08%
Placa	1	0,04%	0% (0% - 0,1%)	0,08%
Estresse	1	0,04%	0% (0% - 0,1%)	0,08%
Bolhas	1	0,04%	0% (0% - 0,1%)	0,08%
Bolhas	1	0,04%	0% (0% - 0,1%)	0,08%
Hemoptise	1	0,04%	0% (0% - 0,1%)	0,08%

FONTE: O autor (2017).

**APÊNDICE B – LISTA DE *STOP WORD*´S UTILIZADA NA ELABORAÇÃO
DAS *WORD CLOUD*´S**

“com de 1 2 3 4 5 6 7 8 9 10 meses dias refere anos semanas pos 2x em ha
mae para da por dias noite hj esq dir nasal por tem pior estar desde teve 12 sempre
45 dos comorbidades uso red resp há hs fez filha apos que ndn nao ou med ontem
direita esquerda queixa ao pcte od oe hoje cx tto plenitude leve relata mais melhora
piora predsim avamys cxs enc pelo dr allegra nasonex pela mãe teste outras não
paciente doença s#“

ANEXO A – ANÁLISE DO COMITÊ DE ÉTICA



HOSPITAL PARANAENSE DE
OTORRINOLARINGOLOGIA
LTDA - IPO



PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Extração de palavras-chave em textos livres não estruturados de prontuários eletrônicos e comparação de diagnósticos a partir de um protocolo de otorrinolaringologia

Pesquisador: ROGERIO HAMERSCHMIDT

Área Temática:

Versão: 2

CAAE: 79919417.4.0000.5529

Instituição Proponente: Instituto Paranaense de Otorrinolaringologia Ltda - IPO

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 2.414.444

Apresentação do Projeto:

Projeto avaliado anteriormente através do Número do Parecer: 2.409.885

Projeto de conclusão de curso do Programa de Pós-Graduação em Clínica Cirúrgica da Universidade Federal do Paraná .

O estudo será uma revisão retrospectiva de prontuários eletrônicos e será realizado no Hospital IPO.

Trata-se da extração de palavras em textos livres, não estruturados, inseridos em prontuário eletrônico e a comparação do resultado da frequência das mais prevalentes com um protocolo de coleta de dados de uma determinada especialidade, otorrinolaringologia, organizado em um sistema eletrônico denominado SINPE© - Sistema Integrado de Protocolos Eletrônicos, desenvolvido em pesquisas anteriormente publicadas. O resultado desta comparação será cruzado com o diagnóstico final de cada atendimento, codificado pela Classificação Internacional de Doenças, 10ª edição (CID-10). Busca-se demonstrar se estas palavras do protocolo são prevalentes no prontuário e capazes de definir o diagnóstico do paciente.

Endereço: REPUBLICA ARGENTINA 826/2080

Bairro: AGUA VERDE

CEP: 80.620-010

UF: PR

Município: CURITIBA

Telefone: (41)3094-5751

Fax: (41)3314-1500

E-mail: nep@ipo.com.br



OUVIDO • NARIZ • GARGANTA

Continuação do Parecer: 2.414.444

HOSPITAL PARANAENSE DE
OTORRINOLARINGOLOGIA
LTDA - IPO



O projeto apresenta-se bem estruturado, simples e claro para o entendimento. Se comprovada sua hipótese, ajudará na elaboração de outros estudos científicos futuros.

Objetivo da Pesquisa:

O objetivo está claro e bem disposto. Procurará identificar palavras chaves mais prevalentes para cada diagnóstico em otorrinolaringologia, que possa identificar tal patologia, em textos livres, criando uma nuvem de palavras chaves de identificação.

Os objetivos apresentados pelos pesquisadores são:

Objetivo Geral:

Extrair palavras em textos livres não estruturadas em prontuário eletrônico e comparar sua frequência com um protocolo já existente de palavras, para confirmação de doenças de maior prevalência em uma base de dados.

Objetivos Específicos - Detectar as palavras mais frequentes em texto livre nas doenças de maior prevalência; - Criar world cloud's (nuvens de palavras) das palavras mais frequentes por doença de maior prevalência; - Identificar as palavras do protocolo eletrônico SINPE© nos textos livres e sua comparação com as doenças de maior prevalência.

