



UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
CENTRO DE CIÊNCIAS EXATAS E DA TERRA
PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS E COMPUTAÇÃO
DOUTORADO EM CIÊNCIA DA COMPUTAÇÃO



Um Framework para Análise Multidimensional de Intervenções em Saúde Pública

Rafael de Moraes Pinto

Natal-RN
Set 2022

Rafael de Moraes Pinto

Um Framework para Análise Multidimensional de Intervenções em Saúde Pública

Tese apresentada ao Programa de Pós-Graduação em Sistemas e Computação do Departamento de Informática e Matemática Aplicada da Universidade Federal do Rio Grande do Norte como requisito para obtenção do título de Doutor em Ciência da Computação.

Área de pesquisa:
Engenharia de Software

Orientador:

Prof. Dr. Lyrene Fernandes da Silva

Coorientador:

Prof. Dr. Ricardo Alexsandro de Medeiros Valentim

PPGSC – PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS E COMPUTAÇÃO

CCET – CENTRO DE CIÊNCIAS EXATAS E DA TERRA

UFRN – UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE

Natal-RN

Set 2022

Universidade Federal do Rio Grande do Norte - UFRN
Sistema de Bibliotecas - SISBI
Catalogação de Publicação na Fonte. UFRN - Biblioteca Central Zila Mamede

Pinto, Rafael de Moraes.

Um framework para análise multidimensional de intervenções em saúde pública / Rafael de Moraes Pinto. - 2022.

146 f.: il.

Tese (doutorado) - Universidade Federal do Rio Grande do Norte, Centro de Ciências Exatas e da Terra, Programa de Pós-Graduação em Sistemas e Computação. Natal, RN, 2022.

Orientadora: Dra. Lyrene Fernandes da Silva.

Coorientador: Dr. Ricardo Aleksandro de Medeiros Valentim.

1. Framework - Engenharia de Softwares - Tese. 2. Análise de Séries Temporais - Tese. 3. Intervenção em Saúde Pública - Tese. I. Silva, Lyrene Fernandes da. II. Valentim, Ricardo Aleksandro de Medeiros. III. Título.

RN/UF/BCZM

CDU 004.41(043.2)

RAFAEL DE MORAIS PINTO

“Um Framework para Análise Multidimensional de Intervenções em Saúde Pública”

Esta Tese foi julgada adequada para a obtenção do título de Doutor(a) em Ciência da Computação e aprovada em sua forma final pelo Programa de Pós-Graduação em Sistemas e Computação do Departamento de Informática e Matemática Aplicada da Universidade Federal do Rio Grande do Norte.



Prof.^a Dr.^a ANNE MAGALY DE PAULA CANUTO
Vice-coordenadora do PPgSC

Banca Examinadora

DocuSigned by:



Examinador(a) Externo(a): **Dr.^a LYANE RAMALHO CORTEZ**

DocuSigned by:

Lyrene Fernandes da Silva

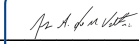
Presidente: **Dr. LYRENE FERNANDES DA SILVA**

DocuSigned by:



Examinador(a) Externo(a): **Dr. PLÁCIDO ANTÔNIO DE SOUZA NETO**

DocuSigned by:



Examinador(a) Externo(a): **Dr. RICARDO ALEXSANDRO DE MEDEIROS VALENTIM**

DocuSigned by:

Thaiza Gois Farias de Moura Santos Lima

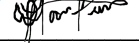
Examinador(a) Externo(a): **Dr. THAISA GOIS FARIAS DE MOURA SANTOS LIMA**

DocuSigned by:



Examinador(a) Interno(a): **Dr. UIRÁ KULESZA**

DocuSigned by:



Examinador(a) Externo(a): **Dr. WAGNER DE JESUS MARTINS**

DocuSigned by:

Ximena Pamela Claudia Diaz Bermúdez

Examinador(a) Externo(a): **Dr.^a XIMENA PAMELA CLÁUDIA DÍAS BERMÚDEZ**

DocuSigned by:

Rafael de Moraes Pinto

Discente: **RAFAEL DE MORAIS PINTO**

Natal, 02 de setembro de 2022

À minha esposa Talita, pelo amor, paciência e incentivo nos momentos difíceis ao longo desta jornada. Aos meus filhos, Gabriel e Miguel, por serem simplesmente meus filhos e absorverem um pouco do meu melhor a cada dia. Aos meus pais, Rubens e Gloriete, minha avó Marlene e tia Bereka que, com muito carinho e incentivo, não mediram esforços para que eu chegasse até esta etapa de minha vida.

Agradecimentos

Finalmente, após toda a trajetória para a conclusão desta tese, deixo aqui os registros de agradecimentos formais a todos aqueles que diretamente me ajudaram na conclusão desta etapa, contribuindo para meu amadurecimento científico e pessoal.

Primeiramente, sou grato a Deus pelas oportunidades e pelas pessoas que colocou em meu caminho. Sem Ele, nada disso teria acontecido.

Sou muito grato a minha esposa Talita pelo amor, carinho, paciência e companheirismo ao longo desses últimos 23 anos. Aos meus filhos Gabriel e Miguel pela compreensão em meus períodos de ausência.

Agradeço aos meus pais, Rubens e Gloriete, pelo alicerce na minha base acadêmica, sempre acreditando na importância da boa educação. E neste sentido, agradeço a minha avó Marlene, tia Bereka e tia Liene por estarem sempre presentes em minha vida acadêmica; aos meus irmãos Daniel, Rubinho e Luciene, pelo apoio e incentivo neste período.

Lyrene e Ricardo, meus orientadores e parceiros nessa longa caminhada, sempre me motivando e traçando as metas para alcançarmos este objetivo.

Lyrene, sou grato pelas longas reuniões de orientação, pelo aprendizado e “rigor” sempre cobrado em busca do melhor. Posso dizer que levo seus ensinamentos comigo.

Ricardo, obrigado pelas inúmeras oportunidades. Tentei não desperdiçar nenhuma e aprendi bastante com seu positivismo e metas “impossíveis” (pois no impossível a concorrência é menor). Sou profundamente grato pela oportunidade de imergir neste trabalho na Universidade de Athabasca (UA) no Canadá.

Ainda no aspecto de orientação desta tese, não posso deixar de agradecer a Thaisa Santos Lima pelas longas reuniões e ensinamentos, enquanto eu ainda estava no Canadá e nas escritas dos artigos. Nossas discussões certamente evoluíram na qualidade dos artigos e desta tese.

Oportunamente, agradeço ao professor Caó e a equipe do Internacional (Carla e Isabelle) que fizeram a ponte entre o LAIS e a Universidade de Athabasca, para que eu pudesse atravessar e escrever parte desta história.

Do outro lado desta ponte, na Universidade de Athabasca, agradeço aos professores Neil Fassina e Shawn Fraser por permitir e facilitar meu acesso e imersão na Universidade; ao professor Vive Kumar pelas reuniões durante minha estadia, onde pude aprender mais sobre análise de dados, bem como aos colegas David Boulanger, e Liliana Quyen Tang que participavam das reuniões e compartilhavam seus conhecimentos comigo.

Durante esta caminhada, reuniões e missões de cooperação mais curtas ocorreram visando aprofundamento das ideias e discussão dos resultados. Cada estudo foi desenvolvido a base de muito esforço e companheirismo, cada um dando seu melhor.

Neste sentido sou grato ao colega Juciano pelos ensinamentos na área da comunicação e as nossas reuniões em busca da análise dos resultados, bem como à colega Ana Cláudia pela proposta das categorias exploratórias utilizadas neste trabalho.

Aos colegas Sidemar e Thiago pelas discussões quase semanais, ainda no período que eu estava no Canadá, a fim de analisarmos as melhores técnicas para processamento dos dados.

Ao amigo Raphael Fontes, pelo companheirismo na missão de Portugal, onde pudemos analisar os dados das notícias online.

Aos professores portugueses Hugo Gonçalo, Paulo Gil, Jorge Henriques e Cezar Teixeira pela recepção na Universidade de Coimbra e discussão dos resultados.

Ao professor Rifat Atun, da Universidade de Harvard, pelas sábias palavras em cada um dos artigos que atuou junto conosco, sempre explicando o seu ponto de vista em termos de análise de impacto e sistemas de saúde resilientes, destacando sempre a importância do Hermes para análise de políticas públicas para o sistema único de saúde (SUS).

Agradeço também aos Professores que aceitaram o convite para compor a banca examinadora desta tese de doutorado: Uirá Kulesza, Plácido Neto, Ximena Pamela, Wagner Martins e Lyane Ramalho. Suas considerações foram enriquecedoras para a evolução do trabalho.

À todos os colegas que prestigiaram minha defesa.

À todos os pesquisadores e colaboradores do LAIS pela presteza nos processos internos.

RESUMO

Promover a conscientização, estimular a adoção de atitudes e comportamentos saudáveis são alguns dos objetivos das intervenções em saúde pública. Porém, para analisar o alcance de uma intervenção é necessário ir além dos dados epidemiológicos, uma vez que este conjunto, por si só, pode não demonstrar a real magnitude dos resultados. É necessário discutir outras fontes de dados, variáveis de interesse e dimensões que podem ser alcançadas pela intervenção. Assim, avaliar o alcance de uma intervenção em saúde pública, em uma perspectiva multidimensional por meio da abordagem de séries temporais, pode ser útil para orientar o desenvolvimento de intervenções mais eficazes na resposta da saúde pública. Nesse contexto, esta tese tem como objetivo desenvolver um *framework* de avaliação multidimensional de intervenções em saúde pública, explorando diferentes variáveis de interesse que são possivelmente impactadas pelas intervenções. Esse *framework* é suportado por um ecossistema de software, denominado Hermes, que é responsável por processar os dados em um ciclo de vida completo e mostrar seus resultados em um painel visual que permita aos tomadores de decisão avaliar o efeito ao longo do tempo antes e depois das intervenções, além de analisar possíveis correlações entre variáveis de interesses. Para compreender o estado atual da arte e orientar a pesquisa neste domínio, realizamos uma revisão sistemática da literatura que explora o uso de abordagens de tecnologia da informação para analisar o impacto de campanhas em saúde pública. Sintetizamos variáveis de interesse, dados de campanha, técnicas e ferramentas empregadas para avaliar intervenções em saúde pública. Em seguida, realizamos três estudos distintos para avaliar uma intervenção em saúde, chamada “Sífilis Não!”, lançada no Brasil. A utilização dos métodos computacionais implementados no ecossistema Hermes, possibilitou a análise de dados de natureza heterogênea para examinar a resposta da saúde pública e seus resultados em todo o país, além de fornecer uma poderosa ferramenta no planejamento e monitoramento da intervenções em saúde, que pode ser transferida para outros desafios da saúde pública. Os estudos mostraram evidências de mudança de comportamento na população considerando aspectos epidemiológicos e de comunicação, além de avaliar a correlação entre as variáveis de interesse levantadas durante esta pesquisa, fortalecendo a necessidade de uma análise multidimensional. Por fim, também analisamos dados epidemiológicos e de comunicação da hepatite no Brasil, a fim de realizar um estudo de caso utilizando o *framework* proposto fora do contexto da sífilis. Os resultados desta tese contribuem para uma análise mais abrangente do alcance das intervenções em saúde pública e, assim, permitem aos formuladores de políticas re-analisar as estratégias de conscientização desenvolvidas para alertar as pessoas sobre os cuidados com a saúde e mudanças comportamentais, bem como melhor direcionar o uso de recursos de forma mais eficaz.

Palavras-chave: framework, análise de séries temporais, intervenção em saúde pública.

ABSTRACT

Promoting awareness, increasing knowledge, and encouraging the adoption of healthy attitudes and behaviors are some of the objectives of public health interventions. However, to analyze the scope of an intervention, it is necessary to go beyond the epidemiological data since this set, by itself, may not demonstrate the absolute magnitude of the results. It is necessary to discuss other data sources, variables of interest, and dimensions that the intervention can achieve. Thus, assessing the scope of a public health intervention from a multidimensional perspective through the time series approach can help guide the development of more effective interventions in the public health response. In this context, this thesis presents a framework for the multidimensional evaluation of public health interventions, exploring variables of interest that are possibly impacted by interventions. This framework is supported by a software called Hermes, responsible for processing the data in a complete lifecycle and showing its results in a visual dashboard that allows decision makers to assess the effect over time before and after campaigns and analyze possible correlations between variables of interest. To understand the current state of the art and guide research in this domain, we conducted a systematic literature review that explores the use of information technology approaches to analyze the impact of public health campaigns. This study summarizes variables of interest, campaign data, techniques, and tools used to evaluate public health interventions. We then conducted three studies to analyze a health intervention launched in Brazil called “Sífilis Não!”. The use of computational methods implemented in Hermes enabled the analysis of data of heterogeneous nature to examine the public health response and its results throughout the country, in addition to providing a powerful tool in planning and monitoring health interventions, which can be transferred to other public health challenges. Furthermore, the studies showed evidence of behavior change in the population considering epidemiological and communication aspects, in addition to evaluating the correlation between the variables of interest raised during this research, strengthening the need for a multidimensional analysis. Finally, we also analyzed epidemiological and communication data on hepatitis in Brazil in order to carry out a case study using the proposed framework outside the context of syphilis. The results of this thesis contribute to a more comprehensive analysis of the scope of public health interventions and thus allow policymakers to re-examine awareness-raising strategies developed to alert people to health care and behavioral changes, as well as better target use resources effectively.

Keywords: framework, time series analysis, public health intervention, public health.

Lista de ilustrações

Figura 1 – Visão geral do <i>framework</i> de análise multidimensional.	20
Figura 2 – Linha do tempo da tese.	21
Figura 3 – Ilustração de análise de séries temporais interrompidas, com mudança de nível de tendência para pontos de dados antes e depois da intervenção.	32
Figura 4 – Framework de Análise Multidimensional.	45
Figura 5 – Diagrama de atividade do <i>framework</i> de análise multidimensional.	46
Figura 6 – Tipos de dados brutos identificados neste estudo.	50
Figura 7 – Arquitetura do ecossistema Hermes em detalhes.	54
Figura 8 – Modelo de dados do Hermes abrangendo as dimensões implementadas.	57
Figura 9 – Interface para análise de notícias no ecossistema Hermes.	59
Figura 10 – Comparação entre notícias identificadas no mesmo cluster originadas de aids.gov.br e replicadas em blog.saude.gov.br.	61
Figura 11 – Comparação entre notícias identificadas no mesmo cluster originadas de aids.gov.br e replicadas em portal.fiocruz.br.	62
Figura 12 – O método de decomposição dividiu os dados em quatro partes: dados originais, tendência, sazonalidade e resíduo aleatório.	64
Figura 13 – Anúncios da Campanha “Sífilis Não!” produzidos e divulgados pela televisão, plataformas de streaming, mídia impressa, revistas e eventos.	69
Figura 14 – Linha do tempo de ações que mostra como as macrocategorias da campanha “Sífilis Não!” foram distribuídas.	73
Figura 15 – Quantidade diária de ações por macrocategoria.	73
Figura 16 – Atividade de pesquisa relativa ao termo ‘sífilis’ no Google entre 2017 e 2019.	74
Figura 17 – Atividade de pesquisa relativa ao termo ‘sífilis’ decomposta em tendência e sazonalidade.	75
Figura 18 – Quantidade de postagens de notícias <i>online</i> sobre sífilis indexadas pelo Google entre janeiro de 2015 e dezembro de 2019.	76
Figura 19 – Notícias <i>online</i> indexadas pelo Google usando o termo ‘sífilis’ decompostas em tendência e sazonalidade.	76
Figura 20 – Número de alunos inscritos nos cursos relacionados à sífilis no AVASUS, entre fevereiro de 2019 e setembro de 2020.	77
Figura 21 – Quantidade de teste de triagem realizado pelo serviço de atenção primária.	79
Figura 22 – Quantidade de testes de triagem decompostos em tendência e sazonalidade.	80
Figura 23 – Número de casos notificados de sífilis adquirida, sífilis em gestante e sífilis congênita por mês/ano decompostos em tendência e sazonalidade.	82

Figura 24 – Série temporal interrompida parametrizada como um modelo de regressão segmentado, incluindo o impacto imediato e a mudança de tendência.	87
Figura 25 – Alterações nas taxas de sífilis congênita na região Norte, 20 meses antes e após a intervenção.	89
Figura 26 – Alterações nas taxas de sífilis congênita na região Nordeste, 20 meses antes e após a intervenção.	91
Figura 27 – Alterações nas taxas de sífilis congênita na região Sul, 20 meses antes e após a intervenção.	92
Figura 28 – Alterações nas taxas de sífilis congênita na região Sudeste, 20 meses antes e após a intervenção.	93
Figura 29 – Alterações nas taxas de sífilis congênita na região Centro-Oeste, 20 meses antes e após a intervenção.	94
Figura 30 – Alterações nas taxas de sífilis congênita no Brasil, 20 meses antes e após a intervenção.	95
Figura 31 – Fluxo de trabalho realizado para categorização de notícias.	101
Figura 32 – Mapa de calor de categorias por trimestre entre 2015 e 2019.	106
Figura 33 – Gráfico de tendência e sazonalidade dos testes de detecção de hepatite viral entre 2015 e 2019.	112
Figura 34 – Gráfico de tendências e sazonalidade de tratamentos para Hepatites Virais entre 2015 e 2019.	113
Figura 35 – Gráfico de tendência e sazonalidade do interesse da população pelo termo 'hepatite' usando o mecanismo de pesquisa do Google.	115
Figura 36 – Articulações Sociais do Projeto “Sífilis Não!” realizado entre março de 2018 e setembro de 2019.	141
Figura 37 – Indicadores da primeira campanha de conscientização do Projeto “Sífilis Não!” no Facebook de fevereiro a novembro de 2018.	142
Figura 38 – Indicadores da segunda campanha de conscientização do Projeto “Sífilis Não!”.	143
Figura 39 – Ações de intervenção realizadas por Apoiadores de Pesquisa e Intervenção, que realizaram trabalhos específicos com gestores locais de saúde.	144
Figura 40 – Indicadores da trilha de treinamento “Sífilis e outras ISTs” que oferece 40 cursos em formato autoinstrucional na plataforma AVASUS.	145

Lista de tabelas

Tabela 1 – Formulário de Extração de Dados.	34
Tabela 2 – Lista de verificação para critérios de avaliação de qualidade.	35
Tabela 3 – Fontes de dados e variáveis de interesse analisadas por estudo.	37
Tabela 4 – Fontes de dados <i>online</i> disponíveis e implementadas no Hermes.	56
Tabela 5 – Exemplo hipotético de conjunto de dados de entrada para processamento de análise de correlação.	65
Tabela 6 – Exemplo hipotético do resultado de processamento da análise de correlação entre variáveis de interesse.	65
Tabela 7 – Dados obtidos pelo ecossistema Hermes.	70
Tabela 8 – Taxa de crescimento de alunos matriculados por semestre letivo.	78
Tabela 9 – Exames realizados por ano e taxa de crescimento por 1.000 habitantes.	79
Tabela 10 – Taxas de incidência de sífilis em gestantes, sífilis congênita e distribuição de penicilina no Brasil, entre 2016 e 2019. Fonte de dados: *SINAN/MoH, **SISMAT/MoH	81
Tabela 11 – Resultados da análise de séries temporais comparando as taxas de sífilis congênita por 1.000 nascidos vivos entre municípios prioritários e não prioritários.	90
Tabela 12 – Resultados da análise de séries temporais comparando as taxas de sífilis congênita por 1.000 nascidos vivos entre municípios prioritários e não prioritários.	91
Tabela 13 – Distribuição quantitativa das categorias exploratórias encontradas na amostra de treinamento a partir dos resultados da busca realizada pelos pesquisadores.	102
Tabela 14 – Páginas da Web categorizadas após a extração do conteúdo do texto principal.	103
Tabela 15 – Principais bigramas encontrados para cada categoria exploratória na amostra de treinamento.	104
Tabela 16 – Distribuição das categorias exploratórias encontradas nas notícias analisadas sobre sífilis por ano, entre 2015 e 2019.	106
Tabela 17 – Número de notícias a cada ano categorizadas por qualidade.	107
Tabela 18 – Correlação entre o número de testes e o número de notícias de acordo com sua qualidade.	107
Tabela 19 – Dados obtidos pelo ecossistema Hermes para avaliar a hepatite no Brasil, ao longo do tempo.	111
Tabela 20 – Número de notícias sobre hepatite indexadas pelo Google entre 2015 e 2019. *Taxa de crescimento em relação ao ano anterior.	114

Tabela 21 – Resultado do processamento de correlação entre variáveis de interesse usando três métodos diferentes.	115
Tabela 22 – Comparação de trabalhos relacionados com esta tese.	122

Abbreviations and Acronyms

Infecções Sexualmente Transmissíveis (ISTs)
Questões de Pesquisa (QP)
Projeto “Sífilis Não!”(PSN)
Ministério da Saúde (MS)
Sistema Único de Saúde (SUS)
Processamento de Linguagem Natural (PLN)
Aprendizado de Máquina (ML)
Séries Temporais Interrompidas (ITS)
Revisão Sistemática de Literatura (SLR)
Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)
Statistical Analysis System (SAS)
ANalysis Of VAriance (ANOVA)
Framework de Análise Multidimensional (FAM)
Cursos Online Abertos Massivos (MOOCs)
Produto Interno Bruto (PIB)
Índice de Desenvolvimento Humano (IDH)
Model-View-Controller (MVC)
Google Custom Search Engine (GCSE)
Sistema de Informações Ambulatoriais do Sistema Único de Saúde (SIA/SUS)
Sistema de Informações Hospitalares do Sistema Único de Saúde (SIH/SUS)
Natural Language Toolkit (NLTK)
Term Frequency – Inverse Document Frequency (TF-IDF)
Density-Based Spatial Clustering of Applications with Noise (DBSCAN)
t-distributed Stochastic Neighbor Embedding (t-SNE)
Latent Semantic Indexing (LSI)
Principal Component Analysis (PCA)
Lexico for Adapted Inference (LeIA)
Valence Aware Dictionary and sEntiment Reasoner (VADER)
Organização Pan-Americana da Saúde (OPAS)
Sistema Integrado de Gerenciamento de Materiais (SISMAT)
Sistema de Informação de Agravos de Notificação (SINAN)
Penicilina G benzatina (BPG)
Atenção Primária à Saúde (APS)

Sumário

1	Introdução	17
1.1	Declaração do Problema	18
1.2	Justificativa	19
1.3	Objetivos	19
1.4	Metodologia	20
1.5	Contribuições	22
1.6	Organização do Documento	26
2	Background e Revisão da Literatura	28
2.1	Campanhas de Saúde Pública para mudar o Comportamento da Saúde	28
2.2	Métodos de Análise Estatística e Aprendizado de Máquina	29
2.3	Análise de Séries Temporais	30
2.4	Revisão da Literatura	31
2.4.1	Metodologia	32
2.4.2	Processo de Pesquisa	33
2.4.3	Extração de Dados	33
2.4.4	Avaliação da Qualidade do Estudo	34
2.4.5	Resultados	36
2.4.6	Conclusões	42
3	Um Framework para Analisar Intervenções em Saúde Pública	44
3.1	Framework de Análise Multidimensional	44
3.2	Ciclo de Vida dos Dados	45
3.3	Dimensões	50
3.3.1	Campanha	50
3.3.2	Comunicação	51
3.3.3	Educação	51

3.3.4	Vigilância Epidemiológica	52
3.3.5	Finanças	52
3.3.6	Sociodemografia	52
4	Hermes - Um Ecossistema de Software para Analisar Intervenções em Saúde	53
4.1	Arquitetura do Ecossistema Hermes	53
4.1.1	Orquestrador	54
4.1.2	Aquisição de Dados	54
4.1.3	Análise de Dados Avançada	57
4.1.3.1	Limpeza e Transformação	58
4.1.3.2	Aprendizado de Máquina	59
4.1.3.3	Análise de Séries Temporais	63
4.1.3.4	Análise de Correlação	64
4.1.4	Interface do Usuário	66
4.2	Conclusões	66
5	Uma Intervenção na Saúde Pública em Resposta à Epidemia de Sífilis	67
5.1	Contextualização	67
5.2	Uma Campanha de Mídia de Massa como Parte da Intervenção de Saúde Pública	68
5.3	Fontes e Dados Heterogêneos	69
6	Estudo Analítico	71
6.1	Análise Experimental	71
6.2	Evidências da Mudança de Comportamento na Saúde	72
6.3	Discussão	83
7	Um Estudo de Séries Temporais para Analisar a Eficácia de uma Intervenção em Saúde Pública	85
7.1	Métodos	85
7.2	Resultados	88
7.3	Discussão	95

8	Notícias Online e suas Influências na Resposta à Saúde Pública . . .	100
8.1	Métodos	101
8.2	Resultados	105
8.3	Discussão	107
9	Estudo sobre Hepatites Virais	110
9.1	Métodos	110
9.2	Resultados	112
9.3	Discussão	114
10	Conclusões	118
10.1	Trabalhos Relacionados	120
10.2	Trabalhos Futuros	122
	Referências	124
	Apêndices	134
	APÊNDICE A Catálogo com dados de campanhas e propriedades extras	135
	APÊNDICE B Intervenções do Projeto “Sífilis Não!” e linha do tempo	141

1 Introdução

As intervenções em saúde estão relacionadas a qualquer atividade realizada com o objetivo de melhorar a saúde humana (i) prevenindo doenças, (ii) curando ou reduzindo a gravidade ou duração de uma doença existente ou (iii) restaurando a função perdida por doença ou lesão (SMITH; MORROW; ROSS, 2015). Assim, as intervenções podem ser classificadas em duas grandes categorias: preventivas e terapêuticas.

As intervenções preventivas estão relacionadas a vacinas, intervenções nutricionais, intervenções maternas e neonatais, educação e mudança de comportamento, alterações ambientais e medicamentos para prevenção de doenças. Enquanto as intervenções terapêuticas estão relacionadas ao tratamento de doenças infecciosas, diagnósticos para orientar a terapia e controle de doenças crônicas (SMITH; MORROW; ROSS, 2015). Por exemplo, as intervenções em saúde pública são essenciais para combater as Infecções Sexualmente Transmissíveis (ISTs) com foco na prevenção, tratamento e conhecimento sobre a doença e formas de transmissão (CRAIG et al., 2017).

As campanhas de saúde pública dos últimos anos têm sido utilizadas para conscientizar a população, aumentar o conhecimento e incentivar a adoção de hábitos saudáveis (DORFMAN; ERVICE; WOODRUFF, 2002; WAKEFIELD; LOKEN; HORNIK, 2010). Para o desenvolvimento deste trabalho, as campanhas de saúde pública foram exploradas na literatura e utilizadas como estudo de caso, uma vez que atuam como um conjunto de intervenções preventivas e terapêuticas coordenadas que objetivam a mudança de comportamento da população.

As campanhas tradicionais de saúde têm transmitido mensagens através de diversos canais para atingir grandes públicos, como televisão, rádio, mídia impressa, revistas, transmissões ao vivo durante eventos, cartazes, cartilhas informativas, adesivos, redes sociais e plataformas de streaming.

Os estudos sobre os efeitos das campanhas de saúde pública consideram principalmente variáveis relacionadas aos campos epidemiológicos, como teste de triagem (NADARZYNSKI et al., 2019; DOWSHEN et al., 2015), casos de notificação e compra de medicamentos (WAKEFIELD et al., 2008), pois aparentam ser mais realistas para análise de impacto de uma intervenção de saúde bem sucedida.

No entanto, outras variáveis de interesse podem ser analisadas para medir o alcance da campanha em outras dimensões, por exemplo: Comunicação (e.g., notícias online, interesse em um determinado tópico em sites de pesquisa e redes sociais), Educação (e.g., o surgimento de cursos e publicações científicas sobre o tema) e Finanças (por exemplo, custos de internação e tratamento).

A escolha das variáveis de interesse depende do papel desempenhado pelas partes interessadas (*stakeholders*). Por exemplo, enquanto um secretário de saúde pode se interes-

sar por variáveis epidemiológicas, um secretário de educação tem sua atenção voltada para variáveis educacionais. Por sua vez, os gestores de campanhas públicas podem se interessar pelas variáveis de comunicação, uma vez que a *Internet* pode proporcionar discussão e engajamento gratuitos sobre o tema levantado pela campanha.

1.1 Declaração do Problema

Para analisar o alcance de uma campanha de saúde pública, é necessário ir além dos dados epidemiológicos, pois esse conjunto, por si só, pode não demonstrar a real magnitude dos resultados. Assim, é necessário discutir outras fontes de dados, variáveis de interesse e as múltiplas dimensões que uma campanha pode atingir.

Além disso, existem desafios metodológicos na medição de resultados e no estabelecimento de uma relação causal entre dimensões distintas e os períodos de intervenção, incluindo desenhos de pesquisa, medição de qualidade e avaliação de resultados (ØVRETT-VEIT; GUSTAFSON, 2003). Métodos inadequados para analisar dados de séries temporais levam a resultados imprecisos e conclusões errôneas ou enganosas (EWUSIE et al., 2020).

A dimensão da Educação, por exemplo, possui uma gama de indivíduos que podem ter sido impactados pela campanha e levam, além da conscientização, o repasse de informações por meio de pesquisas e publicações alcançando outros indivíduos (professores e alunos, profissionais de saúde, por exemplo). Novos cursos podem ser oferecidos e transformar esse ambiente acadêmico/científico. Além disso, com conhecimento mais aprofundado, os profissionais de saúde podem aplicar seus conhecimentos à população e ao ambiente de trabalho.

Por outro lado, a dimensão da comunicação, por exemplo, possui uma rica fonte de dados em conteúdo gerado pela Internet. Redes sociais, registros de atividades em sites de pesquisa e notícias *online* espontâneas são variáveis de interesse que devem ser analisadas em conjunto com as de outras dimensões para verificar possíveis correlações.

Nossa hipótese é poder avaliar muitas variáveis de interesse para fornecer insights a diferentes *stakeholders*, permitindo que eles aprendam continuamente como fazer melhor seu trabalho. Compreender o alcance de uma campanha pode ajudar a direcionar melhor o uso de recursos e esforços de forma mais eficaz.

Assim, esta tese apresenta um *framework* para a avaliação multidimensional do alcance de intervenções em saúde pública, propondo e explorando variáveis de interesse não usuais que possivelmente são impactadas pelas campanhas. Essa estrutura é suportada por um ecossistema de software chamado Hermes, responsável por processar os dados em um ciclo de vida completo dos dados e mostrar seus resultados em um painel visual permitindo que os tomadores de decisão avaliem o efeito ao longo do tempo antes e depois das campanhas.

1.2 Justificativa

Revisamos a literatura que explora o uso da tecnologia da informação para analisar o impacto das campanhas de saúde pública. Este estudo está detalhado na Seção 2.4 e foi publicado no periódico *Frontiers in Public Health* (PINTO et al., 2021). A análise concentrou-se na identificação de variáveis de interesse, técnicas e ferramentas. Observamos que os estudos buscaram avaliar o impacto das campanhas por meio de análises quantitativas e qualitativas, utilizando softwares estatísticos para explorar, visualizar, modelar e fazer inferências nos dados.

De acordo com esta pesquisa, dentre as variáveis de interesse mais exploradas podemos citar: (i) a quantidade de acessos ao site da campanha, (ii) conscientização e conhecimento do assunto por meio de questionários, (iii) interesse da população por meio de sites de busca, (iv) engajamento através de redes sociais, e (v) dados de exposição publicitária em comerciais de televisão.

No entanto, essas variáveis de interesse geralmente são avaliadas individualmente. Não identificamos nos estudos uma abordagem que possa ser reaplicada para diferentes campanhas utilizando dados heterogêneos de diferentes fontes e dimensões. A análise de séries temporais para diferentes dimensões ajudaria a entender o alcance da campanha e permitiria a análise da correlação entre as variáveis de interesse.

1.3 Objetivos

O principal objetivo desta tese é desenvolver um *framework* para análise multidimensional do alcance de intervenções em saúde pública, genérico o suficiente para conectar dimensões inesperadas e variáveis de interesse.

Nossa hipótese: Uma análise multidimensional das intervenções em saúde pública é necessária para entender a relevância das variáveis de interesse ao longo do tempo e suas correlações, a fim de permitir que os tomadores de decisão reavaliem as ações de resposta a crises de saúde pública ao projetar intervenções futuras.

A Figura 1 ilustra uma visão geral do *framework* de análise multidimensional. Primeiramente, precisamos verificar quais fontes de dados estão disponíveis para serem analisadas (A). Em seguida, um conjunto de ações (B) move os dados por meio de cinco etapas que formam o ciclo de vida dos dados: Adquirir, Limpar, Usar, Publicar e Preservar. Essas ações transformam os dados brutos em informações por meio de representações gráficas, disponibilizadas aos tomadores de decisão (C), que podem se interessar em analisar o alcance da intervenção por meio de diferentes pontos de vista, como secretários de saúde, secretários de educação, formuladores de políticas públicas e profissionais de marketing. Por fim, os dados transformados são preservados em bases de dados específicas representadas pelas dimensões de interesse (D), por exemplo, Campanha, Comunicação,

Educação e Vigilância Epidemiológica.

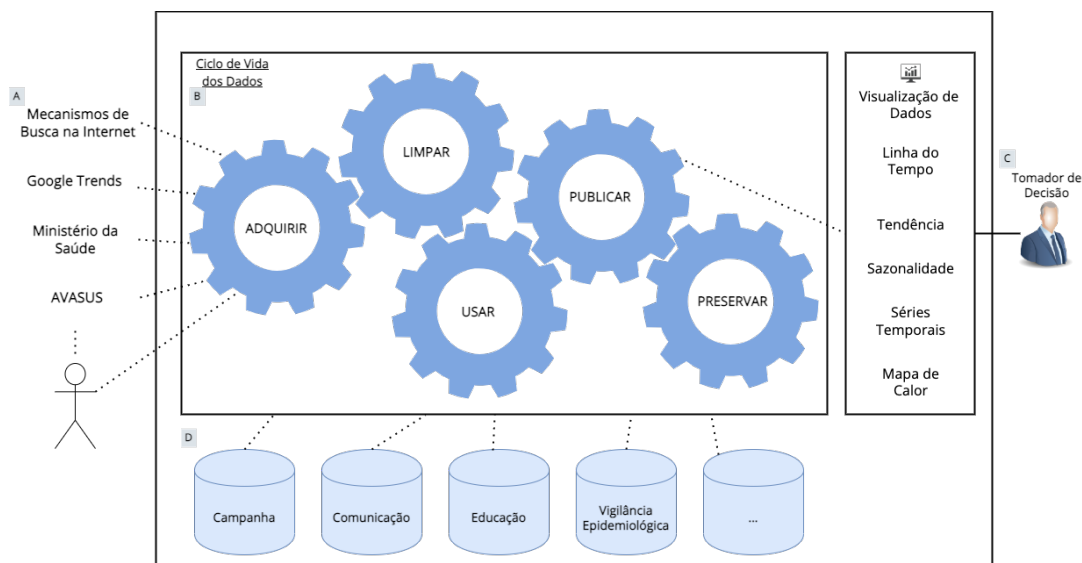


Figura 1 – Visão geral do *framework* de análise multidimensional.

Assim, a definição do *framework* foi orientada pelas seguintes Questões de Pesquisa (QP):

- QP1: Quais dimensões e variáveis de interesse são analisadas para avaliar o impacto das campanhas de saúde pública?
- QP2: Quais técnicas ou ferramentas são utilizadas para apoiar a análise do impacto da campanha?
- QP3: Como apoiar a análise do alcance de uma campanha de saúde pública baseada em aspectos multidimensionais?

1.4 Metodologia

Esta pesquisa aplicada visa desenvolver um *framework* conceitual e um sistema de informação para avançar no estado da arte em avaliações de campanhas de saúde pública. A abordagem de avaliação central é baseada em métodos quantitativos que seguem um fluxo de trabalho analítico de descoberta de informações a partir dos dados.

Esta pesquisa faz parte de uma intervenção nacional de combate à sífilis no Brasil por meio da “Pesquisa Aplicada para Integração Inteligente Orientada ao Fortalecimento das Redes de Atenção para Resposta Rápida à Sífilis”, conhecido como Projeto “Sífilis Não!”(PSN), lançado em 2017 pelo Ministério da Saúde (MS). Este Projeto teve abrangência nacional e foi operacionalizado por eixos de resposta à sífilis, de forma interfederativa, no Sistema Único de Saúde (SUS). O eixo da Comunicação teve como objetivo disseminar o tema sífilis em todo o país (VALENTIM et al., 2021). Em 2018, foi lançada uma campanha

nacional de comunicação cujo tema foi: “Teste, Trate e Cure”. A campanha, mais conhecida como campanha “Sífilis Não!”, teve como objetivo lembrar as pessoas sobre a importância da prevenção da sífilis.

Assim, os estudos de caso desta tese são baseados na análise de dados heterogêneos cujo ponto de interseção é o Projeto “Sífilis Não!”. Para responder às questões de pesquisa e atender aos objetivos definidos para esta tese, revisamos a literatura e realizamos quatro estudos operacionais. A Figura 2 apresenta a linha do tempo percorrida durante estes estudos.

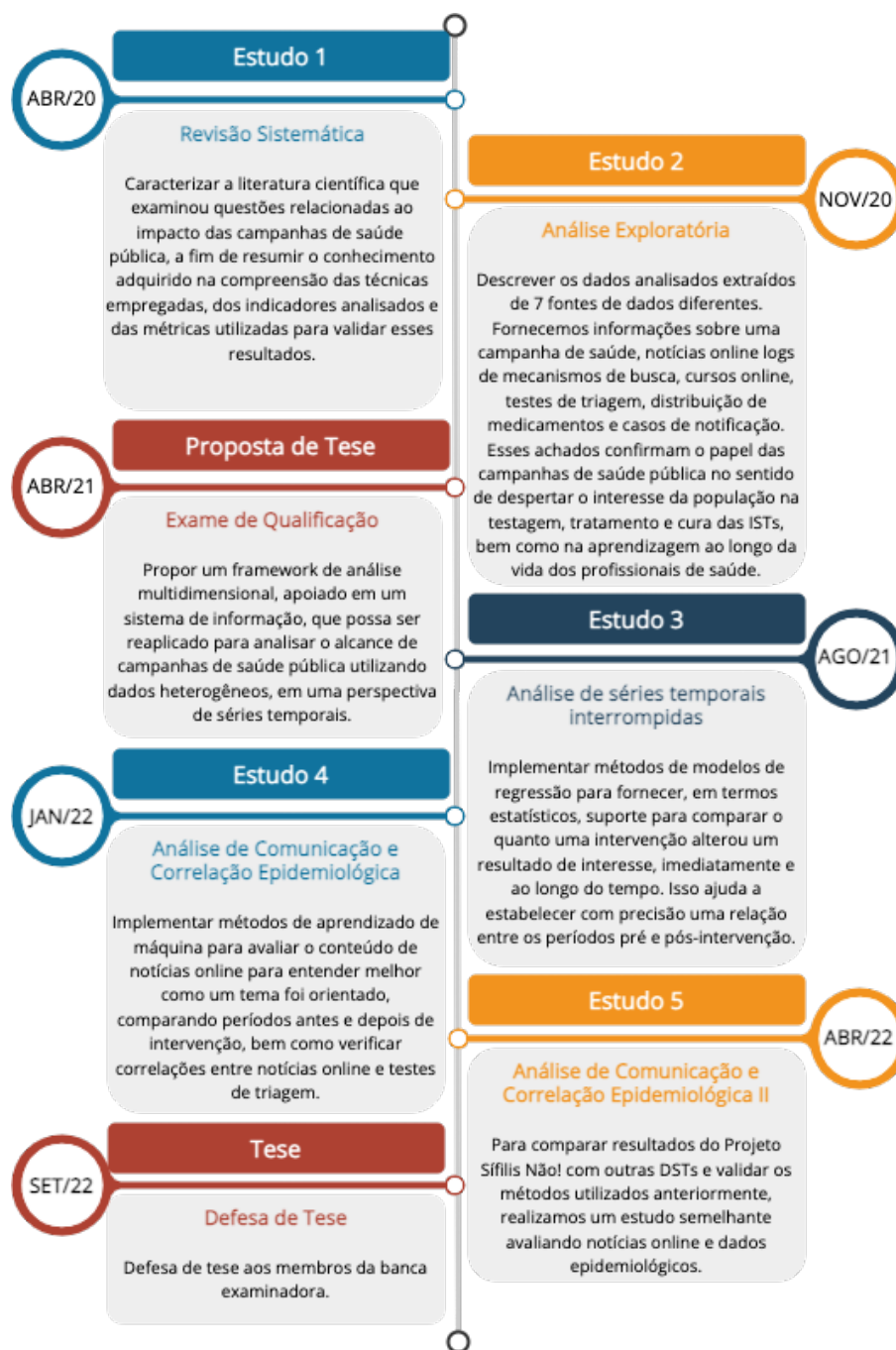


Figura 2 – Linha do tempo da tese.

Primeiramente, realizamos uma revisão sistemática da literatura (Estudo 1) para fornecer uma análise das abordagens de tecnologia da informação que avaliam o impacto das campanhas de saúde pública. A análise concentrou-se na identificação de variáveis de interesse, técnicas e ferramentas. Ajudou a compreender o estado da arte atual para orientar a pesquisa neste domínio. As QP1 e QP2 foram respondidas por meio deste estudo.

A QP3 foi respondida através do Estudo 2. Trata-se de um estudo exploratório baseado na intervenção da mídia de massa que conseguiu discutir quatro dimensões distintas (campanha, educação, comunicação e vigilância epidemiológica) para analisar seu alcance. Este estudo nos ajudou a validar preliminarmente o *framework* proposto. Além disso, desenvolvemos o ecossistema Hermes para dar suporte ao ciclo de vida dos dados e usar algoritmos de aprendizado de máquina para análise de séries temporais.

Os estudos 1 e 2 foram realizados durante a cooperação internacional na Universidade de Athabasca (Canadá). Essa imersão técnico-científica ajudou a impulsionar o conhecimento em mineração e análise de dados relacionados à intervenção em saúde envolvendo os campos da comunicação e da saúde.

O próximo passo foi apresentar a proposta de tese à banca de qualificação, que ocorreu em abril de 2021. Então, conforme planejado, aprofundamos nossa pesquisa avaliando séries temporais (Estudo 3), processamento de linguagem natural e métodos de aprendizado de máquina (Estudo 4) para validar o *framework* e entender como as variáveis de interesse podem estar correlacionadas.

Finalmente, realizamos um estudo avaliando notícias *online* e dados epidemiológicos sobre Hepatite (Estudo 5) para comparar resultados do Projeto “Sífilis Não!” com outras Doenças Sexualmente Transmissíveis e validar os métodos utilizados anteriormente.

Esses estudos foram essenciais para fornecer, em termos estatísticos, subsídios para comparar o quanto uma intervenção alterou uma variável de interesse imediatamente e ao longo do tempo.

Na próxima seção, são discutidas as contribuições obtidas em cada estudo realizado.

1.5 Contribuições

Os resultados desta pesquisa foram publicados em 4 artigos científicos, além dos outros três em revisão em periódicos internacionais. Nos capítulos seguintes, detalhamos esses artigos e os periódicos onde foram publicados. Além disso, as seguintes contribuições desta tese podem ser destacadas:

- Descrevemos variáveis de interesse, dimensões e ferramentas utilizadas em trabalhos que utilizam a tecnologia da informação para avaliar campanhas de saúde pública, apresentando uma Revisão Sistemática da Literatura;

- Disponibilizamos um catálogo com dados de campanhas e propriedades extras que não constavam na taxonomia encontrada na literatura;
- Fornecemos uma lista para verificação de critérios de avaliação de qualidade para melhor identificar a completude nos dados de campanha publicados em artigos;
- Definimos um modelo de dados para organizar elementos de dimensões distintas e padronizar como eles se relacionam em uma perspectiva de séries temporais. Neste modelo de dados, consideramos dados relacionados a campanhas, educação, interesse ao longo do tempo na *Internet*, notícias *online* e procedimentos em saúde;
- Compartilhamos dados coletados de diferentes plataformas, para validar o *framework*, em repositórios públicos. Esses dados permitem que outros pesquisadores usem e comparem com outras abordagens;
- Os dados da Campanha “Sífilis Não!” foram extraídos de planilhas e documentos oficiais, preservados e compartilhados, permitindo que outros pesquisadores utilizem e comparem com outras abordagens;
- Definimos um modelo de classificação para avaliar a qualidade das notícias *online* relacionadas à sífilis além de validar a correlação estatística com testes sorológicos;
- Identificamos mudanças nas tendências da sífilis congênita através de um estudo de caso usando algoritmos de Regressão Linear;
- Extraímos sentimentos de notícias *online* e identificamos correlação entre conteúdo indutor de atitude e testes de triagem;
- Testamos a correlação estatística entre diferentes variáveis de interesse:
 1. Google Trends vs Dados da Campanha: Correlação não encontrada;
 2. Dados da Campanha vs Notícias Online: Correlação não encontrada;
 3. Dados de Campanha vs Testes de Triagem: Correlação não encontrada;
 4. Google Trends vs Notícias Online: Correlação não encontrada;
 5. Google Trends vs Testes de Triagem: Correlação encontrada;
 6. Notícias Online vs Testes de Triagem: Correlação encontrada;
- Um ecossistema de software (Hermes) foi desenvolvido usando algoritmos de Processamento de Linguagem Natural (PLN) e Aprendizado de Máquina (ML) para automatizar tarefas e processos definidos pelo *framework*;

Destacamos as contribuições desta pesquisa agrupando por estudos, como segue:

Estudo 1: Systematic Review on Information Technology Approaches to Evaluate the Impact of Public Health Campaigns: Real Cases and Possible Directions. Publicado na *Frontiers in Public Health*, 2022.

Neste estudo, revisamos a literatura que explora o uso de abordagens de tecnologia da informação para analisar o impacto de campanhas de saúde pública. A análise concentrou-se na identificação de variáveis de interesse, técnicas e ferramentas. Começamos a revisão com 276 artigos e os reduzimos para 17 dos mais relevantes.

Como contribuições para este estudo, podemos citar (i) ferramentas, técnicas, fontes de dados e variáveis de interesse sintetizadas; (ii) fornecemos um catálogo com dados reais de campanhas e propriedades extras que não foram relatadas na taxonomia encontrada na literatura, resumidas com as principais informações: nome da campanha, tipo de anúncio, área do tópico, localização do país, público-alvo, período, nível de organização, lançada por, e fontes de dados exploradas; e (iii) fornecemos insights sobre possíveis indicadores negligenciados na avaliação de impacto das campanhas.

É importante ressaltar que não identificamos uma abordagem suportada por um sistema de informação que possa ser reaplicado para diferentes campanhas usando dados heterogêneos nos estudos.

Estudo 2: Analyzing the Reach of Public Health Campaigns Based on Multidimensional Aspects: the case of the syphilis epidemic in Brazil. Publicado na *BMC Public Health*, Springer, 2021.

O objetivo deste estudo foi discutir o quão expansivo pode ser o alcance de uma campanha de saúde pública. Analisamos 7 fontes de dados de diferentes perspectivas para avaliar uma campanha de saúde do mundo real chamada “Sífilis Não!” lançada no Brasil entre novembro de 2018 e maio de 2019. Primeiro, apresentamos um *framework* de análise multidimensional e implementamos um ecossistema de software (Hermes) para processar esses dados em uma perspectiva de série temporal avaliando o efeito ao longo do tempo antes e depois da campanha. Em seguida, analisamos e descrevemos dados relacionados à campanha, notícias online, logs de atividade de mecanismos de busca na *Internet*, cursos online, testes de triagem, distribuição de medicamentos e casos de notificação.

O estudo 2 mostrou que o *framework* pode permitir uma avaliação mais abrangente do alcance das campanhas de saúde pública e, assim, permitir que os formuladores de políticas re-analisem as estratégias de conscientização desenvolvidas para alertar as pessoas sobre cuidados de saúde e mudanças comportamentais.

Como contribuições deste estudo, podemos citar o catálogo de dados, estruturado e extraído de relatórios e documentos oficiais da campanha “Sífilis Não!” e inserção no ecossistema Hermes, permitindo o processamento e análise desses dados. Além disso, as

outras 7 fontes de dados obtidas e processadas pelo Hermes estão disponíveis para outros pesquisadores explorarem.

Além disso, discutimos os resultados analisados e a influência positiva do Projeto “Sífilis Não!” como indutor da resposta da saúde pública no combate à sífilis no Brasil, pois conseguiu implementar a agenda de testagem, tratamento e cura.

Estudo 3: Use of Interrupted Time Series Analysis in Understanding the Course of the Congenital Syphilis Epidemic in Brazil. Publicado na The Lancet Regional Health – Americas, 2022.

O Estudo 3 foi fundamental para confirmar o impacto da intervenção em saúde pública produzida pelo Projeto “Sífilis Não!” por meio da dimensão Vigilância Epidemiológica. Portanto, um estudo quase experimental usando análise de Séries Temporais Interrompidas (ITS) foi realizado para avaliar impactos imediatos e mudanças na tendência da sífilis congênita nacional antes e depois do Projeto a partir de setembro de 2016 até dezembro de 2019. Os dados foram avaliados considerando as taxas de sífilis congênita por 1.000 nascidos vivos em todos os municípios prioritários (n=100) abrangidos pelo Projeto e em municípios não prioritários (n=5.470) das cinco macrorregiões do Brasil.

Os municípios prioritários apresentaram uma redução mais significativa (mudança de tendência) do que os não prioritários. O modelo de regressão linear revelou mudanças de tendência após a intervenção, com ambos os grupos de municípios apresentando queda na média mensal de casos por 1.000 nascidos vivos, com redução de -0,21 (IC 95% -0,33 a -0,09; $\rho=0,0011$) nos municípios prioritários e de -0,10 (IC 95% -0,19 a -0,02; $\rho=0,0216$) nos municípios não prioritários.

Neste estudo, usando ITS, discutimos as evidências importantes sobre a direção, o momento e a magnitude dos efeitos das intervenções introduzidas como parte do PSN na sífilis congênita no Brasil. Os resultados sugerem que o Projeto “Sífilis Não!” influenciou as tendências da sífilis congênita no Brasil a partir de 2018, com maiores reduções alcançadas nos municípios prioritários.

É importante ressaltar que este artigo foi na capa da revista The Lancet na edição de março¹, e uma publicação editorial intitulada “Eliminating congenital syphilis: time to act” (Eliminando a sífilis congênita: hora de agir) destacou os resultados do estudo.

Estudo 4: How Online News Affects Public Health Response to Syphilis in Brazil: Text Mining Mediated Analysis.