O objetivo está claro e bem delimitado. Tem o intuito de, através de pesquisa de palavras mais usadas pelos médicos para descrever nos prontuários as doenças, desenvolver um sistema que possa identificar estas palavras para confirmar o diagnóstico das patologias, facilitando posteriormente pesquisa de dados.

Avaliação dos Riscos e Benefícios:

O projeto por se tratar de um estudo de revisão de prontuários não apresenta risco físico aos participantes da pesquisa.

Os riscos poderiam ser de possível exposição dos dados dos pacientes, porém os autores se comprometem a tomar as devidas medidas para evitar que isto aconteça.

O projeto não apresenta nenhum benefício direto aos participantes porém caso a hipótese venha a ser confirmada pode trazer benefícios para a comunidade médica e para pacientes futuros e ainda servir como uma ferramenta a mais de pesquisa, através de palavras chaves dos textos livres em prontuários médicos.

Comentários e Considerações sobre a Pesquisa:

Lembramos ao pesquisador que os procedimentos só poderão ser iniciados a partir da emissão do parecer do Comitê de Ética em Pesquisa em Seres Humanos de acordo com a resolução do

Endereço: REPUBLICA ARGENTINA 826/2080
Bairro: AGUA VERDE **CEP:** 80.620-010
UF: PR **Município:** CURITIBA
Telefone: (41)3094-5751 **Fax:** (41)3314-1500 **E-mail:** nep@ipo.com.br



HOSPITAL PARANAENSE DE
OTORRINOLARINGOLOGIA
LTDA - IPO



Continuação do Parecer: 2.414.444

Ministério da Saúde resolução 466/12.

Considerações sobre os Termos de apresentação obrigatória:

Solicita dispensa do TCLE por ser de análise de dados progressos, de prontuários médicos, sem intervenção direta aos pacientes.

Recomendações:

Recomenda-se que se descreva no materiais e métodos para que servem os softwares Microsoft SQL; Epi Info-CDC e SPSS-IBM-, e não simplesmente cita-los e explicar que apenas serão usados. Isto irá ajudar no entendimento de como os autores irão chegar ao desfecho proposto.

Conclusões ou Pendências e Lista de Inadequações:

Após as correções efetuadas pelo pesquisador o CEP considera o projeto Aprovado. Pedimos que o pesquisador verifique as recomendações

Considerações Finais a critério do CEP:

Projeto recebido no CEP na data de 08/11/2017.

Lembramos que não foi identificado no projeto a indicação da instituição proponente deste estudo. Lembramos também que de acordo com a regulamentação das pesquisas com seres humanos é de total responsabilidade do pesquisador o correto preenchimento das informações prestadas na plataforma Brasil.

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_1026328.pdf	01/12/2017 14:04:31		Aceito
Outros	CARTA_retorno_parecer.pdf	01/12/2017 14:03:12	ROGERIO HAMERSCHMIDT	Aceito
Brochura Pesquisa	Projeto_versao_CEP_IPO_1_2017_versao2.pdf	01/12/2017 13:54:39	ROGERIO HAMERSCHMIDT	Aceito
Projeto Detalhado / Brochura Investigador	ProjetoCEP_IPO_1_2017.pdf	09/11/2017 09:19:42	ROGERIO HAMERSCHMIDT	Aceito
Outros	Curriculo_Lattes_Claudio_Jose_Beltrao.pdf	08/11/2017 18:35:25	ROGERIO HAMERSCHMIDT	Aceito
Outros	Curriculo_Lattes_Rogério_Hamerschmidt.pdf	08/11/2017 18:35:04	ROGERIO HAMERSCHMIDT	Aceito
Outros	TCUD_termo_de_compromisso_de_ut	08/11/2017	ROGERIO	Aceito