O estudo 4 é uma análise aprofundada das notícias *online* e como essas notícias afetam a resposta da saúde pública no Brasil. O ecossistema Hermes automatizou a

¹ <[https://www.thelancet.com/journals/lanam/issue/vol7nonull/PIIS2667-193X\(21\)X0008-7](https://www.thelancet.com/journals/lanam/issue/vol7nonull/PIIS2667-193X(21)X0008-7)>

aquisição e processamento de notícias online. Os dados foram coletados de 2015 a 2019 usando a API Google Search, procurando páginas da web que contenham o termo 'sífilis'.

Este estudo analisou o volume e a qualidade do conteúdo de notícias *online* sobre sífilis no Brasil e examinou seu efeito sobre o número de testes de triagem ao longo do tempo. Aplicamos técnicas de PLN para extrair padrões de categorias de notícias online. A presença e a combinação de tais categorias nos textos determinaram a qualidade das notícias, classificadas como notícias de alta, média e baixa qualidade. Além disso, foi examinada a correlação entre a qualidade das notícias e o volume de testes de sífilis usando o Coeficiente de Correlação de Spearman.

Os achados demonstram que estratégias efetivas de comunicação que levam à disseminação de informações de alta qualidade são essenciais para aumentar a adesão às ações de políticas públicas de saúde.

Estudo 5: Um Estudo de Caso Fora do Contexto da Epidemia de Sífilis no Brasil.

Neste estudo, aplicamos as lições aprendidas em estudos anteriores para validar as técnicas e métodos utilizados anteriormente em um contexto fora da epidemia de sífilis no Brasil. Escolhemos a hepatite porque, além de ser uma doença sexualmente transmissível, obtivemos dados epidemiológicos (exames de triagem e tratamento) e de comunicação (registros de atividade de mecanismos de busca e notícias online) dessa doença. O objetivo deste estudo foi observar possíveis mudanças nas tendências e correlações com variáveis de interesse em diferentes dimensões.

Os resultados mostraram estabilidade no tratamento da hepatite, aliada à queda no número de exames. Além disso, observamos um crescimento ano a ano no volume de notícias online. No entanto, como não identificamos intervenções públicas e nacionais no combate a essa doença, esse volume crescente e espontâneo mostrou-se muito pequeno em relação às notícias sobre a sífilis. Os dados do Google Trends mostraram picos e declínios ao longo dos anos.

Por fim, não foi encontrada correlação estatisticamente significativa entre as variáveis de interesse. Isso possivelmente indica que outra variável impulsionou o declínio nos testes e estabilização do tratamento. Nossa hipótese é que o aumento das vacinas disponibilizadas pelo SUS tenha modificado esses resultados. No entanto, não obtivemos esses dados para análise.

1.6 Organização do Documento

O restante deste documento está organizado da seguinte forma.

- O capítulo 2 apresenta os principais conceitos, técnicas e ferramentas necessárias para a compreensão desta tese e os resultados alcançados através da Revisão Sistemática da Literatura.
- O capítulo 3 apresenta o *framework* para a avaliação multidimensional de intervenções em saúde pública. Além disso, detalhamos seus objetivos, atividades, arquitetura, variáveis de interesse e dimensões.
- O capítulo 4 mostra o ecossistema Hermes, que é responsável por suportar e validar o *framework* e coletar e processar dados de fontes heterogêneas.
- O capítulo 5 apresenta a intervenção de saúde pública contra a sífilis no Brasil; a campanha de comunicação de massa lançada para conscientizar a população sobre a epidemia de sífilis e as fontes de dados utilizadas pelos estudos de caso com o objetivo de avaliar a resposta de saúde pública.
- O capítulo 6 apresenta os resultados obtidos pela análise da campanha “Sífilis Não!”, considerando quatro dimensões: campanha, comunicação, educação e vigilância epidemiológica.
- O capítulo 7 apresenta um estudo de série temporal realizado com o objetivo de estimar os efeitos da intervenção ao longo do tempo comparando as taxas de sífilis congênita por 1.000 nascidos vivos nos municípios prioritários cobertos pelo PSN e nos municípios não prioritários nas cinco macro-regiões do Brasil: Norte, Nordeste, Sul, Sudeste e Centro-Oeste.
- O capítulo 8 descreve um estudo de caso explorando notícias *online* para entender essa variável de interesse na dimensão da comunicação. O estudo de caso foi realizado em busca de evidências para avaliar as correlações entre notícias *online* e testes de triagem, destacando a importância desse tipo de mídia para a conscientização da população sobre a mudança de hábitos, em busca de exames, tratamentos e curas de doenças infecciosas.
- O capítulo 9 apresenta o estudo sobre hepatite para demonstrar a eficácia do *framework* proposto, utilizando um estudo de caso fora do escopo do projeto “Sífilis Não!”.
- O capítulo 10 apresentamos as conclusões, trabalhos relacionados e trabalhos futuros.

2 Background e Revisão da Literatura

Este capítulo apresenta o background e conceitos essenciais usados no escopo da tese sobre campanhas de saúde pública, análise estatística, métodos de aprendizado de máquina, análise de séries temporais, bem como técnicas e ferramentas atualmente usadas para avaliar campanhas de saúde. Apresentamos também os resultados alcançados por meio da Revisão Sistemática da Literatura para analisar abordagens de tecnologia da informação que avaliam o impacto das campanhas de comunicação em saúde pública como método estratégico de promoção da saúde.

2.1 Campanhas de Saúde Pública para mudar o Comportamento da Saúde

As campanhas de saúde pública são amplamente utilizadas para promover a conscientização, aumentar o conhecimento e incentivar a adoção de atitudes e comportamentos saudáveis (DORFMAN; ERVICE; WOODRUFF, 2002; BUCCHI; TRENCH, 2014).

Tais campanhas têm como principal objetivo conscientizar a população sobre cuidados com a saúde sexual (GABARRON et al., 2019; NADARZYNSKI et al., 2019; BULL et al., 2012; SYRED et al., 2014; DOWSHEN et al., 2015; AHMED et al., 2013; GAMAGE et al., 2011), percepção de risco ao tabagismo (ZHAN et al., 2017; WAKEFIELD et al., 2008; GRANDPRE et al., 2003; EMERY et al., 2014; ALLEM et al., 2017), e sobrepeso e obesidade (KITE et al., 2018b; O'HARA et al., 2016; KITE et al., 2018a). Além de outros como vacinação infantil (LAMPOS et al., 2015) e triagem neonatal & programas de biobanco (PLATT et al., 2016).

Para entender melhor as características de uma campanha de saúde pública, Dorfman, Ervice e Woodruff (2002) propôs uma taxonomia de campanhas de comunicação, que inclui três eixos: propósito, escopo e maturidade. Por propósito, os autores entendem que o objetivo de uma campanha deve afetar diretamente os indivíduos, ou as políticas coletivas, que moldam o comportamento da sociedade. Escopo refere-se à parte mais visível de uma campanha de comunicação: seu tamanho e extensão (Em que região? Em que período? É local, estadual ou nacional?). A maturidade pode vir com o tempo (as campanhas podem ser eventos únicos ou podem durar anos). Pode se tornar mais formal, com objetivos mais explícitos, materiais bem desenvolvidos e integração mais profunda nas atividades gerais de uma organização.

2.2 Métodos de Análise Estatística e Aprendizado de Máquina

A evidência estatística para avaliar se uma variável de interesse foi alterada ao longo do tempo é fundamental para tirar conclusões formais corretas sobre o impacto de uma intervenção. Nesta tese, usamos métodos de aprendizado de máquina e processamento de linguagem natural baseados em modelos estatísticos para melhorar a qualidade dos resultados, garantindo evidências sobre se a variável de interesse representa uma mudança real, evitando apenas uma representação gráfica.

Os seguintes métodos e técnicas foram usados nesta tese e serão discutidos em seus respectivos capítulos.

- **Decomposição Clássica:** Esta técnica é útil para decompor dados de séries temporais sobrepostas, discriminando e separando componentes individuais como tendências, sazonalidade e resíduos aleatórios (PALIT; POPOVIC, 2006).
 - Esta técnica foi utilizada para decompor grandezas medidas relacionadas a variáveis de interesse ao longo do tempo, permitindo uma análise de mudança de tendências e avaliação de padrões cíclicos de sazonalidade.
- **TF-IDF (Term Frequency-Inverse Document Frequency):** É uma medida estatística usada na recuperação de informações e aprendizado de máquina que pode determinar a importância de representações de *strings* (por exemplo, palavras e frases) em um documento entre uma coleção de documentos (SALTON; BUCKLEY, 1988).
 - Esta técnica foi utilizada para identificar parâmetros de qualidade (definidos como categorias exploratórias) de notícias online. A presença e a combinação de tais categorias em uma única notícia definem sua qualidade da seguinte forma: baixa qualidade (0–2 categorias), média qualidade (3–4) e alta qualidade (5 ou mais).
- **Análise de Sentimentos:** Pode ser definida como a tarefa de extrair informações subjetivas sobre sentimentos de texto, fala, tweets e fontes de banco de dados por meio do Processamento de Linguagem Natural (PLN). A análise de sentimento envolve a classificação de opiniões no texto em categorias como “positivo”, “negativo” ou “neutro” (KHARDE; SONAWANE et al., 2016)
 - Esta técnica foi utilizada para avaliar a influência dos aspectos sentimentais das notícias *online* para explorar suas possíveis relações com outras variáveis de interesse.

- Agrupamento de Textos: Pode ser definido como a tarefa de agrupar documentos semelhantes, marcando os textos discrepantes fora do grupo (BLOKH; ALEXANDROV, 2017).
 - Esta técnica foi usada para identificar quais notícias foram as mais disseminadas ou que serviram de base para a produção de novas notícias, ou seja, notícias que efetivamente provocaram o efeito “intermedia agenda-setting” (MILONE, 2012; ARDÈVOL-ABREU; ZÚÑIGA; MCCOMBS, 2020).
- Coeficiente de Correlação de Spearman, Pearson e Kendall: São técnicas estatísticas usadas para medir a associação ordinal entre duas grandezas de medidas (HAUKE; KOSSOWSKI, 2011).
 - Esta técnica foi utilizada para medir a correlação entre variáveis de interesse ao longo do tempo identificadas em nossos estudos operacionais.
- Análise de Regressão Segmentada: É um método para estimar estatisticamente o quanto uma intervenção afeta a variável de interesse medida imediatamente e ao longo do tempo (WAGNER et al., 2002).
 - Esta técnica foi utilizada para estimar a mudança nas tendências e o impacto imediato da intervenção “Sífilis Não!” em municípios prioritários e não prioritários.

2.3 Análise de Séries Temporais

Uma Série Temporal é uma sequência ordenada de pontos de dados coletados regularmente em intervalos de tempo. A análise de uma determinada série temporal visa principalmente estudar seus componentes para entender melhor o processo pelo qual os dados da série temporal são gerados (LUO; VIDAL; ACHO, 2014). Os componentes significativos das séries temporais são estacionariedade, linearidade, tendência e sazonalidade. Abaixo apresentamos suas definições.

- **Estacionariedade** pode ser definido como uma propriedade estatística de um processo gerando uma série temporal que não muda ao longo do tempo. Não significa dizer que a série não muda ao longo do tempo, apenas que a forma como ela muda não muda ao longo do tempo (PALACHY, 2019).
- **Linearidade** de uma série temporal mostra que a forma da série depende de seu estado. Em outras palavras, determina o padrão da série temporal. Se uma série temporal for linear, ela pode ser representada por uma função linear dos valores presentes e passados. As séries temporais não lineares podem ser representadas pelos modelos não lineares ou bilineares correspondentes (LUO; VIDAL; ACHO, 2014).

- **Tendência**, como o próprio nome diz, indica a tendência geral dos dados para diminuir ou aumentar em uma direção específica durante um período (BOYD, 1998). Segundo Wu et al. (2007) a tendência mais comumente vista é a tendência simples, que é uma linha reta ajustada aos dados.
- **Sazonalidade** de uma série temporal é demonstrado através do padrão flutuante periódico, repetitivo e previsível nos dados (TURNER et al., 2020).

Para avaliar os efeitos ao longo do tempo de uma intervenção que ocorreu em um período delimitado no tempo, a Série Temporal Interrompida (ITS) é considerada um desenho de estudo quase-experimental mais robusto. As intervenções podem ser definidas como atividades que previnem doenças ou que promovem a saúde em um grupo de pessoas e podem ser caracterizadas pelo que foi feito, como foi realizado, a quem foi direcionado e onde foi realizado (BRISS et al., 2000).

A análise de Regressão Segmentada é um método para modelar estatisticamente os dados de séries temporais interrompidas, permitindo avaliar o quanto uma intervenção alterou uma variável de interesse imediatamente e ao longo do tempo (WAGNER et al., 2002; ANSARI et al., 2003). As variáveis de interesse podem ser expressas como contagens, médias, proporções ou taxas. Uma análise de regressão segmentada divide o período de tempo em segmentos pré e pós-intervenção. As interceptações e inclinações separadas são estimadas em cada segmento. Por fim, são realizados testes estatísticos de mudanças em interceptos e inclinações pré-pós-intervenção (TALJAARD et al., 2014).

A Figura 3 ilustra uma análise de série temporal interrompida, com uma mudança no nível dos pontos de dados antes e depois da interceptação. Isso constituiria um efeito de intervenção. O eixo horizontal representa o tempo decorrido, enquanto o eixo vertical apresenta a variável de interesse. A linha azul mostra os dados coletados ao longo do tempo e a linha pontilhada verde mostra a intervenção que ocorreu em um momento específico. Após a intervenção, podemos observar uma mudança no nível da variável de interesse. Além disso, é possível observar a tendência da variável de interesse nas linhas pontilhadas vermelhas nos períodos pré e pós-intervenção.

2.4 Revisão da Literatura

Realizamos uma revisão sistemática para examinar a literatura científica que explora o uso de abordagens de tecnologia da informação para analisar o impacto de campanhas reais de saúde pública. As Questões de Pesquisa 1 e 2 apresentadas no capítulo 1 nos orientaram neste estudo. O texto contido nesta seção foi organizado em forma de artigo intitulado “Systematic Review on Information Technology Approaches to Evaluate the Impact of Public Health Campaigns: Real Cases and Possible Directions” e foi publicado na *Frontiers in Public Health* (PINTO et al., 2021).

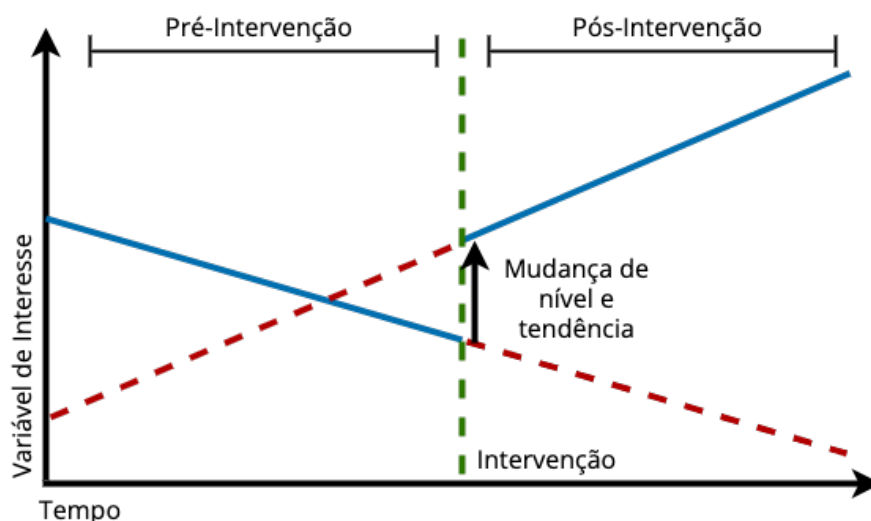


Figura 3 – Ilustração de análise de séries temporais interrompidas, com mudança de nível de tendência para pontos de dados antes e depois da intervenção.

2.4.1 Metodologia

A metodologia de pesquisa para a realização desta Revisão Sistemática de Literatura (SLR) foi realizada a partir das diretrizes propostas por [Keele et al. \(2007\)](#) e [Petersen et al. \(2008\)](#). Embora as diretrizes definidas no PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)) ([MOHER et al., 2009](#); [MOHER et al., 2015](#)) não tenham sido usadas neste estudo, as diretrizes utilizadas seguem etapas semelhantes, como (i) os artigos foram identificados por meio de busca em banco de dados; (ii) os artigos foram selecionados de acordo com os critérios de seleção referentes ao título e resumo; (iii) a elegibilidade de cada artigo na íntegra foi classificada de acordo com critérios pré-determinados (avaliação da qualidade); (iv) os artigos aprovados foram incluídos na revisão sistemática.

Estrutturamos as questões de pesquisa usando os critérios PICOC (População, Intervenção, Comparação, Resultado e Contexto), conforme sugerido por [Keele et al. \(2007\)](#).

- População: campanhas de saúde pública.
- Intervenção: o uso da tecnologia da informação para avaliar campanhas de saúde.
- Comparação: caracterização das medidas utilizadas para apoiar a avaliação da campanha de saúde.
- Resultados: caracterização do output produzido, técnicas emergentes, áreas não usuais e questões em aberto.
- Contexto: qualquer caso em que abordagens de informação tecnológica foram usadas para avaliar campanhas de saúde.

Isso leva às seguintes questões de pesquisa:

- QP1: Quais dimensões e variáveis de interesse são analisadas para avaliar o impacto das campanhas de saúde pública?
- QP2: Quais técnicas ou ferramentas são utilizadas para apoiar a análise do impacto da campanha?

2.4.2 Processo de Pesquisa

A base de dados Scopus foi a principal fonte de artigos científicos para o presente estudo. Selecionamos esta biblioteca digital porque ela contém publicações dos principais periódicos e anais de conferências, e é uma das maiores bases de dados de resumos bibliográficos e citações de literatura de pesquisa (BAAS et al., 2020). Além disso, pesquisas bibliográficas recentes a indicaram como o banco de dados mais abrangente e fácil de usar (HARZING; ALAKANGAS, 2016; MONGEON; PAUL-HUS, 2016).

Em maio de 2020, realizamos uma busca preliminar com termos relacionados aos temas de pesquisa com base na População e na Intervenção previamente definidos. A busca gerou muitos resultados. No entanto, poucos foram relevantes. Repetimos esse processo até encontrarmos uma string de pesquisa com um conjunto de resultados relevantes. Para maximizar esta busca, não incluímos os critérios de “Comparação”, “Resultados” e “Contexto” no termo de pesquisa. Consideramos o critério “Contexto” nos Critérios de Inclusão (CI), enquanto os Critérios de “Resultados” e “Comparação” definiram a estratégia de extração de dados.

O conjunto de termos de pesquisa incluiu três aspectos: (i) *“campaign*”* AND (ii) (*“public health”* OR *“communicable disease”* OR *“disease transmission”* OR *“transmitted infection”* OR *“disease outbreak*”* OR *“illness outbreak*”* OR *“infectious disease*”* OR *“disease surveillance”* OR *“disease epidemiology”*) AND (iii) (*“impact”* OR *“correlation”* OR *“assess”* OR *“effectiveness”* OR *“efficacy”* OR *“evaluation”* OR *“analysis”*).

A busca automatizada buscou esses termos em títulos, resumos e palavras-chave na área de Ciência da Computação. Para aumentar a qualidade dos estudos, eles deveriam estar publicados em periódicos revisados por pares. Além disso, foi realizada uma busca inversa definida como “backward snowball” (GREENHALGH; PEACOCK, 2005) por meio da pesquisa nas referências bibliográficas de todos os artigos selecionados.

2.4.3 Extração de Dados

Para extrair os dados dos estudos primários identificados, desenvolvemos um modelo (Tabela 1) para registrar as informações primárias em uma planilha. Para caracterizar uma campanha pública no âmbito do nosso trabalho, utilizamos a taxonomia proposta por Dorfman, Ervice e Woodruff (2002). Além disso, no entanto, adicionamos outras

propriedades que acreditamos serem relevantes: nome da campanha, tipo de anúncio, área do tópico, público-alvo, quem lançou ou patrocinou e valor gasto. Essas propriedades ajudam a diferenciar as campanhas e fornecem aos pesquisadores dados adicionais para uma melhor análise.

Tabela 1 – Formulário de Extração de Dados.

	Item Extraído	Descrição
Artigo	Artigo ID	Valor Inteiro
	Título	Nome do artigo
	Autor	Conjunto de nomes dos autores
	Ano	Ano da publicação
Campanha	Nome da Campanha	Nome ou título da campanha
	Tipo de anúncio	Tipo de anúncio usado na campanha
	Area	Tópico da area abordada na campanha
	Público-alvo	Público em geral; Público Alvo; Formadores de políticas
	País da campanha	País onde a campanha foi realizada
	Período de tempo	Período de duração da campanha
	Nível de Organização	Local, estadual ou nacional
	Lançado/Patrocinado por	Representante que lançou ou patrocinou a campanha
Valor gasto	Valor total gasto na campanha	
Avaliação	Fontes de dados	Fontes de dados exploradas pelo estudo
	Categoria das fontes de dados	Categoria que melhor caracteriza a fonte de dados
	Indicadores Analisados	Indicadores explorados pelo estudo
	Ferramentas ou Tecnologias	Técnicas, métodos, ferramentas ou outras soluções implementadas
	Resultados	Principais resultados reportados

2.4.4 Avaliação da Qualidade do Estudo

A avaliação da qualidade é essencial em revisões sistemáticas para determinar o rigor e a relevância dos estudos primários e deve ser aplicada de forma semelhante nos diferentes tipos de estudos identificados (WOHLIN *et al.*, 2013).

Para avaliar a qualidade dos estudos, desenvolvemos um modelo de seis características das campanhas de comunicação, Tabela 2 (3^a a 8^a linhas). Além disso, consideramos apenas os estudos classificados como Pesquisa de Avaliação, conforme definido por Wieringa *et al.* (2006) ou seja: técnicas, métodos, ferramentas ou outras soluções são implementadas e avaliadas na prática.

Para cada questão, a qualidade do estudo é avaliada como “Sim”, “Parcialmente” ou “Não”, que são pontuados com os valores 1, 0,5 e 0, respectivamente. Observamos que, no geral, a qualidade dos estudos foi boa, variando de 0,75 a 1. A Tabela 2 mostra os critérios de avaliação da qualidade encontrados para os estudos selecionados (S1 a S17). Um catálogo identificando os estudos primários desta pesquisa é fornecido no Apêndice A, com base em todas as propriedades identificadas.

Tabela 2 – Lista de verificação para critérios de avaliação de qualidade.

Perguntas	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17
1. A campanha está claramente caracterizada no estudo?																	
A campanha tem um nome?	1	0	1	1	1	1	0	0.5	1	1	1	1	0.5	0.5	0.5	1	1
A campanha tem anúncios?	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
A campanha tem um público-alvo?	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1
A campanha tem um período de tempo?	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1
A campanha tem um nível de organização?	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
A campanha tem um representante/patrocinador?	0	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1
Total	0.67	0.83	1	1	1	1	0.50	0.80	1	0.83	1	1	1	1	1	1	1
2. A classificação do desenho do estudo está claramente definida como Pesquisa de Avaliação?																	
As técnicas, métodos, ferramentas ou outras soluções são avaliadas na prática?	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Os resultados são investigados?	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Total	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Qualidade Geral	0.83	0.92	1	1	1	1	0.75	0.90	1	0.92	1	1	1	1	1	1	1

2.4.5 Resultados

No total, 276 artigos foram encontrados pela busca automática e a técnica “backward snowball”. No entanto, após a aplicação dos critérios de inclusão/exclusão, 17 artigos permaneceram no conjunto de artigos relevantes. Acreditamos que esses 17 artigos apresentam os critérios necessários para responder às questões de pesquisa. A tabela 3 resume os principais resultados extraídos dos 17 estudos mais relevantes. Discutimos esses resultados a seguir.

QP1: Quais dimensões e variáveis de interesse são analisadas para avaliar o impacto das campanhas de saúde pública?

Os 17 estudos avaliaram diversos tipos de variáveis de interesse, como quantidade de acesso ao site da campanha, conscientização e conhecimento sobre o assunto por meio de questionários, interesse da população através de mecanismos de busca online, engajamento por meio das Redes Sociais, e dados de exposição publicitária de comerciais de televisão.

A maioria dos estudos buscou cruzar os dados obtidos com dados socioeconômicos e censitários. A fonte de dados utilizada para obter as variáveis de interesse pode ser categorizada da seguinte forma: questionário (n=10), rede social (n=10), testes de triagem para diagnóstico de Infecções Sexualmente Transmissíveis (ISTs) (n=2), comerciais de televisão (n= 2), site da campanha (n=1), produtos farmacêuticos (n=1), leis para proibição de fumo em restaurantes (n=1) e preços do tabaco (n=1). A tabela 3 mostra as fontes de dados, variáveis de interesse, dimensões e métodos usados por cada estudo.

Tabela 3 – Fontes de dados e variáveis de interesse analisadas por estudo.

Estudo	Referência	Fonte de Dados	Variáveis de Interesse	Dimensão	Método Usado
S1	Lamos et al.	Motor de busca	Consultas de pesquisa do Bing geolocalizadas nos locais vacinados de destino.	Comunicação	Os autores apresentaram uma estrutura estatística para estimar a prevalência de uma campanha de intervenção na população a partir de dados da Internet.
		Serviço de rede social	Número de postagens no Twitter.	Comunicação	
S2	Gabarron et al.	Serviço de rede social	Número de visitantes/visitas, tempo gasto.	Comunicação	Os autores analisaram o impacto de uma página de fãs no Facebook, uma campanha publicitária no Facebook e cartazes por meio do painel de estatísticas do Facebook e do Google Analytics.
S3	Zhan et al.	Comerciais de televisão Serviço de Redes Sociais	Tweets relacionados aos anúncios televisivos.	Comunicação Comunicação	Os autores apresentaram uma estrutura estatística (Advertising Social Influence Estimation-ASIE) que prevê a probabilidade de usuários postarem tweets influenciados tanto pela transmissão de TV quanto pelos amigos no ambiente online rede social.
S4	Nadarzynski et al.	Questionário	Perguntas sobre atitudes em relação à promoção da saúde sexual nas mídias sociais, preferências pelo conteúdo das campanhas promocionais e possíveis barreiras ao envolvimento com a promoção da saúde nas mídias sociais.	Epidemiologia	Os autores analisaram entrevistas presenciais qualitativas e o envolvimento com a página de teste de clamídia por meio do painel de estatísticas do Facebook e do Google Analytics, bem como estatística descritiva para avaliar a quantidade de testes de clamídia solicitados durante o período de intervenção.
		Serviço de rede social Testes de DST	Número de visitantes/visitas e ações online. Número de testes de clamídia solicitados durante o período da campanha.	Comunicação Epidemiologia	
S5	Kite et al.	Questionário	Dúvidas sobre a experiência e satisfação com uma APS no Facebook. CEP auto-relatado dos participantes, altura e peso.	Comunicação	Os autores realizaram testes estatísticos para comparar grupos de usuários do Facebook e exploraram seu perfil para examinar as características dos fãs da página, bem como analisaram uma pesquisa online para investigar como os usuários estavam interagindo com a página da campanha e outras páginas de saúde em Facebook.
S6	Wakefield et al.	Questionário	A prevalência de tabagismo foi estimada a partir de questionários.	Epidemiologia	O grupo usou análise de média móvel integrada autorregressiva de séries temporais em um software estatístico (SAS) para estimar o efeito da publicidade antitabaco e das políticas de tabaco sobre prevalência mensal de tabagismo.
		Comerciais de televisão	Ocorrências de todos os anúncios relacionados ao tabaco que aparecem na televisão.	Comunicação	
		Preços do tabaco	O custo do cigarro foi medido com a razão entre o preço médio de varejo recomendado por maço de cigarros e os ganhos médios semanais.	Políticas Públicas	

		Produtos farmacêuticos	Uso populacional de produtos farmacêuticos para cessação do tabagismo.	Epidemiologia	
		Leis de restaurantes livres de fumo	A exposição da população às leis antifumo foi expressa como a porcentagem da amostra total que estava sujeita a tais leis.	Políticas Públicas	
S7	Grandpre et al.	Questionário	Perguntas sobre a percepção dos participantes em relação às mensagens da campanha e sua intenção comportamental pós-campanha.	Comunicação	Os autores realizaram testes estatísticos para analisar três variáveis primárias de interesse (avaliação da fonte, avaliação da mensagem e intenção comportamental), com base na teoria da reatância psicológica.
S8	Bull et al.	Questionário	Perguntas sobre o comportamento de cuidados de saúde sexual. Características demográficas fornecidas pelos participantes.	Epidemiologia	Os autores analisaram dados relacionados ao engajamento com a página de intervenção do Facebook por meio do Google Analytics e uma pesquisa online, usando um software estatístico (SAS).
		Serviço de rede social	Engajamento com a página do Facebook.	Comunicação	
S9	Emery et al.	Serviço de rede social	Aceitação, rejeição e desconsideração da mensagem de cada tweet identificado como relevante para as Dicas.	Comunicação	Os autores apresentam uma análise de mensagens do Twitter sobre uma campanha de saúde usando uma estrutura analítica.
S10	Allem et al.	Serviço de rede social	As mensagens foram rotuladas como campanha anti, pró ou neutra.	Comunicação	Os autores analisaram o conteúdo das mensagens do Twitter sobre uma campanha de saúde por meio de análise estatística.
S11	Syred et al.	Serviço de rede social	Volume de interação. O número total de fãs, postagens no mural e comentários ao longo do tempo, dados demográficos dos fãs. Números de acesso ao site e padrões de visualização.	Comunicação	Os autores examinaram dados quantitativos e qualitativos em uma página do Facebook sobre uma campanha de saúde. O Google Analytics foi usado para descrever o número de pessoas que usam a página e os padrões de visualização.
S12	Dowshen et al.	Serviço de rede social	Dados relacionados a números de “curtidas”, visitas e número de “seguidores”.	Comunicação	O grupo usou estatísticas descritivas padrão para avaliar uma campanha de saúde: rastreando o uso de sites/mídias sociais, pesquisa on-line e comparando taxas de testes de IST.
		Site da campanha	Dados relacionados ao engajamento no site da campanha.	Comunicação	
		Questionário	Informações sobre idade, como ouviram falar da campanha, avaliaram o conhecimento sobre ISTs e se a campanha influenciou a intenção de fazer o teste.	Comunicação e Epidemiologia	
		Testes de DST	Verificação de testes de DST pré e pós-campanha.	Epidemiologia	
S13	Ahmed et al.	Questionário	Perguntas sobre variáveis demográficas, altura e peso, como souberam do estudo, história sexual, experiência e conhecimento sobre IST.	Epidemiologia	O grupo avaliou a viabilidade de usar SNSs para recrutar mulheres jovens para completar uma pesquisa relacionada à saúde. Os dados foram analisados por meio de um software estatístico (STATA).

S14	Platt et al.	Serviço de rede social	Engajamento com o Facebook: curtidas na página, visualizações, postagens e engajamento no álbum de fotos (curtidas, conexões, compartilhamentos, conversas e comentários).	Comunicação	Os autores apresentaram uma estrutura para examinar um espectro de resultados de engajamento do Facebook, desde a observação até a conversa. Os dados foram fornecidos pelo painel do Facebook.
S15	Gamage et al.	Questionário	Número de clientes que entraram em contato com o serviço após a exposição dos anúncios e tipo de anúncio utilizado.	Comunicação	Os autores apresentaram um estudo descritivo simples avaliando a eficácia de diferentes métodos publicitários.
S16	O'Hara et al.	Questionário	Perguntas sobre variáveis demográficas, fatores de risco, se o entrevistado estava ciente da campanha “Troque”, atitudes e comportamentos em relação à dieta e exercícios, e intenções e ações comportamentais.	Comunicação e Epidemiologia	Os autores avaliaram uma campanha de saúde por meio de questionários telefônicos seriados transversais. Os dados foram analisados usando um software estatístico (STATA).
S17	Kite et al.	Questionário	Perguntas sobre conscientização, conhecimento, atitudes e intenções da campanha. Comportamento atual e mudança de comportamento recente. Variáveis demográficas, categoria do índice de massa corporal e pontuação do índice de risco.	Comunicação e Epidemiologia	Os autores apresentaram um estudo de desenho de corte e sed modelos lineares generalizados mistos em um software estatístico (SAS) para examinar a conscientização da campanha, conhecimento, atitudes, intenções e comportamentos ao longo do tempo.

Apesar da variedade de variáveis de interesse analisadas, podemos observar uma predominância relacionada a i) avaliação do engajamento dos usuários no assunto da campanha por meio das redes sociais, como Facebook e Twitter (LAMPOS et al., 2015; GABARRON et al., 2019; ZHAN et al., 2017; NADARZYNSKI et al., 2019; BULL et al., 2012; EMERY et al., 2014; ALLEM et al., 2017; SYRED et al., 2014; DOWSHEN et al., 2015; PLATT et al., 2016); e ii) avaliação do conhecimento, referentes às atitudes e o comportamento das pessoas por meio de questionários (NADARZYNSKI et al., 2019; KITE et al., 2018b; WAKEFIELD et al., 2008; GRANDPRE et al., 2003; BULL et al., 2012; DOWSHEN et al., 2015; AHMED et al., 2013; GAMAGE et al., 2011; O'HARA et al., 2016; KITE et al., 2018a) antes e/ou após a divulgação da campanha.

As redes sociais foram usadas para recrutar participantes para responder a questionários (KITE et al., 2018b; BULL et al., 2012; AHMED et al., 2013), avaliando o alcance e o conhecimento das pessoas impactadas. Os autores argumentaram que essa abordagem produziu bons resultados devido à geolocalização precisa do usuário e dados demográficos (por exemplo, sexo, idade, educação, ocupação, idioma falado).

Além disso, seis estudos (GABARRON et al., 2019; NADARZYNSKI et al., 2019; KITE et al., 2018b; BULL et al., 2012; DOWSHEN et al., 2015; PLATT et al., 2016) destacaram que os anúncios do Facebook ajudaram significativamente a aumentar o número de pessoas que visitam a página da campanha e conseqüentemente o número de usuários que poderiam melhorar seu conhecimento e comportamento de cuidados de saúde por causa da campanha.

Os anúncios televisivos foram avaliados em duas campanhas anti-tabaco usando o tempo de transmissão e sua classificação Nielsen¹. Zhan et al. (2017) usou classificações de TV correlacionadas com mídias sociais (Twitter), enquanto Wakefield et al. (2008) correlacionou classificações de TV com questionários, preços de tabaco, uso de produtos farmacêuticos para parar de fumar e leis de restaurantes livres de fumo. Em ambos os casos, os resultados indicaram sucesso nas abordagens utilizadas para avaliar o impacto da campanha.

Em relação aos testes de triagem de ISTs Nadarzynski et al. (2019) e Dowshen et al. (2015) indicam que sua campanha atingiu seu objetivo, ou seja, mudar o comportamento de cuidado sexual e atingir um grande número de pessoas. De fato, a demanda por pedidos de teste aumentou no período pós-campanha.

O agrupamento das variáveis de interesse em dimensões (áreas de contexto) proporcionou uma melhor compreensão de qual dimensão foi mais explorada. Nossos achados mostram a predominância de três dimensões: comunicação (n=20), epidemiologia (n=10) e políticas públicas (n=2).

A dimensão comunicação foi atribuída quando a variável de interesse buscou avaliar o engajamento na internet/redes sociais ou pesquisas questionaram sobre o conhecimento

¹ <<https://www.nielsen.com/ca/en/about-us/>>

do usuário sobre a campanha de saúde. A dimensão epidemiologia foi atribuída quando a variável de interesse buscava avaliar padrões de condições de saúde e doença em uma população definida. Por fim, a dimensão de política pública foi detectada quando a variável de interesse se referia a políticas de controle, como mudanças nos impostos sobre produtos e novas leis.

QP2: Quais técnicas ou ferramentas são utilizadas para apoiar a análise do impacto da campanha?

Considerando que as redes sociais (Facebook, Twitter e Instagram) foram amplamente explorados nos estudos, as principais ferramentas foram os *dashboards* dos respectivos sites (explorados por Gabarron et al. (2019), Nadarzynski et al. (2019), Platt et al. (2016), Dowshen et al. (2015), Syred et al. (2014), Bull et al. (2012)). Os painéis de redes sociais geralmente fornecem dados suficientes para verificar o envolvimento de sites hospedados de forma privada.

No entanto, analisar o conteúdo gerado pelo usuário em uma rede social, como o Twitter, e inferir correlações com os dados da campanha requer uma avaliação detalhada do conteúdo. De acordo com os estudos realizados por Zhan et al. (2017), Allem et al. (2017), Emery et al. (2014), os dados do Twitter foram obtidos por meio de um provedor licenciado (Gnip²). Este fornecedor fornece acesso em tempo real a 100% de todos os tweets e metadados. Essas análises e correlações foram apoiadas por modelos lineares gerais usados para avaliar a relação entre dados demográficos e conscientização da campanha. Além disso, o Google Analytics foi usado para identificar o número de usuários únicos, total de visitas, visualizações de página, visualizações únicas e duração média de visitas para cada página do site da campanha.

Em relação ao software para análise estatística e ciência de dados para explorar, visualizar, modelar e fazer inferências nos dados, O'Hara et al. (2016), Dowshen et al. (2015), Ahmed et al. (2013) usou Stata enquanto Kite et al. (2018a), Bull et al. (2012), Wakefield et al. (2008) usou Statistical Analysis System (SAS). NVivo, um software de análise de dados qualitativos, foi usado por Kite et al. (2018b) para análise textual dos questionários e por Syred et al. (2014) para capturar o conteúdo da página da campanha, usando um método de contagem simples para descrever o conteúdo do usuário e moderador, a *thread* de discussão e as intervenções do moderador. Dowshen et al. (2015) usou um aplicativo de gerenciamento de dados e administração de pesquisas baseado na web chamado REDCap³.

Em Kite et al. (2018b), todas as análises estatísticas foram realizadas com o SPSS Statistics 22, uma plataforma de software que oferece análises estatísticas avançadas, uma vasta biblioteca de algoritmos de aprendizado de máquina, análise de texto, extensibilidade de código aberto, integração com big data e implantação em aplicativos.

² <<http://www.gnip.com>>

³ <<https://www.project-redcap.org/>>

Lamos et al. (2015) propôs uma estrutura estatística para transformar conteúdo gerado pelo usuário publicado em plataformas web para avaliação de impacto de uma intervenção orientada para a saúde e Grandpre et al. (2003) usou uma técnica estatística chamada ANOVA (ANalysis Of VAriance) para testar todas as variáveis definidas no estudo.

2.4.6 Conclusões

Este estudo forneceu uma visão geral dos artigos que avaliam o impacto das campanhas de saúde pública. A análise concentrou-se na identificação de variáveis de interesse, técnicas e ferramentas. Por fim, discutimos os resultados e descobrimos novas variáveis de interesse.

As campanhas de saúde pública promovem estrategicamente a conscientização, aumentam o conhecimento e incentivam a população-alvo a adotar atitudes e comportamentos desejáveis. Conforme observado, seu impacto deve ser medido em diversas dimensões como: i) comunicação (engajamento em redes sociais, questionários avaliando o conhecimento do usuário sobre a campanha), ii) epidemiologia (teste de triagem de doenças, notificação de casos, questionários avaliando o conhecimento do usuário sobre a doença/questões de saúde) e iii) aplicação de políticas (estratégias de leis sobre promoção da saúde).

Em geral, observamos uma lacuna na avaliação do impacto das campanhas de saúde pública em relação aos dados *online* (ou seja, notícias online) e outros conteúdos da *Internet* gerados pelos usuários. O Google Search e o Yahoo devem ser usados em conjunto com o buscador Bing para verificar o aumento de notícias relacionadas à campanha. Além disso, várias perguntas devem ser feitas, como: O que a análise desse conteúdo pode trazer? As campanhas estão ajudando a aumentar as notícias espontâneas sobre o tema relacionado? Ao avaliar uma população ao longo do tempo, o Google Trends é sem dúvida uma excelente ferramenta para demonstrar como a população procurou saber sobre um determinado tópico (DREHER et al., 2018; LING; LEE, 2016; GLYNN et al., 2011). No entanto, não foi explorado no contexto desses trabalhos.

Além disso, bancos de dados de artigos científicos como Springer Nature, Wiley Blackwell, Taylor & Francis, IEEE, American Physical Science e Elsevier e seus indexadores como Scopus, Web of Science (WoS) e Google Scholar poderiam ser usados para demonstrar o interesse da academia em desenvolver pesquisas sobre o tema da campanha em uma comparação de tempo (antes, durante e após a campanha). Essas variáveis de interesse podem apresentar uma nova dimensão, a Educação.

Assim, não identificamos nos estudos uma abordagem sistemática apoiada em um sistema de informação que possa ser reaplicado para diferentes campanhas utilizando dados heterogêneos de várias dimensões. Além disso, poucos estudos estabeleceram uma correlação entre a intervenção (campanha) e as variáveis de interesse analisadas ao longo do tempo.

Essa análise multidimensional pode fornecer informações mais abrangentes das campanhas de saúde pública para entender seu alcance, encontrar correlações entre diferentes variáveis de interesse e instigar análises mais aprofundadas.

3 Um Framework para Analisar Intervenções em Saúde Pública

Este capítulo apresenta o *framework* para a análise multidimensional das intervenções de saúde pública. Discutimos seus objetivos, atividades, arquitetura, variáveis de interesse e dimensões em detalhes.

3.1 Framework de Análise Multidimensional

Um *framework* conceitual é uma organização de ideias que pode explicar melhor a sequência de fluxos, hipóteses ou relações sobre o fenômeno a ser estudado de maneira lógica e integrada (CAMP, 2001; OSANLOO; GRANT, 2016; GERVASI; MASTROGIACOMO; FRANCESCHINI, 2020).

Assim, o Framework de Análise Multidimensional (FAM) é um modelo para a informação de análise temporal do alcance das intervenções em saúde. Foi desenvolvido considerando o contexto do Projeto “Sífilis Não!” (em resposta rápida à epidemia de sífilis no Brasil), e considerando aspectos de uma campanha de saúde do mundo real e a revisão de literatura. Por fim, foi implementado no sistema Hermes (PINTO et al., 2020; PINTO et al., 2021; PINTO et al., 2022) que gerencia um ciclo de vida completo dos dados em quatro dimensões.

O FAM pretende abranger até agora seis dimensões de informação: campanha, comunicação, educação, vigilância epidemiológica, finanças e sociodemografia. Isso porque encontramos essas dimensões na literatura e seus conjuntos de dados geralmente estão disponíveis. A Figura 4 ilustra os caminhos conceituais pelos quais a intervenção em saúde pública pode ser analisada. Inicialmente, precisamos obter os dados correspondentes para cada dimensão a ser analisada. A dimensão possui uma ou mais variáveis de interesse, e essas variáveis são compostas por um conjunto de dados, detalhado na seção 3.3.

A Campanha (A) é a primeira dimensão identificada no *framework*. Seus dados representam quais ações foram realizadas durante um período, sobre determinado tema, para um público-alvo e quanto foi gasto. A dimensão Comunicação (B) analisa o alcance da intervenção na Internet. A dimensão Educação (C) visa analisar a extensão do aprendizado científico por meio de atividades de aprendizagem ao longo da vida para profissionais de saúde. Por fim, a dimensão Vigilância Epidemiológica (D) visa analisar a intervenção por meio de variáveis de interesse relacionadas à atenção primária à saúde.

Dessa forma, o FAM define um fluxo de dados exploratório que intercala pré-processamento de dados, análise descritiva de sazonalidade e tendência, e análise de correlações buscando identificar novas informações e explicar aquelas já reveladas.

O *framework* se propõe a analisar os dados a partir de uma perspectiva de séries temporais, para entender como esses dados se comportam ao longo do tempo. Nesta fase, os dados são processados através de um ciclo de vida completo dos dados, conforme explicado na Seção 3.2. Por fim, os dados devem ser transformados e disponibilizados por meio de um *Dashboard* (E), para que os Tomadores de Decisão (F) possam utilizá-los para planejar ou implementar políticas em suas áreas.

No entanto, a escolha das dimensões e tipos de análise depende das necessidades dos tomadores de decisão. Além disso, os métodos computacionais utilizados para processar os dados dependem dos tipos de análise a serem realizados e dos conjuntos de dados disponíveis. Desta forma, novas dimensões e variáveis de interesse ou outros tipos de análise poderão ser integrados ao *framework* (G).

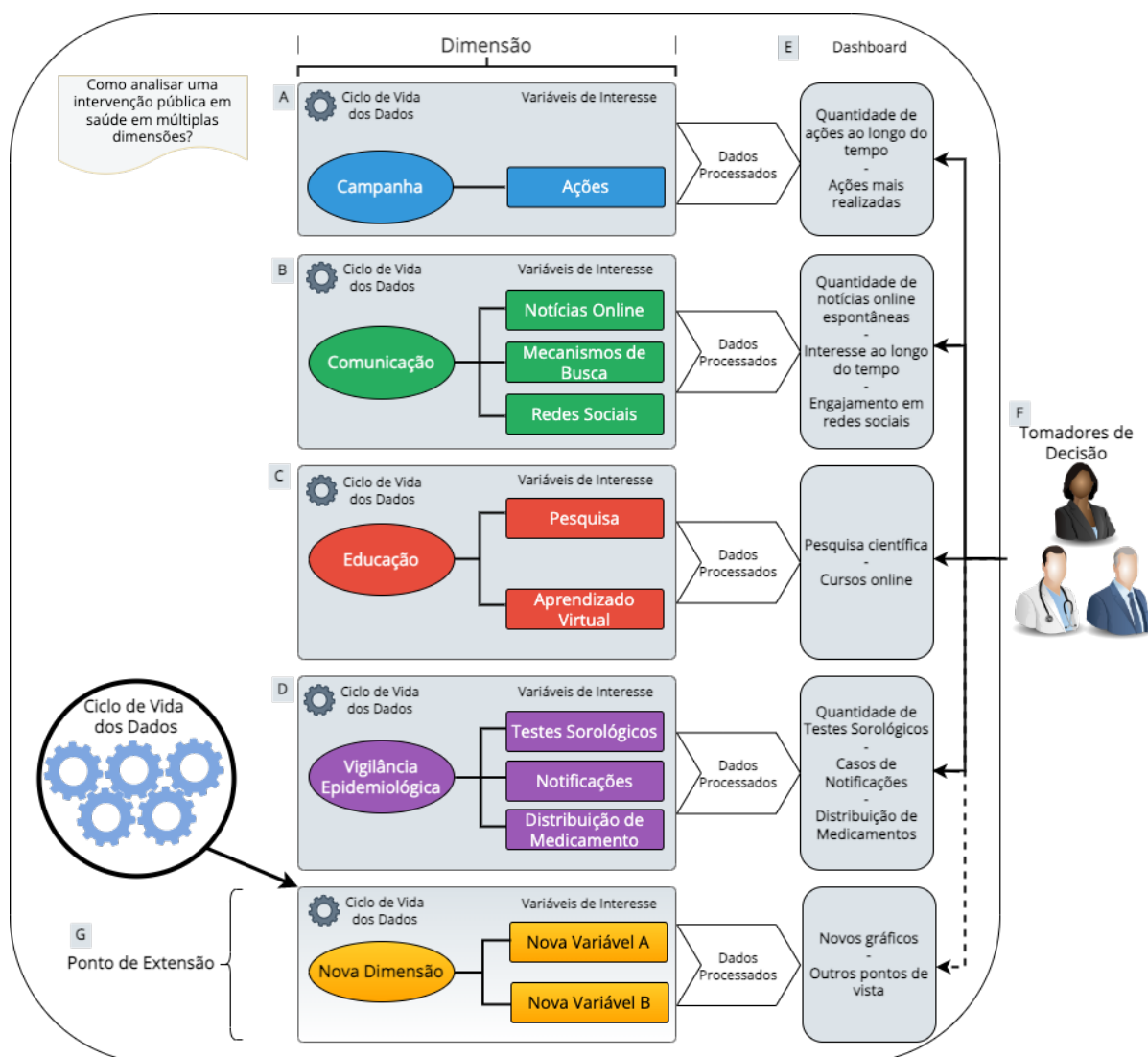


Figura 4 – Framework de Análise Multidimensional.

3.2 Ciclo de Vida dos Dados

De acordo com Berman et al. (2018), os dados têm um ciclo de vida, como organismos vivos, desde o nascimento, passando por uma vida ativa, até a “imortalidade” ou alguma forma de expiração. O ciclo de vida dos dados é fundamental para entender as oportunidades e os desafios de aproveitar ao máximo os dados digitais.

A Figura 5 ilustra a sequência de atividades realizadas através do *framework* apresentado, com base no ciclo de vida dos dados proposto por Berman et al. (2018). O processo consiste em cinco atividades básicas: adquirir (caixa amarela), limpar (caixa verde), usar (caixa azul), publicar (caixa laranja) e preservar (caixa cinza). Nas subseções a seguir, explicamos em detalhes cada tarefa e fluxo de trabalho.

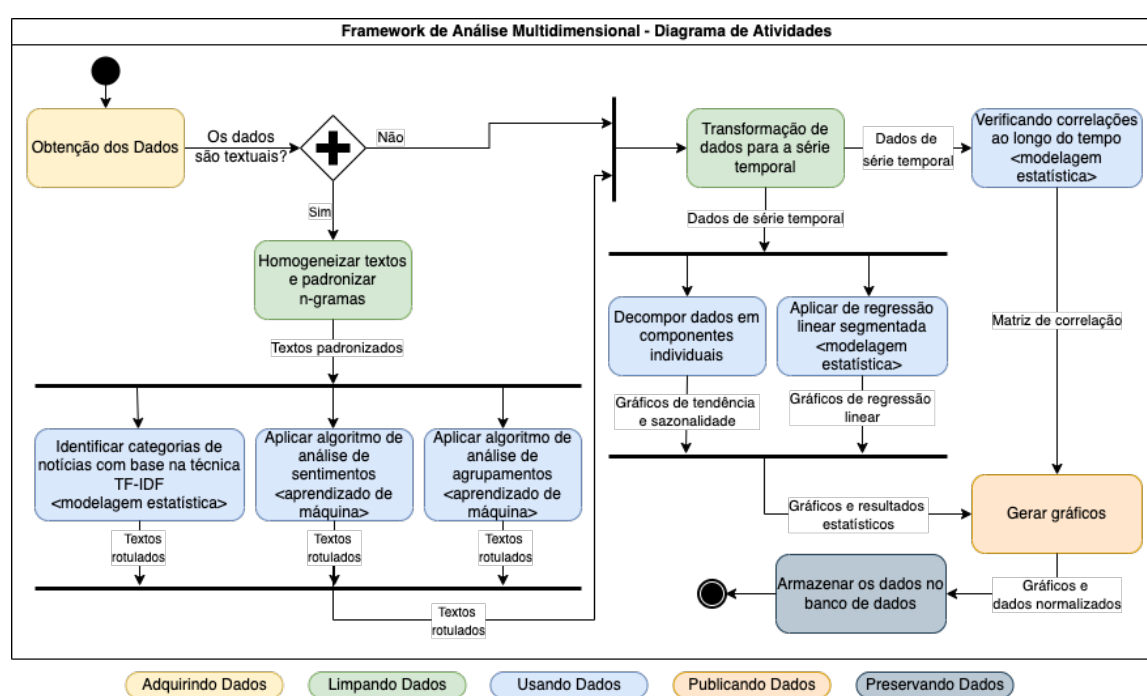


Figura 5 – Diagrama de atividade do *framework* de análise multidimensional.