Endereço: REPUBLICA ARGENTINA 826/2080

Bairro: AGUA VERDE

CEP: 80.620-010

UF: PR

Município: CURITIBA

Telefone: (41)3094-5751

Fax: (41)3314-1500

E-mail: nep@ipo.com.br



OUVIDO • NARIZ • GARGANTA

**HOSPITAL PARANAENSE DE
OTORRINOLARINGOLOGIA
LTDA - IPO**


Continuação do Parecer: 2.414.444

Outros	ilizacao_de_dados.pdf	18:34:36	HAMERSCHMIDT	Aceito
Outros	Pedido_isencao_termo_consentimento.pdf	08/11/2017 18:34:16	ROGERIO HAMERSCHMIDT	Aceito
Outros	Lista_da_Equipe_da_Pesquisa_Clinica.pdf	08/11/2017 18:33:59	ROGERIO HAMERSCHMIDT	Aceito
Outros	Declaracao_vinculo_institucional.pdf	08/11/2017 18:33:43	ROGERIO HAMERSCHMIDT	Aceito
Outros	Declaracao_do_orientador_do_colaborador.pdf	08/11/2017 18:33:27	ROGERIO HAMERSCHMIDT	Aceito
Outros	Declaracao_de_confidencialidade.pdf	08/11/2017 18:33:01	ROGERIO HAMERSCHMIDT	Aceito
Outros	Declaracao_de_compromisso_com_os_resultados.pdf	08/11/2017 18:32:35	ROGERIO HAMERSCHMIDT	Aceito
Outros	Cumprimento_das_Resolucoes_pelo_Investigador.pdf	08/11/2017 18:32:09	ROGERIO HAMERSCHMIDT	Aceito
Outros	Carta_de_encaminhamento.pdf	08/11/2017 18:31:49	ROGERIO HAMERSCHMIDT	Aceito
Outros	Carta_de_compromisso.pdf	08/11/2017 18:31:17	ROGERIO HAMERSCHMIDT	Aceito
Outros	Carta_de_autorizacao_da_instituicao.pdf	08/11/2017 18:30:59	ROGERIO HAMERSCHMIDT	Aceito
Folha de Rosto	Folha_de_rosto.pdf	08/11/2017 18:29:01	ROGERIO HAMERSCHMIDT	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

CURITIBA, 04 de Dezembro de 2017

Assinado por:
Evaldo Dacheux de Macedo Filho
(Coordenador)

Endereço: REPUBLICA ARGENTINA 826/2080

Bairro: AGUA VERDE

CEP: 80.620-010

UF: PR

Município: CURITIBA

Telefone: (41)3094-5751

Fax: (41)3314-1500

E-mail: nep@ipo.com.br

ANEXO B – CONTRATO DE LICENÇA DE USO DO SOFTWARE SINPE

Contrato de licença de uso do software Sinpe

Termo de Concessão de Direito de Uso Sinpe

(IMPORTANTE E LEIA COM ATENÇÃO)

Este é um contrato de licença de usuário final para o software Sinpe – Sistema Integrado de Protocolos Eletrônicos. Ele representa um acordo legal firmado em 01/04/2015 entre CLAUDIO JOSE BELTRAO, brasileiro, casado, portador do CPF 771.259.409-00 doravante designado como CONTRATADO e Osvaldo Malafala, brasileiro, casado, portador do CPF 007.016.709-59 doravante designado o CONTRATANTE. Ao utilizar o software Sinpe, seja por qualquer meio ou dispositivo o CONTRATADO estará concordando com este contrato. Caso não concorde poderá devolvê-lo.

Licença do Software:

O Sinpe é protegido por leis e tratados internacionais. Ele é licenciado e não vendido. Estes documentos é um contrato completo entre CONTRATADO e o CONTRATANTE relativo ao uso do software Sinpe. Ele prevalece sobre quaisquer comunicações, propostas ou representações orais ou escritas, anteriores ou contemporâneas e sempre as informações deste contrato devem prevalecer.

1. Concessão de licença

Esta seção descreve os direitos concedidos ao instalar e utilizar o software:

- 1.1. O CONTRATADO poderá instalar e utilizar o software em um único computador.
- 1.2. A licença do Sinpe não pode ser compartilhada ou cedida pelo CONTRATADO.
- 1.3. O CONTRATADO não pode utilizar o Sinpe para criar trabalhos que irão ferir, de maneira direta ou indireta, leis locais ou internacionais, a integridade de pessoas e empresas.
- 1.4. O CONTRATADO se compromete a utilizar o Sinpe somente para fins de pesquisas científica e ceder (como contrapartida de uso do software) a propriedade intelectual dos protocolos construídos utilizando o Sinpe ao CONTRATANTE, desde que o Sinpe tenha sido fornecido gratuitamente para este fim.
- 1.5. O CONTRATADO se compromete a construir protocolos utilizando extensa pesquisa bibliográfica sobre o assunto ao qual se propõe sob orientação do CONTRATANTE e que estes protocolos não incorporem, mesmo que de modo inadvertido, conhecimento falacioso.
- 1.6. Caso ocorra alguma infração nos itens 1.3, 1.4 e 1.5 mesmo que inadvertidamente, o CONTRATADO assumirá as responsabilidades por seus atos e irá defender e ressarcir o CONTRATANTE de quaisquer acusações ou prejuízos financeiros.