Adquirindo Dados

O processo inicia a partir da tarefa *Obtenção dos dados*, que podem ser adquiridos manualmente ou automaticamente para qualquer dimensão, desde que disponível.

Caso os dados precisem ser informados manualmente, o cientista de dados deve inserir manualmente esses dados por meio de uma interface amigável ou fazer upload de um arquivo em um formato previamente definido pelo sistema implementado. No entanto, se os dados disponíveis para análise estiverem em plataformas *online* (e.g., site, web services, plataformas virtuais de aprendizagem), o sistema implementado deve obtê-los via microsserviços, que são responsáveis por obter os dados disponíveis na *Internet* por meio de APIs ou Web Scraping.

Web Scraping é uma técnica usada para obter algumas informações de um site automaticamente, em vez de copiá-las manualmente, geralmente no formato HTML. Assim, o web scraping se concentra em transformar dados não estruturados na Web, analisando seus conteúdos para encontrar e extrair dados e estruturas que podem ser armazenadas em um banco de dados ou planilha (VARGIU; URRU, 2013; GLEZ-PEÑA et al., 2014).

Por outro lado, as APIs fornecem uma interface uniforme para realizar a troca de informações entre dois ou mais programas. As APIs evitam analisar HTML e, em vez disso, acessam os dados estruturados diretamente usando formatos padrões como XML e JSON.

Uma vez que os dados são obtidos, seja *online* ou manualmente, passamos para o ponto de decisão *Os dados são textuais?*. Essa verificação é necessária para verificar se os dados são de conteúdo textual ou não. O conteúdo textual deve passar para a tarefa *Homogenizar textos e padronizar n-gramas*, e o conteúdo não textual deve passar para *Transformação de dados para a série temporal*. Ambas as tarefas estão relacionadas à atividade **Limpando Dados**, explicada abaixo.

Limpando Dados

Os dados obtidos vêm de diferentes fontes e têm formatos diferentes. Nesta atividade, várias técnicas de pré-processamento podem ser necessárias dependendo da natureza e características do conjunto de dados. Além disso, esses dados heterogêneos às vezes não são estruturados, dificultando a análise e o processamento para entender o verdadeiro impacto das intervenções projetadas. Portanto, é fundamental homogeneizá-los, padronizá-los e transformá-los.

Assim, se o fluxo passar para *Homogenizar textos e padronizar n-gramas*, uma série de etapas em uma sequência lógica será realizada para padronizá-los e habilitá-los para uma série de novas atividades relacionadas aos conteúdos textuais, o que implica possibilidades de descoberta do conhecimento.

Os seguintes procedimentos são realizados para homogeneizar os dados textuais:

- Alterar a capitalização das palavras para minúsculas;
- Excluir caracteres especiais;
- Remover a pontuação;
- Remover espaços extras;
- Remover acentos e números;
- Remover “stop words” (palavras sem significado, como advérbios e preposições);
- Realizar a lematização e agrupar palavras flexionadas. Com esta etapa, por exemplo, “corre”, “correndo” e “correu” são agrupados no termo “correr”.

Após esses procedimentos, o sistema implementado lista as palavras mais importantes, com base em seu corpus, agrupadas por n-gramas, e o cientista de dados deve decidir qual n-grama (por exemplo, unigram, bi-gram, trigram) deve seguir para a próxima etapa. Neste ponto, três tarefas podem ser realizadas: *Identificar categorias de notícias com base na técnica TF-IDF*, *Aplicar algoritmo de análise de sentimentos* e *Aplicar algoritmo de análise de agrupamentos*, que serão descritos na atividade **Utilização dos Dados**.

Por outro lado, se os dados não estiverem relacionados a um conteúdo textual, o fluxo deve seguir para *Transformar de dados em série temporal*. O núcleo principal da análise de séries temporais é baseado em dois atributos: pontos que representam os próprios dados e intervalos de tempo. Nesse sentido, os dados obtidos na tarefa anterior precisam ser transformados neste formato. Isso significa que o ponto numérico pode ser representado por um valor do tipo Decimal (que facilmente pode converter outros formatos como um Inteiro) e que a data deve ser padronizada, por exemplo, no formato AAAA-MM-DD. Usar um padrão de data é útil para identificar automaticamente se o período tem um intervalo equivalente como: diário, semanal, mensal, anual ou outro.

Usando Dados

O desafio na análise de uma intervenção em saúde é entender como transformar dados brutos em informações relevantes a partir de uma perspectiva de séries temporais. Por exemplo, as notícias *online* são uma rica fonte de informação que serve para alertar, conscientizar e mudar o comportamento da população. No entanto, isso pode ser potencializado descobrindo quais características são intrínsecas aos textos que induzem a população a mudar seus hábitos e buscar tratamento. Da mesma forma, postagens e comentários em redes sociais podem ser coletados e classificados para avaliar o conhecimento da população sobre a dimensão da Comunicação.

Analisar o conteúdo textual dessas variáveis de interesse, e classificá-las em diferentes aspectos, como o nível de qualidade ou o sentimento que elas carregam, são de extrema importância para entender o impacto na população e conseqüentemente em outras variáveis como testes de triagem e tratamento de doenças, pertencentes a outras dimensões.

A transformação desses dados brutos em novos modelos é o aspecto fundamental de sustentação desse *framework* para que possamos buscar correlações entre variáveis de interesse e definir sua importância e como elas podem influenciar umas às outras.

Assim, a atividade **Usando Dados** é responsável por processar o conjunto de dados usando modelos estatísticos e métodos de aprendizado de máquina para gerar gráficos e resultados estatísticos que possibilitem a análise de efeitos ao longo do tempo. Seguindo o fluxo proposto, três tarefas podem ser realizadas para análise de dados textuais:

1. *Identificar categorias de notícias com base na técnica TF-IDF* pode mostrar categorias

'ocultas', presentes nos textos ao longo do tempo e, assim, avaliar se essas categorias influenciam de alguma forma outras variáveis de interesse.

2. *Aplicar algoritmo de análise de sentimentos* pode classificar textos como sentimentos, rotulando-os como “positivo”, “negativo” ou “neutro”. Novamente, extraímos valores 'ocultos' de textos (por exemplo, notícias, postagens em redes sociais, tweets) que ganham outro significado que não o texto puro.
3. *Aplicar algoritmo de análise de agrupamento* ajuda a pesquisar textos semelhantes ao longo do tempo e ajuda a entender qual conteúdo é historicamente relevante. Além disso, esse grupo de textos pode se destacar sobre outros a serem investigados.

É importante ressaltar que o resultado do processo anterior são textos rotulados com categorias encontradas ao longo do tempo. Cada texto tem pelo menos um ou mais rótulos e uma data. Assim, podemos somar e agrupar (por categorias, sentimentos ou agrupamentos) os textos ao longo do tempo, usando a tarefa *Transformar de dados em série temporal*, e avançar para métodos de séries temporais.

Da mesma forma, três tarefas podem ser executadas para análise de séries temporais:

1. *Decompor dados em componentes individuais* facilita a identificação de mudanças nas tendências e sazonalidade dos dados. Uma mudança de tendência pode indicar que algo mudou o curso atual, enquanto a sazonalidade pode explicar determinados comportamentos em um período específico. Esses aspectos são fatores-chave e servem de entrada na configuração de métodos de regressão linear.
2. *Aplicar de regressão linear segmentada* é essencial para estimar estatisticamente se uma variável de interesse mudou após a intervenção e qual foi seu impacto imediato quando esta política pública foi aplicada.
3. *Verificar correlações ao longo do tempo* é o passo final antes de tirar conclusões considerando todo o processo. Nesta tarefa, podemos verificar se uma variável de interesse se correlaciona com outra e é estatisticamente significativa. É importante observar que, neste momento, temos um novo conjunto de dados, limpo, transformado e baseado em métodos computacionais e estatísticos.

Publicando Dados

A representação visual dos dados é primordial para analisar o impacto da intervenção, determinando se um efeito é imediato ou tardio, abrupto ou gradual, e se um efeito é sustentado ou não ao longo do tempo (CARROLL, 2008).

O conjunto de dados previamente obtido, processado e transformado pode ser publicado por meio da tarefa *Gerar gráficos*, permitindo que os tomadores de decisão

realizem suas análises para obter insights sobre suas áreas, comparar resultados entre dimensões e verificar a obtenção de resultados. Ao analisar estes resultados, eles poderão adaptar novas estratégias de prevenção em intervenções de saúde pública para melhor usar seus recursos e esforços.

Preservando Dados

Por fim, após a conclusão de todos os processos, o fluxo prossegue para *Armazenar os dados no banco de dados*. Assim, o conjunto de dados pode ser preservado como um banco de dados estruturado, permitindo a reutilização futura pelo próprio sistema implementado ou atendendo outros sistemas por meio de APIs.

3.3 Dimensões

Durante esta pesquisa, identificamos seis dimensões que agrupam variáveis de interesse em uma perspectiva de séries temporais para analisar intervenções em saúde pública, conforme apresentado na Figura 6: campanha, comunicação, educação, vigilância epidemiológica, finanças e sociodemografia. Nossa abordagem visa coletar cada uma dessas variáveis, separá-las em um modelo de série temporal e mostrar as informações mais relevantes por meio de representação visual. Então, cada *stakeholder* pode analisá-la e tirar conclusões sobre sua área de conhecimento.

Campanha	Comunicação	Educação	Epidemiologia	Finanças	Socio-demografia
<ul style="list-style-type: none"> - título - período - custo - agravo - região <p>Ações</p> <ul style="list-style-type: none"> - tipo de anúncio - período - custo - público-alvo 	<p>Notícias online</p> <ul style="list-style-type: none"> - título - link - data <p>Interesse ao longo do tempo</p> <ul style="list-style-type: none"> - pontuação - data/período <p>Redes sociais</p> <ul style="list-style-type: none"> - postagens - curtidas - compartilhamentos - comentários 	<p>Pesquisa</p> <ul style="list-style-type: none"> - título - data de publicação - autor - universidade - tipo <p>Aprendizado Virtual</p> <ul style="list-style-type: none"> - Id do estudante - data de registro - curso - região - ocupação 	<p>Testes de triagem</p> <ul style="list-style-type: none"> - quantidade - região - data <p>Casos de notificação</p> <ul style="list-style-type: none"> - quantidade - região - data <p>Distribuição de medicamentos</p> <ul style="list-style-type: none"> - quantidade - região - data 	<p>Custos de hospitalização</p> <ul style="list-style-type: none"> - custo - região - data <p>Custos de tratamento</p> <ul style="list-style-type: none"> - custo - região - data 	<ul style="list-style-type: none"> - área territorial - população estimada - densidade demográfica - geolocalização - Gênero sexual - idade - nível de escolaridade - PIB - IDH - renda familiar
Perspectiva da Série Temporal					

Figura 6 – Tipos de dados brutos identificados neste estudo.

3.3.1 Campanha

As campanhas de saúde pública visam influenciar o comportamento por meio de mensagens que inspiram mudanças de hábitos saudáveis. Tais mensagens podem ser comunicadas em uma área específica (ou seja, infecções sexualmente transmissíveis, álcool, tabaco, obesidade), para um público-alvo (ou seja, homens, mulheres, crianças, homens

que fazem sexo com homens, profissionais do sexo), em um período (curto ou longo), com um nível de organização (local, estadual ou nacional).

Essas mensagens podem ser disseminadas em mídias que atingem grandes audiências, por exemplo, comerciais de TV, rádio, mídia impressa, revistas e na Internet, por meio de plataformas de streaming, redes sociais e páginas da web. As campanhas têm um representante que as criou, lançou e pagou seus custos.

3.3.2 Comunicação

A dimensão da comunicação diz respeito ao conteúdo gerado pela Internet, como (i) notícias online, (ii) interesse em um determinado tópico em mecanismos de busca na Internet e (iii) redes sociais.

Notícias online relacionadas a um tema específico pode mostrar o interesse da mídia *online* (jornais, revistas de notícias, blogs, sites governamentais, por exemplo) por compartilhar conhecimento e, conseqüentemente, obter mais audiência.

Mecanismo de busca na internet possuem um conjunto de dados (registros das buscas) que pode medir o interesse em um tópico ou termo de pesquisa ao longo do tempo, onde ele é mais pesquisado ou o que mais as pessoas pesquisam em relação a ele. Pode ajudar a comparar o interesse das pessoas pelo assunto durante o período da campanha, detectando assim qual tipo de ação atraiu mais atenção do público.

Redes Sociais desempenham um papel fundamental no conteúdo gerado pelo usuário. Elas podem atuar como veículos para melhorar a saúde de adolescentes e jovens e desempenhar um papel vital em um contexto educacional (CASTILLO et al., 2020).

3.3.3 Educação

A educação é um direito humano fundamental e um processo transformador em muitos outros aspectos da vida de um indivíduo. Com o advento das tecnologias de informação e comunicação, um aliado essencial na educação formal e informal é a Educação a Distância (ARAÚJO et al., 2019). As variáveis educacionais estão relacionadas aos cursos educacionais e ao crescimento da pesquisa. É um importante indicador para qualificar os profissionais de saúde, demonstrando interesse científico.

Plataformas de *aprendizagem virtual* para promover Cursos Online Abertos Massivos (MOOCs) para profissionais de saúde têm sido utilizadas pelo Ministério da Saúde, no Brasil, como estratégias de aprendizagem ao longo da vida. Portanto, o aumento de MOOCs oferecidos na plataforma AVASUS e profissionais capacitados são fundamentais para a análise dessa variável.

Além disso, o número de pesquisas e publicações pode mostrar o interesse relacionado à qualificação na área acadêmica.

3.3.4 Vigilância Epidemiológica

As variáveis de vigilância epidemiológica visam compreender a consciência da população sobre as mudanças de comportamento. Podemos atentar para variáveis como (i) testes de triagem, (ii) casos de notificação e (iii) distribuição de medicamentos.

Aumentos nos testes de triagem podem significar que a população entende a preocupação com um problema e recebe ajuda. Diretamente relacionados a isso estão os casos de notificação. Aumentar o número de testes de triagem não significa necessariamente um aumento real de casos. Muito provavelmente, o aumento da testagem e conseqüentemente o aumento do número de notificações, pode levar ao descobrimento de subnotificações anteriores. Por fim, podemos verificar a distribuição dos medicamentos, o que significa que a população está sendo tratada e curada.

3.3.5 Finanças

Embora não tenhamos encontrado estudos na revisão sistemática que analisassem os impactos alcançados pelo custo das internações ou tratamento de uma determinada doença, acreditamos que essa seja uma dimensão importante. Os custos de hospitalização poderiam ser evitados com a implementação de uma intervenção de saúde pública eficaz, bem como os custos do tratamento poderiam ser reduzidos.

3.3.6 Sociodemografia

A análise sociodemográfica permite um diagnóstico mais detalhado das variáveis de interesse, avaliando principalmente as desigualdades e especificidades socioeconômicas de cada região, facilitando o planejamento de ações corretivas de acordo com a realidade local. Em relação a esses atributos, podemos citar: área territorial, população estimada, densidade demográfica, geolocalização, sexo, idade, escolaridade, Produto Interno Bruto (PIB), Índice de Desenvolvimento Humano (IDH), renda familiar.

4 Hermes - Um Ecossistema de Software para Analisar Intervenções em Saúde

Um ecossistema de software refere-se a uma coleção de produtos de software com algumas relações simbióticas (MESSERSCHMITT; SZYPERSKI, 2003), usado por um conjunto de *stakeholders* através de uma plataforma tecnológica comum (MANIKAS; HANSEN, 2013). Um ecossistema de software deve ser visto através de múltiplas perspectivas, como tecnologia, comunicação e vigilância epidemiológica.

O Hermes automatiza o *framework* conceitual apresentado no capítulo 3. Seu nome foi inspirado na mitologia grega - Hermes era o mensageiro dos deuses. Assim, o Hermes se concentra em levar conhecimento aos *stakeholders* por meio da coleta, processamento, integração e visualização de dados heterogêneos dessas informações por meio de uma única plataforma, usando processamento de linguagem natural, técnicas de aprendizado de máquina e análise estatística para possibilitar a análise de políticas de saúde pública.

Ele foi desenvolvido em linguagem Python e usou o Django (HILLAR, 2018) como o *framework* model-view-controller (MVC). Além disso, foi construído um módulo externo no Software R (COMPUTING et al., 2013) para processar os dados através de regressão linear segmentada. R é um ambiente de software e linguagem de programação estatística construído para computação estatística e visualização de dados e lida com bibliotecas para análise de séries temporais. O ecossistema conta com um núcleo que gerencia todas as tarefas de download, upload e transformação, mostrando seu resultado em um painel visual.

O Hermes também faz uso extensivo da biblioteca Pandas. Esta biblioteca fornece estruturas de dados e ferramentas para trabalhar com conjuntos de dados comuns a aplicativos de computação matemática e científica, fornecendo métodos integrados e intuitivos para realizar manipulações e análises de dados (MCKINNEY et al., 2011). A núcleo principal da biblioteca Pandas é baseado em os objetos do tipo *Series* e *Dataframe*, que oferecem uma estrutura para organizar tipos diferentes de dados (e.g., textos, inteiros, decimais) em um único objeto de dados (BETANCOURT; CHEN, 2019). Em outras palavras, o *Dataframe* pode ser considerado um contêiner de dicionário de objetos do tipo *Series*.

4.1 Arquitetura do Ecossistema Hermes

A arquitetura do ecossistema Hermes é ilustrada na Figura 7, que resume as bibliotecas usadas em cada componente e como elas se relacionam. Nas subseções seguintes, apresentamos em detalhes as funcionalidades dos componentes.

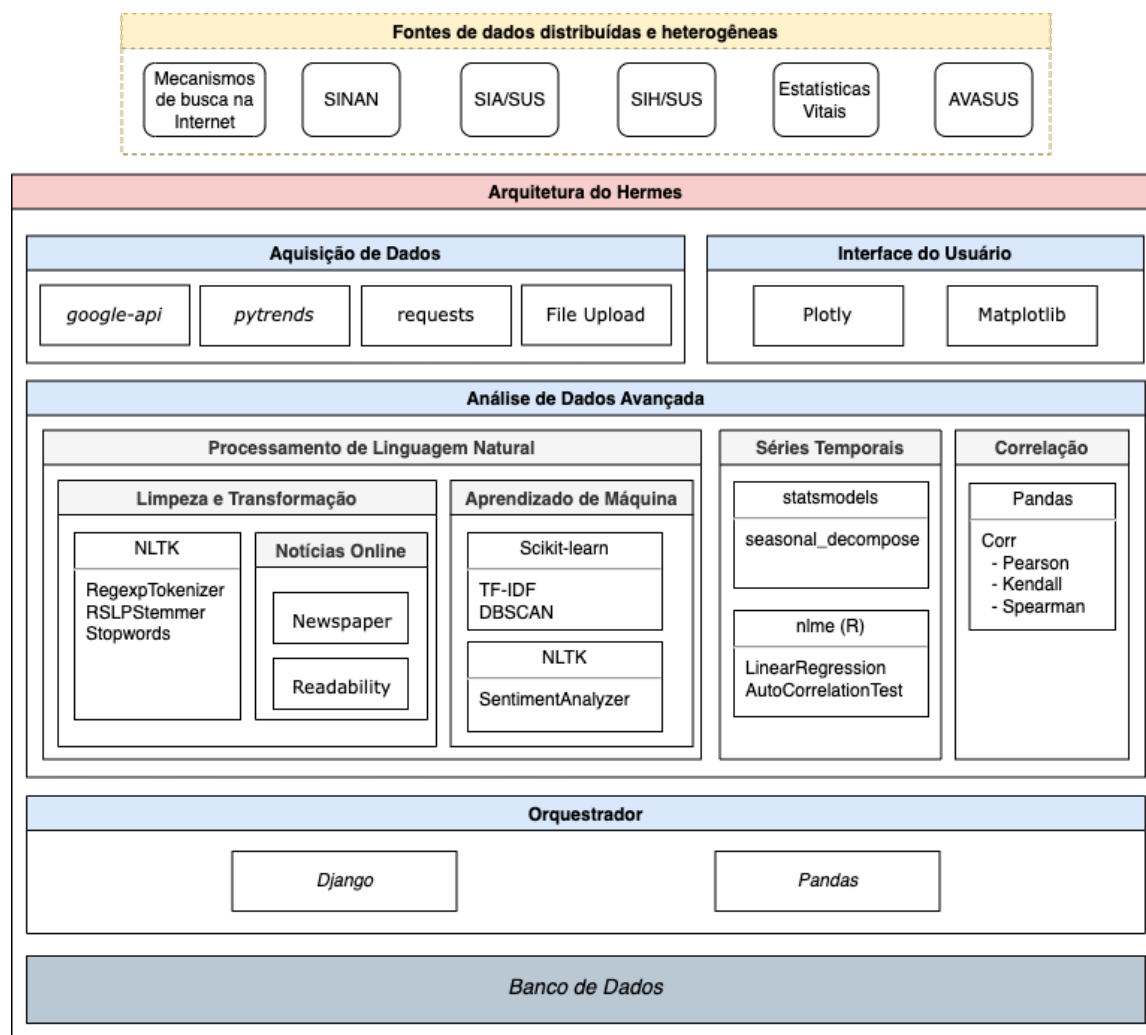


Figura 7 – Arquitetura do ecossistema Hermes em detalhes.

4.1.1 Orquestrador

Considerando o *framework* de implementação do Django, o módulo Orquestrador funciona como um controlador e é responsável pela comunicação entre todos os componentes. Assim, os comandos acionados pelo *stakeholder* são interceptados e coordenados pelo orquestrador, que os direciona para os demais componentes, processando-os através do módulo de análise, persistindo os dados transformados no banco de dados e apresentando-os de volta ao *stakeholder* por meio de um *dashboard*.

4.1.2 Aquisição de Dados

O módulo de aquisição de dados compreende bibliotecas Python responsáveis por obter dados fora do contexto do sistema e encaminhá-los, por meio do Orquestrador, para o módulo de análise.

No módulo de aquisição de dados, atualmente usamos as seguintes bibliotecas:

- **google-api-python-client**: Fornece uma API para serviços de pesquisa do Google, automatizando a recuperação de dados (páginas da web e notícias) para posterior processamento do Hermes. É importante observar que o retorno desses dados compreende os seguintes campos: Data em que a página foi indexada pelo Google, URL e trecho da página localizada.
- **requests**: É uma biblioteca HTTP que permite enviar solicitações HTTP/1.1 facilmente. Essa biblioteca obtém o conteúdo textual de cada página da web retornada pelo Google Search API. Assim, uma solicitação ao endereço URL é feita e um objeto HTML é retornado ao orquestrador. Outra utilização da biblioteca *requests* é obter dados disponíveis em sites abertos ao público em geral, como DataSUS¹. Por exemplo, um comando acionado pelo *stakeholder* pode obter informações (não estruturadas) sobre procedimentos ambulatoriais e hospitalares através da biblioteca *requests*. A transformação de dados não estruturados em um modelo estruturado, considerando a dimensão temporal, é realizada pelo Django em conjunto com a biblioteca Pandas, que será explicada posteriormente.
- **pytrends**: Através desta biblioteca, é possível automatizar o download de dados do Google Trends. Após obter os dados, o *framework* Django os transforma em um modelo e persiste no banco de dados.
- **File Upload**: É permitido o upload de arquivos no Hermes através da interface do usuário em formatos previamente definidos para processamento de dados, como dados de campanha e dados da plataforma AVASUS².

É importante notar que, apesar do **google-api-python-client** fornecer em seu módulo uma opção para retornar todas as páginas web indexadas com o termo de busca no período desejado, um bug foi reportado pela comunidade do Stackoverflow e permaneceu não resolvido durante o período de testes realizados pelo Hermes. Assim, o número de resultados obtidos pelo método ao pesquisar todo o período foi inconsistente com o resultado real. Para solucionar esse problema, o Hermes implementa uma rotina de busca que percorre o período solicitado dia a dia, ou seja, caso seja necessário buscar por páginas da web de 01/01/2019 a 31/12/2019, o método será chamado 365 vezes (uma para cada

¹ <<https://datasus.saude.gov.br/aceso-a-informacao/>>

² <<https://avasus.ufrn.br/>>

dia), retornando corretamente o número de páginas que devem ser encontradas no período total.

Na necessidade de obter dados de outras fontes, é necessário implementar no Hermes como obter os dados (web scraping ou API), onde obtê-los (endereço URL), quais parâmetros são necessários e que tipo de dados de retorno. A tabela 4 mostra as fontes de dados *online* previamente configuradas e disponíveis para uso do Hermes, descritas a seguir:

Tabela 4 – Fontes de dados *online* disponíveis e implementadas no Hermes.

Fonte de Dados	Dimensão	Tipo	Formato de Saída
Mecanismo de pesquisa personalizado do Google	Comunicação	API	JSON
Google Trends	Comunicação	API	pandas Dataframe
AVASUS	Educação	API	JSON
SIA/SUS	Vigilância Epidemiológica	Web Scraping	Tabela HTML
SIH/SUS	Vigilância Epidemiológica	Web Scraping	Tabela HTML
Estatística Vital	Vigilância Epidemiológica	Web Scraping	Tabela HTML

- Para implementar a solicitação através do Google Search API, usamos a biblioteca fornecida pelo Google, chamada Google Custom Search Engine (GCSE), que permite que qualquer pessoa crie e personalize termos de pesquisa para sua organização (GAVALI, 2015). Ela encapsula a URL de acesso através de um método chamado “list()”. Ao chamar o método *list*, é necessário informar os seguintes parâmetros: “developerKey” (chave que o Google fornece, autorizando o software fonte a realizar esta chamada), “search_term” (palavra-chave a ser pesquisado) e “lr” (a restrição de idioma para os resultados da pesquisa). Se a solicitação for bem-sucedida, o servidor responderá com um código de status HTTP 200 OK e os dados de resposta no formato JSON.
- Para dados disponíveis no Google Trends, usamos a biblioteca *pytrends*³. Esta é uma biblioteca não oficial para acesso ao Google Trends que permite uma interface simples para o download automático de relatórios do Google Trends. Da mesma forma, *pytrends* encapsula a chamada para a URL de acesso usando um método chamado “build_payload”. Para executar o método, é necessário informar os parâmetros “kw_list” (palavras-chave para obter os dados), timeframe (data de início ou datas específicas) e geo (geolocalização dos resultados). O formato do resultado é um objeto do tipo Pandas Dataframe.
- A terceira e quarta fontes de dados *online* disponíveis no Hermes (Tabela 4) são o Sistema de Informação Ambulatorial do Sistema Único de Saúde⁴ (SIA/SUS) e Sistema de Informações Hospitalares do Sistema Único de Saúde⁵ (SIH/SUS). O

³ <<https://pypi.org/project/pytrends/>>

⁴ <<https://datasus.saude.gov.br/aceso-a-informacao/producao-ambulatorial-sia-sus/>>

⁵ <<https://datasus.saude.gov.br/aceso-a-informacao/producao-hospitalar-sih-sus/>>

Ministério da Saúde brasileiro administra esses sistemas e seus dados estão disponíveis para consulta pública na Internet. No entanto, não encontramos uma API pública para obter esses dados. Assim, usamos Web Scraping para ler as tabelas HTML disponíveis publicamente e obter os dados. Os dados básicos de entrada são unidade da federação (ou região), mês e ano e procedimento ambulatorial. Uma tabela prévia e específica com todos os procedimentos ambulatoriais pode ser analisada antes do envio desses dados. A saída é uma tabela HTML com dados não estruturados (procedimento ambulatorial, unidade federativa, quantidade).

- Por fim, os dados sobre Estatísticas Vitais podem ser obtidos no Departamento de Informática em Saúde do Ministério da Saúde do Brasil (DATASUS), também disponível para consulta pública na *Internet* ⁶.

Embora os dados de séries temporais sejam o foco de nossa análise, é importante notar que os demais atributos recebidos pelo Hermes não são descartados. Pelo contrário, eles serão úteis para uma análise de dados mais detalhada. A Figura 8 apresenta o modelo de dados utilizado para armazenar as informações e possibilitar sua análise. Em vermelho, é possível observar as classes referentes à dimensão da campanha, em verde dados de educação, em azul claro dados de interesse ao longo do tempo na internet, em rosa as notícias *online* e em amarelo procedimentos em saúde (por exemplo, testes de triagem e casos notificados). Por fim, em roxo, podemos observar a estrutura de dados necessária para analisar a série temporal.

4.1.3 Análise de Dados Avançada

Após obter os dados, o orquestrador transforma os objetos retornados de consultas externas em um DataFrame e, em seguida, interage com o módulo de análise de dados avançada. Este módulo executa um subconjunto de ações com base no tipo de dados. Essas ações estão associadas ao pré-processamento, limpeza e transformação dos dados para o modelo de dados esperado e à análise temporal e de correlação. Em seguida, os seguintes submódulos são acionados dependendo do tipo de dados:

4.1.3.1 Limpeza e Transformação

Todas as rotinas de análise de informações textuais passam pelo módulo Limpeza e Transformação. Para isso, o Hermes usa uma plataforma python chamada NLTK (Natural Language Toolkit). Essa plataforma fornece interfaces para mais de 50 conjuntos de dados linguísticos (chamados corpora), recursos lexicais e bibliotecas de processamento de texto para classificação, tokenização, lematização, análise e raciocínio semântico.

⁶ <<https://datasus.saude.gov.br/nascidos-vivos-desde-1994/>>

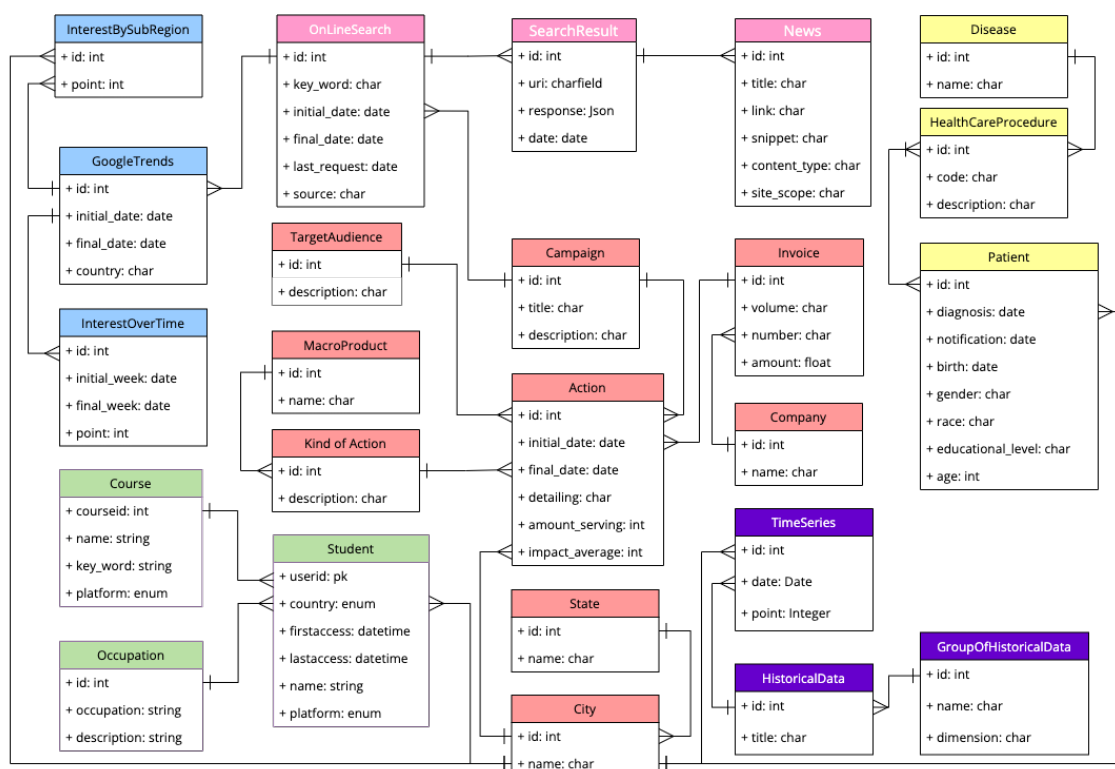


Figura 8 – Modelo de dados do Hermes abrangendo as dimensões implementadas.

Nesta etapa, as notícias são pré-processadas, visando homogeneizar e padronizar esse conteúdo, utilizando as mesmas etapas descritas na seção Limpeza de Dados do capítulo 3, transformando cada notícia em um conjunto de n-grams e compondo um dicionário de palavras representando cada texto.

Notícias Online

O submódulo Notícias Online é responsável por limpar o conteúdo HTML obtido através da biblioteca requests e armazenar o corpo das notícias que serão analisadas. No conteúdo HTML retornado pelas requisições, outros atributos são obtidos de forma não estruturada, como cabeçalho, menu lateral e rodapé.

Assim, o conteúdo HTML deve ser limpo e seus dados transformados de forma estruturada para serem analisados corretamente pelo módulo de análise avançada. Obter o conteúdo principal de uma página da Web corretamente depende de quão bem estruturada ela é. Isto está fora do escopo do Hermes e depende de cada site que postou a notícia. Portanto, mais de uma abordagem foi implementada: (i) Newspaper⁷ e (ii) Readability⁸. Essas abordagens são responsáveis por extrair o conteúdo principal de uma página web e remover as demais seções desnecessárias nesta análise.

Dessa forma, a abordagem que retorna mais conteúdo é responsável por entregar o corpo da notícia ao Orquestrador para armazenar no banco de dados. Por meio de

⁷ <<https://newspaper.readthedocs.io/en/latest/>>

⁸ <<https://pypi.org/project/readability-lxml/>>

uma interface visual, o ator pode verificar o conteúdo extraído da página web e realizar uma extração manual, se necessário. Por exemplo, a figura 9 ilustra uma notícia na parte central e o formulário à direita permite que o *stakeholder* analise a notícia. Os dados disponíveis são utilizados para caracterizar a notícia no sistema e excluí-la caso não atenda aos parâmetros esperados relacionados ao seu conteúdo.



Figura 9 – Interface para análise de notícias no ecossistema Hermes.

4.1.3.2 Aprendizado de Máquina

O submódulo de aprendizado de máquina é responsável por transformar conjuntos de texto pré-processados em informações relevantes por meio de mineração de dados. Três abordagens são implementadas neste módulo para extrair essas informações e análises adicionais por atores interessados.

Classificação de Texto

A classificação de texto é uma das tarefas críticas no aprendizado de máquina supervisionado. Para extrair a informação, transformamos cada n-grama, previamente definido no módulo *Limpeza e Transformação*, em um valor numérico usando a técnica Term Frequency – Inverse Document Frequency (TF-IDF). Essa técnica determina estatisticamente a importância de uma palavra em um corpus em relação a outros textos dentro do mesmo conjunto de dados. O peso da palavra para tal texto depende do número de vezes que ele aparece, mas é compensado pelas frequências de palavras em outros textos dentro do mesmo conjunto de dados (HONG et al., 2013). Portanto, palavras comuns

em todos os textos, como “isto”, “o que” e “se” têm uma classificação baixa, porque não significam muito para aquele texto em particular.

O Scikit-Learn é uma das bibliotecas de aprendizado de máquina Python mais conhecidas e utilizadas, dentre as muitas existentes, criada em código aberto e desenvolvida para apoiar e possibilitar o treinamento de diversas técnicas estatísticas e de aprendizado de máquina tanto para aprendizado supervisionado quanto não supervisionado.

O Hermes utiliza a biblioteca Scikit-Learn para aplicar a técnica TF-IDF e extrair informações de textos, visando identificar, por exemplo, a qualidade da notícia com base em categorias pré-definidas (definição de doenças; indicadores epidemiológicos; prevenção; teste rápido/diagnóstico; consequências; campanha e tratamento, e.g.)⁹. Para isso, um conjunto de notícias foi separado e etiquetado com base em tais categorias. A partir deste conjunto, chamado “amostra de treinamento”, Hermes verifica se outras notícias possuem as mesmas categorias. Assim, com base na quantidade de categorias encontradas em uma notícia, o Hermes define a qualidade dessa notícia da seguinte forma: baixa qualidade (0 a 2 categorias), média qualidade (3 a 4) e alta qualidade (5 ou mais).

É importante destacar que esta funcionalidade, implementada no ecossistema Hermes, é genérica o suficiente para abranger outras análises sob outro conjunto de categorias, uma vez que a “amostra de treinamento” com essas categorias é previamente definida.

Um novo estudo está previsto para ser realizado em breve, e o objetivo é identificar automaticamente as notícias em um conjunto de páginas da web. O critério planejado é identificar o 5W1H: “quando, o quê, quem, onde, por que e como”.

Agrupamento de Textos

Outro algoritmo essencial, contido na biblioteca Scikit-Learn, é DBSCAN (Density-Based Spatial Clustering of Applications with Noise). Este algoritmo tem como premissa criar grupos de dados que se assemelham, além de manter dados diferentes em diferentes grupos (ESTER et al., 1996). Como resultado, ele pode descobrir clusters de diferentes formas e tamanhos a partir de uma grande quantidade de dados analisando ruídos e discrepâncias.

O objetivo de agrupar dados no Hermes é identificar clusters de notícias para encontrar padrões que justifiquem a reprodução de determinado conteúdo jornalístico. Essa identificação de padrões é útil, pois permite manter o tema em pauta por um período prolongado, facilitando a definição de estratégias de comunicação.

A tarefa de agrupamento de notícias não é trivial e outros três algoritmos foram analisados, antes da definição do DBSCAN: t-distributed Stochastic Neighbor Embedding

⁹ Um estudo sobre a qualidade da notícia, com base em sete categorias, foi submetido como um artigo e está em análise no *Journal of Medical Internet Research*, intitulado: “How Online News Affects Public Health Response to Syphilis in Brazil: Text Mining Mediated Analysis”

(t-SNE) (HINTON; ROWEIS, 2002), Latent Semantic Indexing (LSI) (DEERWESTER et al., 1990) e Principal Component Analysis (PCA) (PEARSON, 1901).

O método t-SNE mapeia dados de alta dimensão em um mapa bi ou tridimensional e é baseado no método criado por Hinton e Roweis (2002). A ideia é manter objetos semelhantes próximos e diferentes distantes de acordo com as probabilidades atribuídas.

O segundo método verificado foi o LSI proposto pelo Deerwester et al. (1990), que é aplicado especificamente a um conjunto de documentos de texto. Além disso, utiliza a Singular Value Decomposition (SVD) como principal técnica de transformação.

Por fim, foi utilizada a técnica de redução de dimensionalidade PCA (PEARSON, 1901). Este método visa tentar criar uma nova representação de dados com uma dimensão menor, tentando preservar a variância entre eles. O cálculo do PCA é baseado no uso de autovalores e autovetores.

Realizamos técnicas de visualização de dados para observar os resultados dos quatro algoritmos. No entanto, os algoritmos t-SNE, LSI e PCE não apresentaram resultados satisfatórios, e as imagens mostraram um conjunto de notícias na parte central do gráfico, dificultando a separação dos grupos. Para garantir que esse cluster não fosse apenas um grupo de notícias, verificamos manualmente algumas dessas notícias e o resultado mostrou que a grande maioria era distinta.

Os resultados apresentados pelo DBSCAN, visualizados através das técnicas de redução de dimensionalidade, foram satisfatórios, pois foi possível observar grupos de notícias semelhantes. Além disso, esses grupos foram analisados por meio de uma interface visual no Hermes, onde foi possível comparar os grupos e analisar se os conjuntos apresentavam alguma discrepância.

As figuras 10 e 11 apresentam uma tela Hermes, que compara três notícias identificadas pelo DBSCAN incluídas no mesmo grupo. As figuras apresentam uma notícia original (lado esquerdo) e uma réplica (lado direito). Na Figura 10, uma notícia é identificada em 23 de novembro de 2018, e originada do site aids.gov.br, e outra contendo o mesmo conteúdo textual datada de 27 de novembro de 2018, identificada no blog do site. saude.gov.br. Os trechos vermelhos foram removidos na segunda notícia e o trecho verde foi incluído na segunda notícia.

Da mesma forma, a Figura 11 compara as notícias do site original (aids.gov.br) com as notícias secundárias, publicadas em 13 de dezembro de 2018, no site portal.fiocruz.br. É possível observar que, embora estejam em sites diferentes e com datas diferentes, as notícias são muito parecidas. Esse padrão de similaridade também foi identificado em outros grupos de notícias.

Realizamos um estudo operacional sobre agrupamento de textos, com o objetivo de (i) analisar o grande número de notícias relacionadas à sífilis no Brasil durante 2015–2019; e (ii) identificar quais notícias foram as mais divulgadas ou subsidiaram a produção de novas notícias, ou seja, notícias que efetivamente provocaram o efeito “intermedia agenda-

Visualizar Similaridade no Cluster #12 da Campanha "#SífilisNão!"

Notícia Original	Notícia Secundária
<p>Título: Teste, Trate e Cure: nova campanha de combate à sífilis chama atenção para a prevenção e o tratamento da doença</p> <p>Data: 23/11/18</p> <p>http://www.aids.gov.br/pt-br/noticias/teste-trate-e-cure-nova-campanha-de-combate-sifilis-chama-atencao-para-prevencao-e-o</p>	<p>Título: Nova campanha de combate à sífilis chama atenção para a prevenção e o tratamento da doença</p> <p>Data: 27/11/18</p> <p>http://www.blog.saude.gov.br/index.php/promocao-da-saude/53654-teste-trate-e-cure-nova-campanha-de-combate-a-sifilis-chama-atencao-para-a-prevencao-e-o-tratamento-da-doenca</p>
<p>1 SÍFILIS23</p> <p>2 11</p> <p>3 2018 - 09:4929</p> <p>4 11</p> <p>5 2018 - 16:56 Começa a ser veiculada nesta quarta-feira (22) a campanha Teste, Trate e Cure que tem como objetivo lembrar as pessoas sobre a importância da prevenção, do diagnóstico e do tratamento da sífilis</p> <p>6 A ação tem como alvo as gestantes e suas parcerias sexuais, e a população sexualmente ativa em geral e profissionais de saúde</p> <p>7 A campanha é realizada no âmbito do projeto Sífilis Não, uma parceria entre o Ministério da Saúde e a Universidade do Rio Grande do Norte que tem como objetivo, entre outros, ampliar o acesso da população ao diagnóstico e ao tratamento na rede básica de saúde</p>	<p>1 Começou a ser veiculada a campanha Teste, Trate e Cure que tem como objetivo lembrar as pessoas sobre a importância da prevenção, do diagnóstico e do tratamento da sífilis</p> <p>2 A ação tem como alvo as gestantes e suas parcerias sexuais, e a população sexualmente ativa em geral e profissionais de saúde</p> <p>3 A campanha é realizada no âmbito do projeto Sífilis Não, uma parceria entre o Ministério da Saúde e a Universidade do Rio Grande do Norte que tem como objetivo, entre outros, ampliar o acesso da população ao diagnóstico e ao tratamento na rede básica de saúde</p>

Figura 10 – Comparação entre notícias identificadas no mesmo cluster originadas de aids.gov.br e replicadas em blog.saude.gov.br.

Visualizar Similaridade no Cluster #12 da Campanha "#SífilisNão!"

Notícia Original	Notícia Secundária
<p>Título: Teste, Trate e Cure: nova campanha de combate à sífilis chama atenção para a prevenção e o tratamento da doença</p> <p>Data: 23/11/18</p> <p>http://www.aids.gov.br/pt-br/noticias/teste-trate-e-cure-nova-campanha-de-combate-sifilis-chama-atencao-para-prevencao-e-o</p>	<p>Título: Campanha de combate à sífilis chama atenção para a prevenção e tratamento da doença</p> <p>Data: 13/12/18</p> <p>https://portal.fiocruz.br/noticia/campanha-de-combate-sifilis-chama-atencao-para-prevencao-e-tratamento-da-doenca</p>
<p>1 SÍFILIS23</p> <p>2 11</p> <p>3 2018 - 09:4929</p> <p>4 11</p> <p>5 2018 - 16:56 Começa a ser veiculada nesta quarta-feira (22) a campanha Teste, Trate e Cure que tem como objetivo lembrar as pessoas sobre a importância da prevenção, do diagnóstico e do tratamento da sífilis</p> <p>6 A ação tem como alvo as gestantes e suas parcerias sexuais, e a população sexualmente ativa em geral e profissionais de saúde</p> <p>7 A campanha é realizada no âmbito do projeto Sífilis Não, uma parceria entre o Ministério da Saúde e a Universidade do Rio Grande do Norte que tem como objetivo, entre outros, ampliar o acesso da população ao diagnóstico e ao tratamento na rede básica de saúde</p>	<p>1 Começou a ser veiculada a campanha Teste, Trate e Cure que tem como objetivo lembrar as pessoas sobre a importância da prevenção, do diagnóstico e do tratamento da sífilis</p> <p>2 A ação tem como alvo as gestantes e suas parcerias sexuais, e a população sexualmente ativa em geral e profissionais de saúde</p> <p>3 A campanha é realizada no âmbito do projeto Sífilis Não, uma parceria entre o Ministério da Saúde e a Universidade do Rio Grande do Norte que tem como objetivo, entre outros, ampliar o acesso da população ao diagnóstico e ao tratamento na rede básica de saúde</p>

Figura 11 – Comparação entre notícias identificadas no mesmo cluster originadas de aids.gov.br e replicadas em portal.fiocruz.br.

setting”(MCCOMBS, 2015). Os resultados desta pesquisa são discutidos em um artigo científico, que está em revisão. Embora o autor desta tese não seja o primeiro autor do artigo, os métodos implementados, as descobertas e a discussão surgiram a partir desta tese de doutorado.

Análise de Sentimentos

O submódulo de análise de texto utiliza a biblioteca LeIA (Lexico for Adapted Inference) (ALMEIDA, 2018) que é uma adaptação do léxico e ferramenta para análise de sentimentos VADER (Valence Aware Dictionary and sEntiment Reasoner) adaptado para textos em português, com suporte para emojis e foco na análise de sentimentos de textos expressos em mídias sociais, mas funcional para textos de outros domínios, como notícias de saúde pública.

Nesta tese, implementamos essa biblioteca para entender o sentimento predominante nas notícias e como essas notícias impactam outras variáveis de interesse. Assim, identificamos que notícias que retratam a vulnerabilidade à doença e suas consequências induziram um aumento de exames na atenção primária à saúde nos anos de 2018 e 2019. Os resultados desta pesquisa são discutidos em um artigo científico ainda em revisão e, embora o autor desta tese não seja o primeiro autor do artigo, os métodos implementados, as descobertas e a discussão surgiram a partir desta tese de doutorado.

4.1.3.3 Análise de Séries Temporais

O submódulo de análise de séries temporais do Hermes possui duas técnicas para verificação dados ao longo do tempo: Análise de Decomposição e Regressão Linear Segmentada.

Análise de Decomposição

A análise de decomposição é a abordagem atualmente usada pelo Hermes para reduzir a tendência e dessazonalizar os dados de séries temporais. Essa abordagem pode separar os dados de séries temporais em seus componentes principais: tendência, sazonalidade e o resíduo aleatório restante. Assim, Hermes usa a biblioteca statsmodels do Python (SEABOLD; PERKTOLD, 2010) que fornece uma implementação do método de decomposição em uma função chamada “sazonal_decompose()”. O parâmetro necessário é a própria série, como um Dataframe do Pandas.

Após a decomposição dos dados, os gráficos são construídos e salvos para posterior análise em uma representação visual, com dados agrupados por tipo de gráfico (tendência e sazonalidade) e variáveis de interesse. A Figura 12 apresenta uma amostra de resultados de decomposição sazonal. A primeira subfigura representa os dados originais. A segunda subfigura representa as tendências, o valor crescente ou decrescente na série. A terceira

subfigura representa o componente de sazonalidade, ou seja, o ciclo de repetição na série. Por fim, a quarta subfigura representa o resíduo restante, o que significa o que sobra quando os componentes sazonais e de ciclo de tendência são subtraídos dos dados. Esses gráficos, com dados específicos dos estudos, serão explicados nos capítulos seguintes.

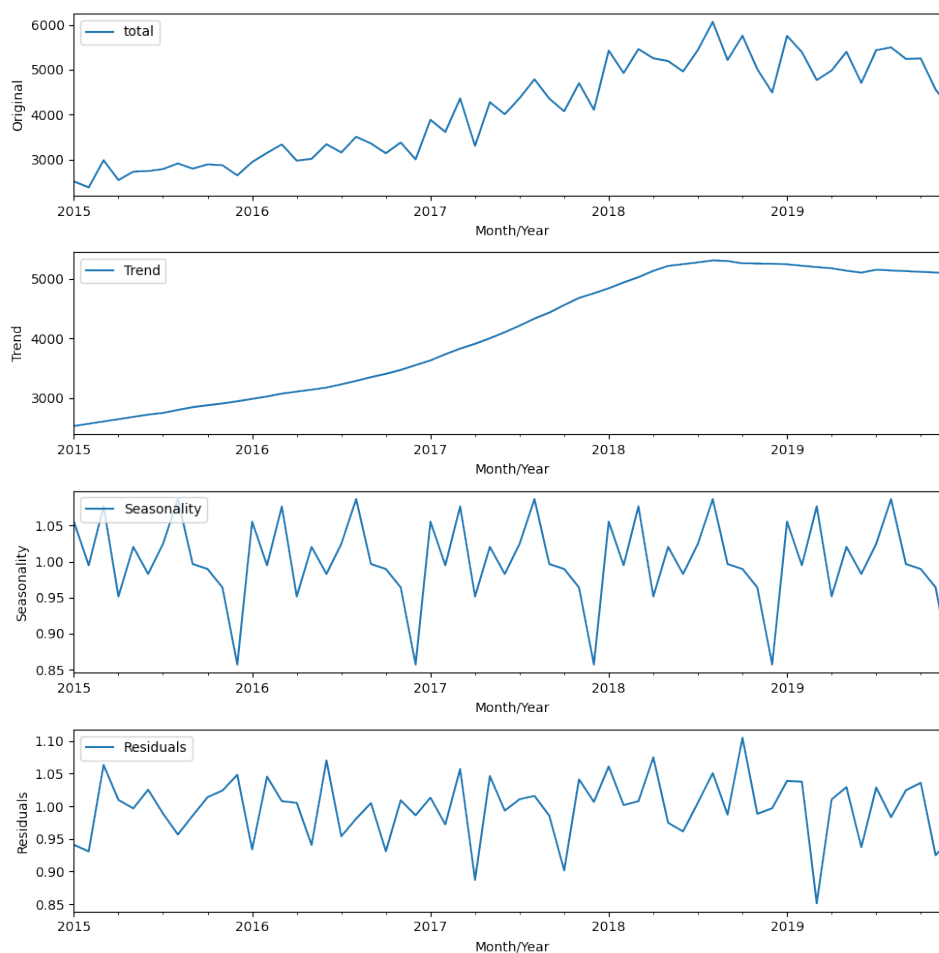


Figura 12 – O método de decomposição dividiu os dados em quatro partes: dados originais, tendência, sazonalidade e resíduo aleatório.