- 1.7. O CONTRATADO deverá indenizar isentar de responsabilidade e defender o CONTRATANTE de e contra quaisquer reivindicações ou ações judiciais, incluído honorários advocatícios que decorram ou resultem do mau uso do Sinpe.
- 1.8. O CONTRATADO concorda que o Sinpe poderá enviar informações sobre a instalação e código de ativação programa através da Internet ou outro meio de comunicação e nenhuma informação pessoal (dados do usuário ou dados de paciente) será enviada.
- 1.9. Todos os direitos que não sejam expressamente concedidos são reservados ao CONTRATANTE.

2. Outros direitos e limitações

- 2.1. É possível que o Sinpe seja provido de tecnologia de proteção contra cópia para impedir a reprodução não autorizada e é ilegal fazer cópias não autorizadas ou ignorar qualquer tecnologia de proteção contra cópia contida nele.
- 2.2. É proibido realizar engenharia reversa, descompilação ou desmontagem do Sinpe ou de qualquer parte que seja fornecido juntamente com ele (como bancos de dados e núcleo), exceto e somente na medida em que esta atividade seja expressamente permitida pela legislação aplicável, não obstante tal limitação.
- 2.3. O Sinpe é licenciado como um único produto e seus componentes não podem ser separados para o uso em mais de um dispositivo.
- 2.4. Este contrato não outorga ou atribui qualquer direito de exploração comercial ou de serviços do Sinpe e nem atribui direitos relativos às marcas comerciais ou de serviços do CONTRATANTE.
- 2.5. É proibido arrendar ou alugar o Sinpe.
- 2.6. O CONTRATANTE poderá lhe fornecer recursos de suporte ao Sinpe como correção de eventuais bugs e fornecimento de documentação para usuário final.
- 2.7. Caso o CONTRATADO não cumpra os termos aqui contidos, o CONTRATANTE poderá rescindir este contrato sem qualquer prejuízo ou perdas de direitos e no caso de uma rescisão o CONTRATADO deverá destruir todas as cópias do Sinpe e seus componentes.
- 2.8. Os trabalhos científicos resultantes da utilização do Sinpe, quer na forma de apresentação oral quer na impressa, deverão incluir como co-autores os nomes Osvaldo Malafaia e José Simão de Paula Pinto, citando no capítulo de material e métodos ou equivalente esta concessão de uso.

3. Exclusão de danos incidentais, consequenciais e outros

Na extensão máxima permitida pela legislação aplicável, em hipótese alguma o CONTRATANTE ou seus fornecedores serão responsáveis por qualquer dano especial, incidental indireto ou consequencial. Estão aqui incluídos danos por lucros cessantes, ou por perda de informações confidenciais ou outras; por interrupção nos negócios; por danos pessoais, por perda de privacidade; por falha no cumprimento de qualquer obrigação, inclusive de boa fé e com cuidados razoáveis; por negligência e por perda financeira de qualquer natureza. Estes fatos podem ser decorrentes de qualquer forma de relacionamento ao uso ou a incapacidade de

utilizar o software; ao fornecimento ou falha no fornecimento de serviços de suporte ou de outros modos sob ou com relação a qualquer disposição desde contrato - mesmo que haja falha, ato ilícito (inclusive negligências) -; responsabilidade restrita; quebra de contrato ou de garantia pelo CONTRATANTE ou qualquer fornecedor, mesmo que o CONTRATANTE, ou qualquer fornecedor tenha sido alertado sobre a possibilidade de tais danos.

Curitiba, 01/04/2015



CONTRATANTE
Osvaldo Malafaia



CONTRATADO
Claudio Jose Beltrão



TESTEMUNHA 1



TESTEMUNHA 2