Análise de Regressão Segmentada

Conforme explicado no capítulo 2, a regressão linear segmentada foi escolhida para avaliar séries temporais por apresentar um modelo estatístico robusto para analisar intervenções de saúde pública, verificando mudanças ao longo do tempo e seus efeitos imediatos (WAGNER et al., 2002; ANSARI et al., 2003).

Assim, o ecossistema Hermes, de forma integrada, envia os dados para o componente R e aguarda o resultado em tabelas (coeficientes estatísticos) e gráficos para comparar antes e depois da intervenção e a verificação da mudança nos dados de tendência.

O componente R processa os dados de entrada e gera artefatos de saída para análise no painel gráfico do Hermes. As tabelas com os coeficientes estatísticos gerados ajudam a

entender o nível de impacto causado em termos numéricos, além de gerar Coeficientes de Determinação e Intervalos de Confiança, a fim de dar robustez aos resultados apresentados pelos gráficos e validar estatisticamente o estudo.

Para estudar e validar este módulo, escrevemos um artigo científico intitulado “Use of Interrupted Time Series Analysis in Understanding the Course of the Congenital Syphilis Epidemic in Brazil” publicado na *The Lancet Regional Health-Americas* (PINTO et al., 2022).

4.1.3.4 Análise de Correlação

Outra característica importante do Hermes é a análise de correlação estatística entre as variáveis de interesse. Usando a função “corr()” do Pandas, o Hermes pode calcular os coeficientes de correlação entre duas ou mais variáveis usando os métodos Pearson, Kendall Tau e/ou Spearman. O resultado desta função é uma matriz de correlação.

O dado de entrada para análise de correlação é um Dataframe, onde a primeira coluna corresponde a uma data, e as demais colunas correspondem às variáveis de interesse que a correlação deve ser medida. A tabela 5 apresenta um exemplo de um conjunto de dados hipotético para verificar a correlação entre notícias *online* (segunda coluna), testes de triagem (terceira coluna) e medicamentos (quarta coluna) durante um determinado período (primeira coluna).

Tabela 5 – Exemplo hipotético de conjunto de dados de entrada para processamento de análise de correlação.

Data	Notícias Online	Testes de Triagem	Medicamentos
01-01-2015	0	9.468	2.489
01-02-2015	1	8.695	1.854
01-03-2015	5	9.671	3.820
01-04-2015	0	9.927	2.341
01-05-2015	2	10.841	2.805

Da mesma forma, a Tabela 6 apresenta uma matriz de correlação hipotética, que exhibe os coeficientes de correlação para diferentes variáveis. Ele descreve a correlação entre todos os pares de valores possíveis na tabela. Os valores dos coeficientes podem variar de +1 a -1, onde +1 indica uma correlação linear positiva perfeita entre duas variáveis, -1 indica uma correlação linear negativa perfeita entre duas variáveis e 0 indica nenhuma correlação linear entre duas variáveis.

Neste exemplo hipotético, a célula destacada abaixo mostra que a correlação entre “Notícias Online” e “Testes de Triagem” é 0,8441, o que indica que eles estão fortemente correlacionados positivamente. Os coeficientes de correlação ao longo da diagonal da tabela são iguais a 1 porque cada variável está perfeitamente correlacionada consigo mesma.

Tabela 6 – Exemplo hipotético do resultado de processamento da análise de correlação entre variáveis de interesse.

	Notícias Online	Testes de Triagem	Medicação
Notícias Online	1.0000	0.8441	-0.3384
Testes de Triagem	0,2201	1,0000	-0,2045
Medicação	-0,3383	-0,2045	1,0000

No âmbito do Projeto “Sífilis Não!”, procurou-se encontrar uma correlação entre o número de testes de triagem realizados mensalmente na atenção primária à saúde e a qualidade das notícias veiculadas no mesmo período. É fundamental destacar que a variável “número de testes de triagem” foi subdividida em 7, um para cada tipo de exame para verificação de sífilis, disponível no portal DataSUS. Assim, a verificação foi mais precisa, identificando qual tipo de teste teve mais correlação com a qualidade da notícia.

Este módulo é fundamental para a compreensão da análise multidimensional apresentada, pois é através dele que podemos observar correlações que induzem as variáveis de interesse analisadas.

4.1.4 Interface do Usuário

A área de interface com o usuário é responsável por disponibilizar gráficos para facilitar o entendimento das informações apresentadas, auxiliando na identificação de padrões e comparação de resultados por meio de um *dashboard*. O Hermes usa a Biblioteca Plotly para fazer gráficos interativos e prontos para publicação, uma poderosa biblioteca no campo de visualização de dados. Além disso, o Hermes usa a biblioteca Matplotlib para criar algumas visualizações estáticas, como gráficos de tendência e sazonalidade.

4.2 Conclusões

Neste capítulo, apresentamos o *framework* e a arquitetura do ecossistema Hermes. É importante destacar que o aspecto chave do *framework* é a análise multidimensional das intervenções em saúde pública. No entanto, o Hermes não estabelece um único método de análise baseado em séries temporais, mas sim a extrapolação de vários métodos juntos, além de extrair dados de onde, aparentemente, não havia informações suficientes, como nas notícias online.

Além disso, a análise de correlação estatística para verificar a causalidade também é um fator chave entre as diversas dimensões propostas. Os capítulos seguintes apresentam estudos operacionais realizados com o apoio do ecossistema Hermes.

5 Uma Intervenção na Saúde Pública em Resposta à Epidemia de Sífilis

Este capítulo apresenta a intervenção de saúde pública contra a sífilis no Brasil, a campanha de comunicação de massa lançada para informar a população sobre a epidemia de sífilis e as fontes de dados utilizadas pelos estudos de caso para avaliar a resposta de saúde pública.

Esta intervenção em saúde foi amplamente explorada nesta tese com o objetivo de validar nosso *framework* e o ecossistema Hermes. Três estudos operacionais foram realizados com o projeto de intervenção “Sífilis Não!” para avaliar os dados produzidos ou impactados pelas ações realizadas.

5.1 Contextualização

Para combater a epidemia de sífilis no Brasil, em outubro de 2016, o Ministério da Saúde (MS) lançou a Agenda de Ações Estratégicas para Redução da Sífilis no Brasil ([TRANSMISSÍVEIS, 2017](#)). Esta Agenda estabeleceu uma lista de prioridades em colaboração com a Organização Pan-Americana da Saúde OPAS e outras instituições como universidades, agências internacionais e representantes estaduais e municipais. Além disso, tal Agenda suscitou uma emenda parlamentar com recursos específicos para implementar uma resposta rápida no Sistema Único de Saúde (SUS).

A Agenda apresenta seis eixos principais (ações e atividades) para a redução da sífilis no Brasil: (i) Educomunicação; (ii) Qualificação da Informação Estratégica; (iii) Fortalecimento da parceria entre o MS e outros atores; (iv) Ampliação dos Comitês de Pesquisa em Transmissão Vertical do HIV (Vírus da Imunodeficiência Humana), Sífilis e Hepatites Virais; (v) Fortalecimento das Redes de Atenção à Saúde; e (vi) Resposta Rápida à Sífilis nas Redes de Atenção à Saúde.

No final de 2017, para apoiar os objetivos da Agenda, o MS convidou os gestores estaduais e municipais a aderirem à estratégia nacional de combate à sífilis por meio do “Projeto de Resposta Rápida à Sífilis nas Redes de Atenção”, conhecido como Projeto “Sífilis Não!”. O Projeto, desenvolvido em cooperação com a Universidade Federal do Rio Grande do Norte (UFRN), mediante o Termo de Execução Descentralizada (TED) número 54/2017, visa reduzir a sífilis adquirida, sífilis em gestantes e sífilis congênita através da ampliação da cobertura de diagnóstico (por meio de testes rápidos) e tratamento oportuno e adequado de gestantes e parceiros sexuais em situação de pré-natal, parto ou aborto. Por meio do Projeto, a epidemia de sífilis tem sido enfrentada através de duas linhas estratégicas: (1) reforço das ações universais do SUS e (2) implementação de ações

específicas para 100 municípios escolhidos pelo MS como prioritários para a resposta à sífilis congênita, que representaram 68,95% do número de casos de sífilis congênita no Brasil em 2015.

A linha universal da intervenção incluiu a aquisição e distribuição de insumos para exames e tratamento (penicilina cristalina e benzatina), ampliação da rede de laboratórios de IST e salas de situação para vigilância epidemiológica, estratégias de educomunicação (VALENTIM *et al.*, 2021), intervenções sociais e campanhas de conscientização realizadas para enfrentar a sífilis nesse período.

Paralelamente, foi concedido apoio institucional aos municípios prioritários, onde os Apoiadores de Pesquisa e Intervenção (RIS) realizaram um trabalho específico com os gestores locais de saúde. O papel dos Apoiadores é potencializar as ações nos municípios prioritários e criar um vínculo entre o Ministério da Saúde e os atores locais (gestores municipais, profissionais de saúde e a população) (SAÚDE., 2020b).

Assim, este trabalho proporcionou cooperação técnica para fortalecer as ações de prevenção voltadas à redução da transmissão vertical da sífilis, como a implementação de comitês focados na investigação de casos de sífilis e reforço do planejamento local de resposta no eixo do Projeto. Os detalhes das intervenções do projeto e cronogramas relevantes são apresentados no Apêndice (Figuras 36 a 40).

5.2 Uma Campanha de Mídia de Massa como Parte da Intervenção de Saúde Pública

Para sensibilizar a população no combate à epidemia de sífilis no país, o MS lançou uma campanha nos meios de comunicação social como parte da intervenção de saúde pública. A campanha “Lembre de se cuidar – Sífilis: #TesteTrateCure”, mais conhecida como “Sífilis Não!”, objetivou relembrar a importância da prevenção da sífilis, destacando os números da epidemia; reforçar a importância do diagnóstico precoce e emergente do tratamento adequado, bem como da prevenção com uso de preservativo para prevenir a transmissão sexual da doença; enfatizar a importância do conceito: Testar, tratar e curar; e desenvolver diferentes estratégias e materiais para levar a informação em linguagem clara e direcionada de acordo com a população.

A campanha “Sífilis Não!” carrega o conceito de acesso universal ao Testar, Tratar e Curar que visa alertar as pessoas para a realização do teste rápido de sífilis ou VDRL em qualquer unidade básica de saúde (UBS) do Sistema Único de Saúde (SUS). Se o resultado for positivo, o paciente deve ser tratado na UBS com o medicamento Penicilina Benzatina e, após o devido acompanhamento médico, ser considerado definitivamente curado.

Para execução desta campanha foi criada uma Câmara Técnica de Comunicação, com a participação de membros do Laboratório de Inovação Tecnológica em Saúde da UFRN (LAIS/UFRN), do Núcleo de Estudos em Saúde Coletiva da UFRN (NESC/UFRN),

representantes da Secretaria de Vigilância em Saúde do Ministério da Saúde (SVS/MS) e da Divisão de Publicidade e Promoção Institucional do Ministério da Saúde.

Essa campanha foi realizada entre novembro de 2018 a maio de 2019. Nesse período, uma grande quantidade de material foi produzida e divulgada em diversos meios com abordagens específicas para diferentes públicos com ações pela televisão, rádio, plataformas de streaming, mídia impressa, revistas, transmissões ao vivo em eventos, cartazes, cartilhas informativas e adesivos. Além disso, na Internet, houve a divulgação de conteúdos relacionados em sites voltados especificamente para gestantes e outros voltados à cobertura de notícias, redes sociais, aplicativos de relacionamento e inserção de conteúdos em páginas digitais de revistas. Além disso, dez influenciadores digitais fizeram postagens (patrocinadas) em suas redes sociais, conforme mostrado na Figura 13.



Figura 13 – Anúncios da Campanha “Sífilis Não!” produzidos e divulgados pela televisão, plataformas de streaming, mídia impressa, revistas e eventos.

5.3 Fontes e Dados Heterogêneos

Para avaliar a intervenção de saúde pública descrita neste capítulo, analisamos várias fontes e dados heterogêneos. Dados da campanha “Sífilis Não!” foram obtidos por meio de cooperação interna entre a Universidade Federal do Rio Grande do Norte (UFRN) e o Ministério da Saúde. Tais dados foram inseridos manualmente através de formulários usando o ecossistema Hermes. Os dados de notícias *online* foram obtidos por meio da API

de pesquisa do Google. O interesse ao longo do tempo foi coletado usando o Google Trends. Os dados de cursos *online* foram coletados através da plataforma AVASUS¹. Os testes de triagem foram obtidos no SIA/SUS via web scraping. As distribuições de medicamentos foram obtidas através do SISMAT (Sistema Integrado de Gerenciamento de Materiais) fornecido pelo MS. Os casos de notificação e os registros de nascidos vivos foram extraídos do SINAN (Sistema de Informação de Agravos de Notificação), disponível na página do MS (SAÚDE., 2020a).

A tabela 7 mostra todos os dados obtidos pelo Hermes. Após o processamento completo do ciclo de vida dos dados, o Hermes disponibiliza as informações através de painéis de visualização para os tomadores de decisão analisarem e obterem insights sobre suas áreas.

Tabela 7 – Dados obtidos pelo ecossistema Hermes.

Dimensão	Dados	Fonte	Forma de obtenção	Período
Campanha	Dados de Campanha	Faturas	Hermes por meio de uma interface com o usuário	Nov, 2018 a Mar, 2019
Comunicação	Notícias Online	Google Search	Hermes consumindo API	Jan, 2015 a Dez, 2019
Comunicação	Interesse ao Longo do Tempo	Google Trends	Hermes consumindo API	Jan, 2017 a Dez, 2019
Educação	Dados de Cursos Online	AVASUS	Hermes via arquivos CVS	Fev, 2019 a Set, 2020
Vigilância Epidemiológica	Testes Sorológicos	SIA/SUS/MoH	Hermes via Web Scraping	Jan, 2017 a Dez, 2019
Vigilância Epidemiológica	Distribuição de Medicamentos	SISMAT/MoH	Hermes via Arquivo CVSs	Mar, 2016 a Dez, 2020
Vigilância Epidemiológica	Casos de Notificação	SINAN/MS	Hermes via Arquivo CVSs	Jan, 2015 a Dez, 2019
Vigilância Epidemiológica	Registros de Nascidos Vivos	SINAN/MS	Hermes via Web Scraping	Jan, 2015 a Dez, 2019

Nos capítulos seguintes, discutimos nossos métodos, descobertas e implicações ao analisar esses dados de uma perspectiva multidimensional para avaliar uma intervenção em saúde pública.

¹ <<https://avasus.ufrn.br/>>

6 Estudo Analítico

Este capítulo apresenta os resultados obtidos pela análise da campanha “Sífilis Não!”, considerando quatro dimensões: campanha, comunicação, educação e vigilância epidemiológica. A análise descritiva realizada neste estudo foi apoiada por nosso *framework* de análise multidimensional. Analisamos a campanha “Sífilis Não!” e sete fontes de dados diferentes. Nas seções a seguir, esclarecemos nossas questões de pesquisa e como nosso *framework*, apoiado pelo ecossistema Hermes, nos ajudou a entender esses dados.

O texto contido neste capítulo foi organizado como um artigo intitulado “Analyzing the Reach of Public Health Campaigns Based on Multidimensional Aspects: the case of the syphilis epidemic in Brazil” e foi publicado no periódico BMC Public Health (PINTO et al., 2021).

6.1 Análise Experimental

Esta análise experimental foi conduzida pelas seguintes questões de pesquisa e foram desenvolvidas usando pelo menos uma questão por dimensão:

- **QP1 - Como as ações da campanha foram agrupadas e distribuídas ao longo do tempo?** Raciocínio: Essa questão ajuda a entender se a campanha abrangeu todo o período ou se há lacunas entre as datas. As lacunas entre as datas podem inviabilizar um estudo de séries temporais que busque correlacionar o alcance diário das ações. Além disso, entender como as ações foram agrupadas ajuda a entender o tipo de público procurado.
- **QP2 - Como se comportou o interesse da população pelo tema ao longo do tempo na Internet? Houve aumento de notícias *online* espontâneas sobre o tema na Internet?** Raciocínio: O comportamento de busca na *Internet* por determinado tema na *Internet* pode ser útil para avaliar o interesse da população em informações sobre saúde, além de possibilitar a comparação do interesse antes e depois da campanha de conscientização. Além disso, as notícias *online* espontâneas podem apoiar ações de campanha, compartilhando informações de forma gratuita.
- **QP3 - A campanha ajudou a promover o aprendizado técnico ou científico, engajando profissionais de saúde, professores e alunos com o tema?** Raciocínio: O engajamento de profissionais de saúde, professores e alunos pode contribuir para melhorar a formação profissional de perspectiva da educação permanente.
- **QP4 - A campanha atingiu seu objetivo principal de estimular a população a realizar exames para diagnóstico de sífilis, tratamento e cura?**

Raciocínio: Comparar o número de exames de sífilis, distribuição de medicamentos e casos de notificação durante o período pré e pós-campanha pode nos trazer o real alcance relacionado à vigilância epidemiológica.

6.2 Evidências da Mudança de Comportamento na Saúde

As questões de pesquisa foram respondidas de acordo com os resultados produzidos pelo Hermes.

QP1 - Como as ações da campanha foram agrupadas e distribuídas ao longo do tempo?

As ações de comunicação foram agrupadas em sete macrocategorias: Mídia Digital, Mídia Outdoor, Mídia Impressa, Rádio, Mídias Sociais, Plataformas de Streaming e Televisão. Essas macrocategorias são as mesmas utilizadas nos relatórios e documentos oficiais da campanha “Sífilis Não!”, conforme segue:

- Mídia Digital: Mensagens exibidas em mídia digital, em sites ou aplicativos da Internet, por meio de banners ou outros formatos publicitários compostos por texto, imagens, vídeos ou áudios;
- Mídia Outdoor: Cartazes em vitrines, pontos de ônibus, outdoors e outras formas de mídia implantadas em espaços urbanos de acesso público;
- Mídia Impressa: Anúncios impressos em jornais e revistas de consumo;
- Rádio: Áudios pré-gravados (spots e depoimentos) veiculados por rádio;
- Mídias Sociais: Mensagens postadas em redes sociais (Facebook, Instagram e Twitter) por influenciadores ou publicações patrocinadas do Facebook e Instagram direcionadas a um público específico (por exemplo, por idade, escolaridade, ocupação);
- Plataformas de Streaming: Vídeos patrocinados em canais do Youtube e áudios pré-gravados veiculados como anúncios no Spotify;
- Televisão: Vídeos pré-gravados sobre sífilis (spots e depoimentos) veiculados pela televisão.

A figura 14 mostra como as macrocategorias da campanha “Sífilis Não!” foram distribuídas ao longo do tempo. As ações iniciaram em 22/11/2018 e terminaram em 31/03/2019 com gaps, ou seja, em alguns períodos não houveram ações todos os dias.

Mídia Outdoor foi a macrocategoria com maior número de ações diárias acumuladas ($n = 4.193$), seguida por Mídia Digital ($n = 1.566$), Rádio ($n = 825$), Televisão ($n = 675$),

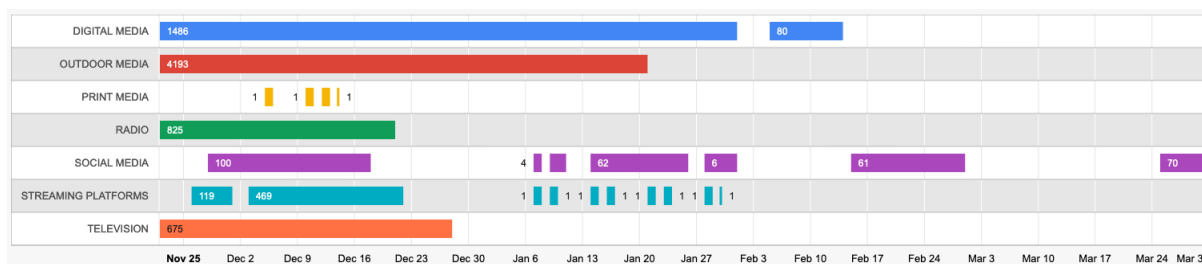


Figura 14 – Linha do tempo de ações que mostra como as macrocategorias da campanha “Sífilis Não!” foram distribuídas.

Plataformas de Streaming (n = 596), Mídias Sociais (n = 318) e Mídia Impressa (n=4). As lacunas podem ser visualizadas nas macrocategorias Mídia Digital, Mídia Impressa, Mídia Social e Plataformas de Streaming.

A figura 15 mostra a quantidade diária de ações por macrocategoria. Embora as ações tenham sido realizadas até o final de março de 2019, é possível observar uma redução abrupta no número de ações diárias realizadas a partir de 23 de dezembro de 2018, quando cessaram as ações de Televisão, Rádio e Mídia Impressa. Entre 1 e 5 de fevereiro, não houve ações nem entre 1 e 25 de março.

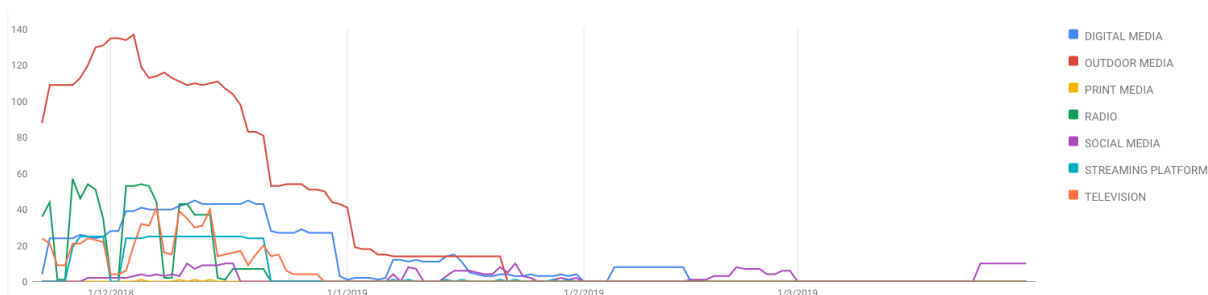


Figura 15 – Quantidade diária de ações por macrocategoria.

QP2 - Como se comportou o interesse da população pelo tema ao longo do tempo na Internet? Houve aumento de notícias *online* espontâneas sobre o tema na Internet?

Para analisar o interesse da população ao longo do tempo na *Internet* obtivemos dados de consulta do Google Trends. Esses dados representam a atividade de pesquisa usando o termo 'sífilis' no Brasil entre janeiro de 2017 e dezembro de 2019. O Google Trends fornece dados de série temporal relacionados a um termo de pesquisa específico realizado no mecanismo de busca do Google. Ele representa o interesse de pesquisa em relação ao ponto mais alto no gráfico para a região e hora especificadas (NUTI et al., 2014).

A figura 16 mostra todos os dados semana a semana. As linhas verticais pontilhadas mostram o período da campanha. O maior volume de atividade de pesquisa foi observado

durante a primeira semana imediatamente após o início das ações (25 de novembro e 7 de dezembro). Para entender melhor esses dados, a análise de decomposição pode ser usada para eliminar tendências e dessazonalizar dados de séries temporais obtidos do Google (PALIT; POPOVIC, 2006).

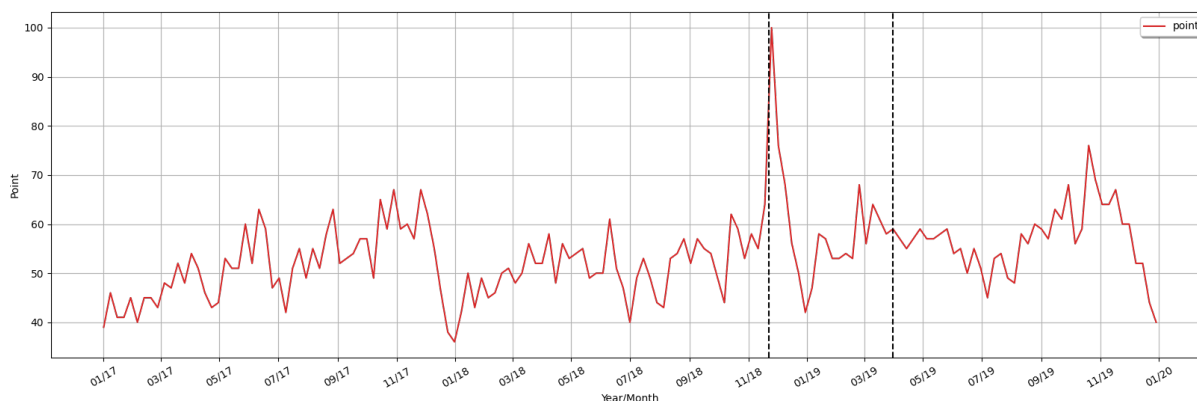


Figura 16 – Atividade de pesquisa relativa ao termo 'sífilis' no Google entre 2017 e 2019.

As tendências podem ser visualizadas quando vemos uma inclinação crescente ou decrescente nos dados de séries temporais, enquanto a sazonalidade está presente quando observamos um padrão de repetição específico em intervalos regulares (TAYEB; KARABATAK; VAROL, 2020). A Figura 17 mostra o gráfico dos dados de tendência e sazonalidade. O gráfico sazonal mostra um padrão de repetição com picos nas buscas pelo termo em outubro, novembro e dezembro, com queda abrupta em janeiro e fevereiro. O gráfico de tendência mostra uma inclinação crescente a partir de junho de 2018, bem como uma inclinação suavemente decrescente em julho de 2019, que continuou subindo depois disso.

Além disso, é importante mencionar que em maio, junho e julho de 2018 foram realizadas ações de capacitação e seminários regionais com gestores dos territórios e estados, bem como com Apoiadores de Pesquisa e Intervenção foram selecionados para atuar em 72 dos 100 municípios prioritários identificados no escopo do Projeto, distribuídos em todas as regiões do Brasil (LUCAS et al., 2019). Essas atividades foram organizadas para promover discussões, trocar experiências e apoiar a campanha, e certamente impulsionaram a busca pelo termo sífilis nos mecanismos de busca da Internet.

Em relação às notícias espontâneas relacionadas ao tema na Internet, o ecossistema Hermes obteve notícias *online* indexadas pelo Google, utilizando o Google Custom Search Engine (GCSE), entre janeiro de 2015 e dezembro de 2019.

A figura 18 mostra a quantidade de postagens de notícias *online* indexadas com o termo 'sífilis' pelo Google entre janeiro de 2015 e dezembro de 2019, no Brasil. É possível observar um aumento gradual no número de notícias *online* por mês e ano. Todos os meses do ano de 2019 (pós- campanha) tiveram mais notícias do que os meses dos anos anteriores, principalmente o mês de outubro. Possivelmente porque o terceiro sábado de outubro foi

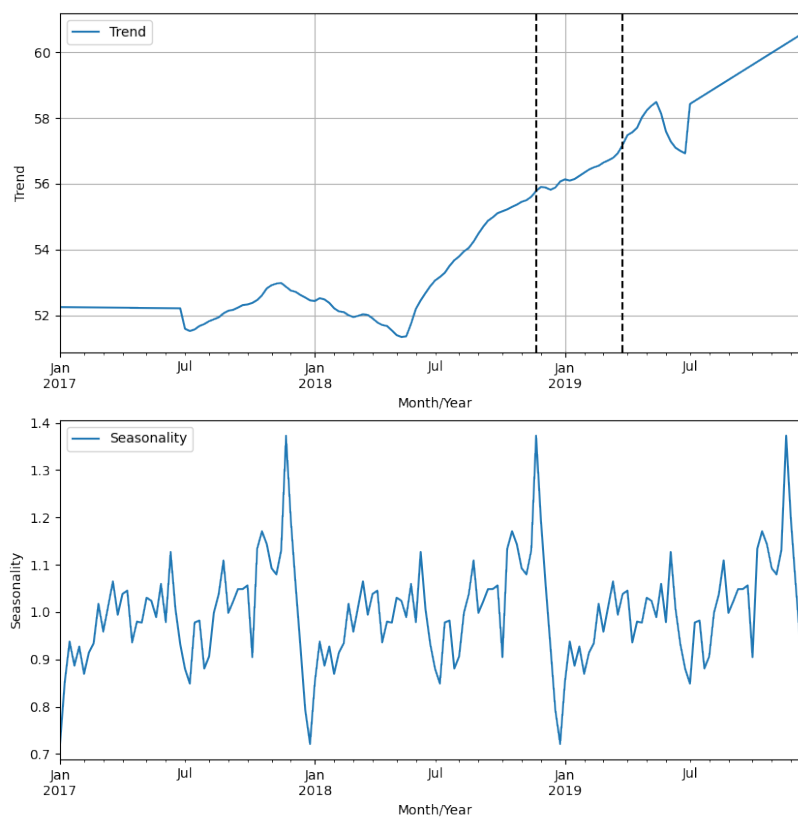


Figura 17 – Atividade de pesquisa relativa ao termo 'sífilis' decomposta em tendência e sazonalidade.

instituído por lei como o Dia Nacional de Combate à Sífilis e Sífilis Congênita (FEDERAL, 2017). A lei buscou estimular a participação dos profissionais de saúde nas atividades da campanha, para enfatizar a importância do diagnóstico e tratamento adequado da sífilis.

Decompondo esses dados em um gráfico de tendência e sazonalidade (Figura 19), o Hermes mostra uma inclinação ascendente a partir de março de 2018, bem como um pico recorde em outubro de 2019. Em relação à sazonalidade, há um claro padrão de repetição com um pico de notícias em outubro e novembro, e um declínio em janeiro e fevereiro.

QP3 - A campanha ajudou a promover o aprendizado técnico ou científico, engajando profissionais de saúde, professores e alunos com o tema?

Na perspectiva da aprendizagem ao longo da vida dos profissionais de saúde, o número de publicações, dissertações e teses pode ser visto como um poderoso indicador do interesse de docentes e discentes pelo tema. No entanto, para este caso real, considerando a abordagem de séries temporais desta pesquisa, outros indicadores de educação foram mais oportunos.

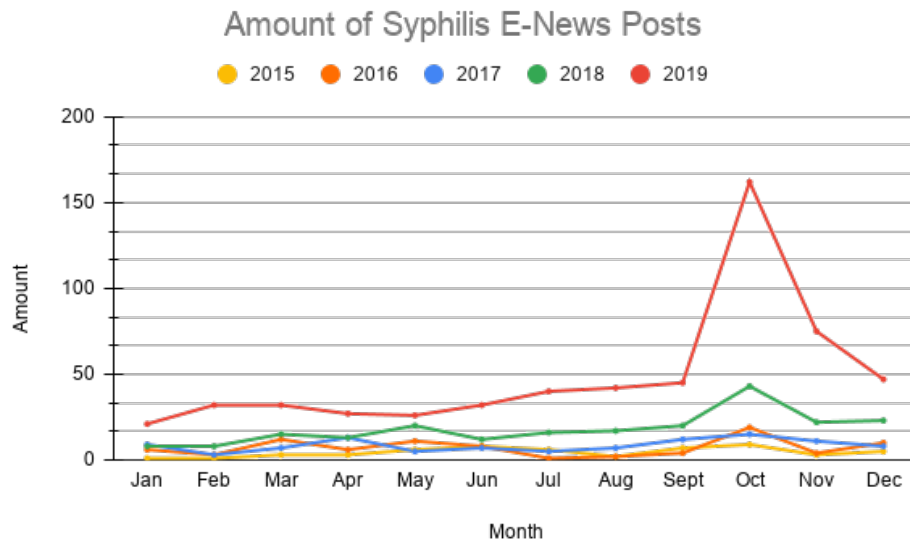


Figura 18 – Quantidade de postagens de notícias *online* sobre sífilis indexadas pelo Google entre janeiro de 2015 e dezembro de 2019.

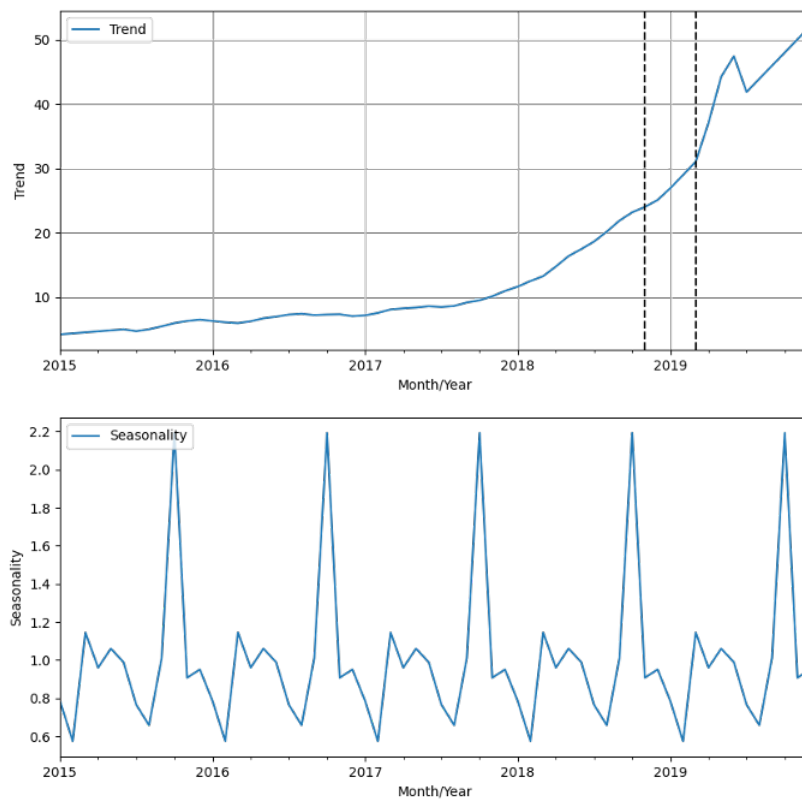


Figura 19 – Notícias *online* indexadas pelo Google usando o termo 'sífilis' decompostas em tendência e sazonalidade.

Dois indicadores oportunos relacionados à aprendizagem ao longo da vida são o número de MOOCs oferecidos pelo governo sobre sífilis e os alunos matriculados neles. Obtemos resultados da plataforma AVASUS, desenvolvida pela Universidade Federal do Rio Grande do Norte, por meio do Laboratório de Inovação Tecnológica em Saúde (LAIS) que visa promover a educação a distância através de ferramentas de ensino e aprendizagem *online* (LOPES¹ et al., 2018).

O primeiro curso *online* sobre sífilis foi lançado no AVASUS em fevereiro de 2019. A figura 20 mostra o número de alunos matriculados nos Cursos relacionados à sífilis no AVASUS, entre fevereiro de 2019 e setembro de 2020. O eixo horizontal representa o tempo decorrido, enquanto o eixo vertical representa o número de alunos matriculados no mês.

No primeiro semestre de 2019, havia apenas dois cursos sobre sífilis (C01 e C02) com média mensal de 310 alunos matriculados, considerando os dois cursos. No segundo semestre de 2019, foram adicionados três novos cursos (C03, C04 e C05), elevando a média mensal de alunos matriculados para 737. No semestre seguinte, 2020.1, foram adicionados 6 novos cursos à plataforma *online* (C06, C07, C08, C09, C10 e C11), elevando a média mensal para 1.907 (o que representa uma taxa de crescimento de 515,16%, quando comparado ao mesmo período do ano anterior - 2019,1).

Por fim, durante os meses de julho, agosto e setembro de 2020 (segundo semestre), foi criado um novo curso de sífilis (C12) e a média mensal de alunos matriculados foi de 1.776, o que representa uma taxa de crescimento de 140,98% em relação ao segundo semestre de 2019. Esses 12 cursos atingiram 22.744 alunos matriculados até setembro de 2020. A tabela 8 mostra a taxa de crescimento de alunos matriculados em relação aos semestres letivos.

Tabela 8 – Taxa de crescimento de alunos matriculados por semestre letivo.

Semestre escolar	Alunos matriculados (média mensal)	Taxa de Crescimento
2019.1	310	..
2019.2	737	..
2020.1	1.907	515,16% (em comparação com 2019.1)
2020.2	1.776	140,98% (em comparação com 2019.2)

Os cursos com maior número de alunos matriculados foram “C01 - Sífilis: senta que lá vem informação!” (n=4.284, desde=fev/19), “C04 - Sífilis congênita: do pré-natal ao tratamento ambulatorial” (n=5.248, desde=Nov/19), e “C05 - Sífilis: patogênese, resposta imune e desenvolvimento de métodos diagnósticos” (n=4.487, desde=Dez/19). Esses três cursos atingiram 14.019 alunos matriculados e atuam exatamente no eixo central das ações do Projeto “Sífilis Não!”, em especial, no que diz respeito ao tratamento e cura dos pacientes.

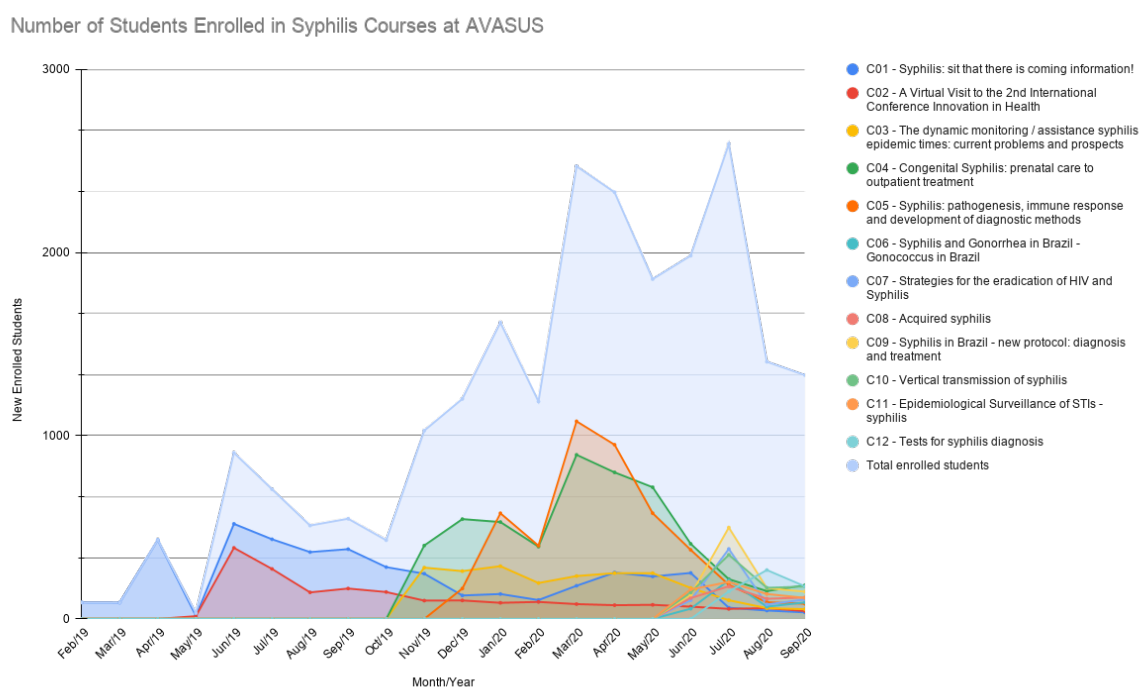


Figura 20 – Número de alunos inscritos nos cursos relacionados à sífilis no AVASUS, entre fevereiro de 2019 e setembro de 2020.

Os dados mostram que a dimensão educação atua como fator-chave na resposta à sífilis. O engajamento de profissionais de saúde, professores e alunos pode ser considerado um importante indicador de transformação da educação em saúde para a sífilis. Como reação em cadeia, os profissionais de saúde com conhecimento dos efeitos adversos da sífilis na vida das pessoas, orientam a população para o correto diagnóstico, tratamento e cura da doença. O Hermes mostra que os indicadores educacionais podem ser uma valiosa contribuição para o fortalecimento do sistema de saúde.

QP4 - A campanha atingiu seu objetivo principal de estimular a população a realizar exames para diagnóstico de sífilis, tratamento e cura?

Para analisar o alcance da campanha na perspectiva da vigilância epidemiológica, analisamos três indicadores: testes de triagem, distribuição de medicamentos e taxas de notificação de casos.

Os dados sobre teste e diagnóstico de sífilis foram coletados entre janeiro de 2015 e dezembro de 2019 do SIA/SUS. A escolha de incluir dois anos anteriores à nossa abordagem de séries temporais se baseia em sua importância para explicar o aumento gradual da testagem de sífilis observado antes de 2017 nas UBS, que passa a ser sustentável a partir de então. Utilizou-se como indicador central o número de exames sorológicos (testes não treponêmicos e treponêmicos) para sífilis nas UBS do Brasil, segundo ano.

Na Figura 21, podemos observar uma quantidade discreta de testes em 2015 ($n=518.859$) e 2016 ($n=911.420$) e um aumento mais óbvio a partir de 2017 ($n=1.448.364$), 2018 ($n=2.068.184$) que continuou até o final de 2019 ($n=2.533.571$). A partir de 2015 é evidente o aumento dos testes de sífilis no Brasil. Estes já vinham ocorrendo de forma gradual e sustentável.

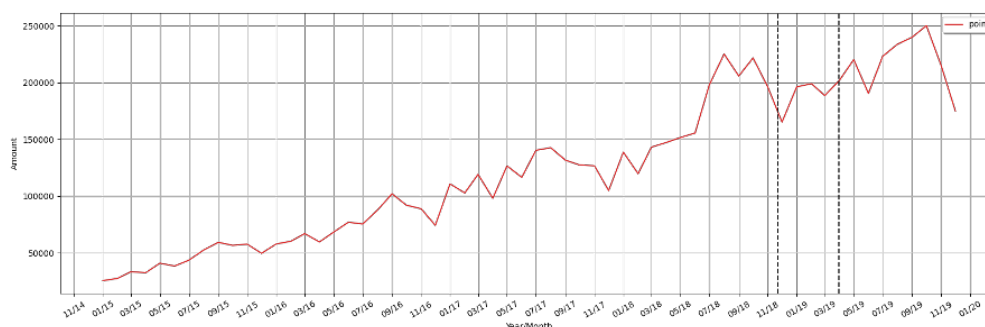


Figura 21 – Quantidade de teste de triagem realizado pelo serviço de atenção primária.

Para analisar o aumento da testagem, é necessário verificar o crescimento populacional no mesmo período. Portanto, a Tabela 9 mostra a população por ano, o número de exames realizados, a taxa de testagem por 1.000 habitantes, o aumento da taxa de testagem considerando a população e o percentual de aumento.

Assim, tendo como base o ano de 2015, o aumento mais significativo é observado nos anos de 2018 (289,09%) e 2019 (375,18%), contabilizando inclusive o crescimento populacional do país. Isso significa que 3,89 vezes mais indivíduos foram testados em 2018 e 4,75 vezes mais em 2019, em todo o país. Estes últimos dois anos se destacam, pois se enquadram no período de execução do Projeto “Sífilis Não!”.

Tabela 9 – Exames realizados por ano e taxa de crescimento por 1.000 habitantes.

Ano	População	Testes realizados	Taxa de testes por 1.000 habitantes	Aumento da testagem por 1.000 habitantes considerando o tamanho da população em relação a 2015	Aumento percentual de testagem por 1.000 habitantes considerando o tamanho da população em relação a 2015
2015	204.500.000	518.859	2,54	-	-
2016	206.200.000	911.420	4,42	1,74	74,21%
2017	207.800.000	1.448.364	6,97	2,75	174,71%
2018	209.500.000	2.068.184	9,87	3,89	289,09%
2019	210.147.000	2.533.571	12,06	4,75	375,18%

A Figura 22 mostra os dados de teste de triagem decompostos em um gráfico de tendência e sazonalidade. A tendência de crescimento se acentuou desde o início de 2018.

Além disso, a sazonalidade mostra queda no número de exames em dezembro e pico em outubro.

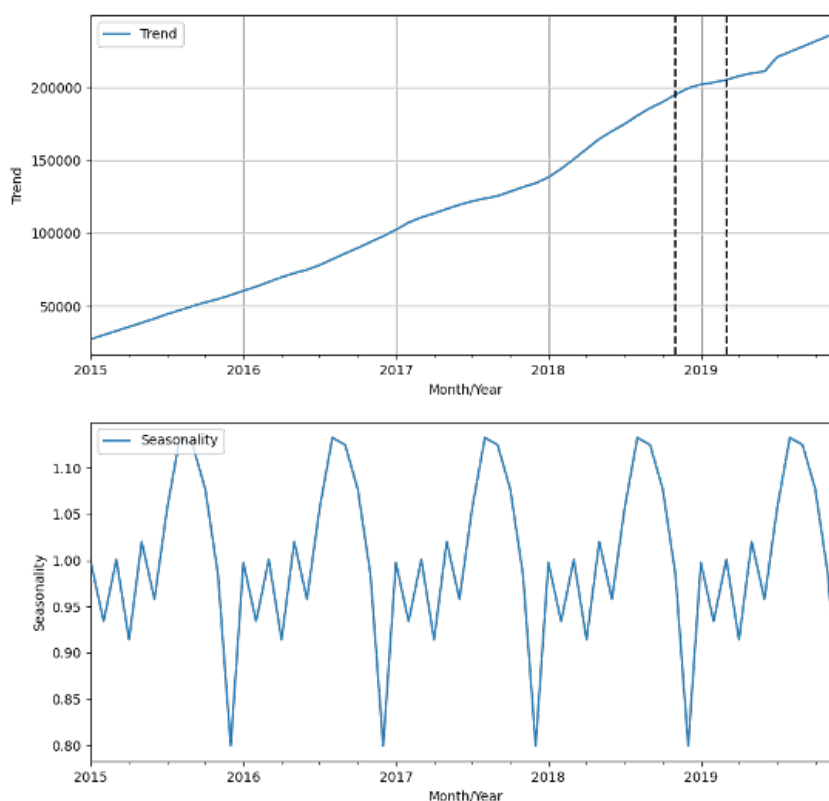


Figura 22 – Quantidade de testes de triagem decompostos em tendência e sazonalidade.

Outro fator importante que destaca o impacto do Projeto “Sífilis Não!” na influência do curso da epidemia no Brasil, pode ser notado ao se observar a distribuição da penicilina. A penicilina G benzatina (BPG) continua sendo o tratamento de escolha para todos os estágios da sífilis (FORRESTEL; KOVARIK; KATZ, 2020) e é o único antibiótico conhecido por atravessar a placenta e prevenir transmissão infantil de sífilis (ORGANIZATION, 2017).

A tabela 10 mostra a taxa de detecção de sífilis adquirida (por 100.000 habitantes), a taxa de detecção de sífilis em gestantes e a taxa de incidência de sífilis congênita (por 1.000 nascidos vivos) entre 2016 e 2019, além da distribuição de doses de BPG no mesmo período. Nesse caso, consideramos 2016 como linha de base, pois os dados de distribuição da penicilina estavam disponíveis apenas a partir de 2016 e agrupados por ano.

Assim, podemos observar que em 2016, o Brasil recebeu níveis de oferta de penicilina superiores a 2018 e próximo a 2019. Em 2017, houve escassez de penicilina benzatina no Brasil e no mundo (ÉPOCA, 2017; S.PAULO, 2017; ORGANIZATION, 2017) e isso pode explicar a menor distribuição neste ano.

Embora tenhamos observado um aumento tanto nas taxas de sífilis congênita no Brasil quanto de sífilis em gestantes entre 2016 e 2018, houve aumento de nascidos vivos sem sífilis no mesmo período. Destacamos os anos de 2018 e 2019 onde a relação entre

Tabela 10 – Taxas de incidência de sífilis em gestantes, sífilis congênita e distribuição de penicilina no Brasil, entre 2016 e 2019. Fonte de dados: *SINAN/MoH, **SISMAT/MoH

Ano	Nascidos Vivos	Taxas de incidência de sífilis em gestantes*	Taxa de incidência de sífilis congênita*	Razão (relação entre sífilis em gestantes e sífilis congênita)	Taxa de detecção de sífilis adquirida por 100.000 habitantes	Distribuição da Penicilina Benzatina**
2016	2.979.259	12,8	7,1	55,54%	44,5	1.557.000
2017	2.857.800	17,4	8,7	50,07%	59,0	648.800
2018	2.923.535	21,4	9,0	41,88%	75,4	1.279.800
2019	2.944.932	20,8	8,2	39,48%	72,8	1.965.550

sífilis em gestantes e sífilis congênita caiu para 41,9% em 2018 e para 39,4% em 2019, mesmo considerando o aumento das taxas de testagem. Isso significa que, em 2019, a probabilidade de uma criança nascer com sífilis era de 39,48%, considerando que sua mãe foi diagnosticada com sífilis.

Da mesma forma, a taxa de detecção de sífilis adquirida por 100.000 habitantes manteve um aumento entre 2016 (n=44,5), 2017 (n=59,0) e 2018 (75,4), mas caiu 2,6 casos por 100.000 habitantes em 2019 (n=72,8). Apesar da pequena diferença no número de casos (menos 5.136 casos em números absolutos), é importante notar o início do declínio dos casos, quando comparado com anos anteriores em que as taxas continuaram a subir.

A Figura 23 mostra a curva inclinada relativa à quantidade de casos notificados de sífilis adquirida, sífilis em gestante e sífilis congênita por mês e ano, decompostos em tendência e sazonalidade. Analisando o gráfico de tendências, observa-se um discreto declínio na quantidade de casos notificados de sífilis adquirida (a) e sífilis em gestantes (c), com queda acentuada nos casos de sífilis congênita (e) a partir de meados de 2018. Quanto à sazonalidade, podemos observar picos de casos notificados na sífilis adquirida e na sífilis em gestantes em agosto (segundo pico em março), e queda em dezembro. Os casos notificados de sífilis congênita apresentam picos em março e maio e queda em dezembro.

Esses dados refletem o desempenho das ações do sistema público de saúde quanto à detecção, tratamento e cura da sífilis adquirida, sífilis em gestante e sífilis congênita.

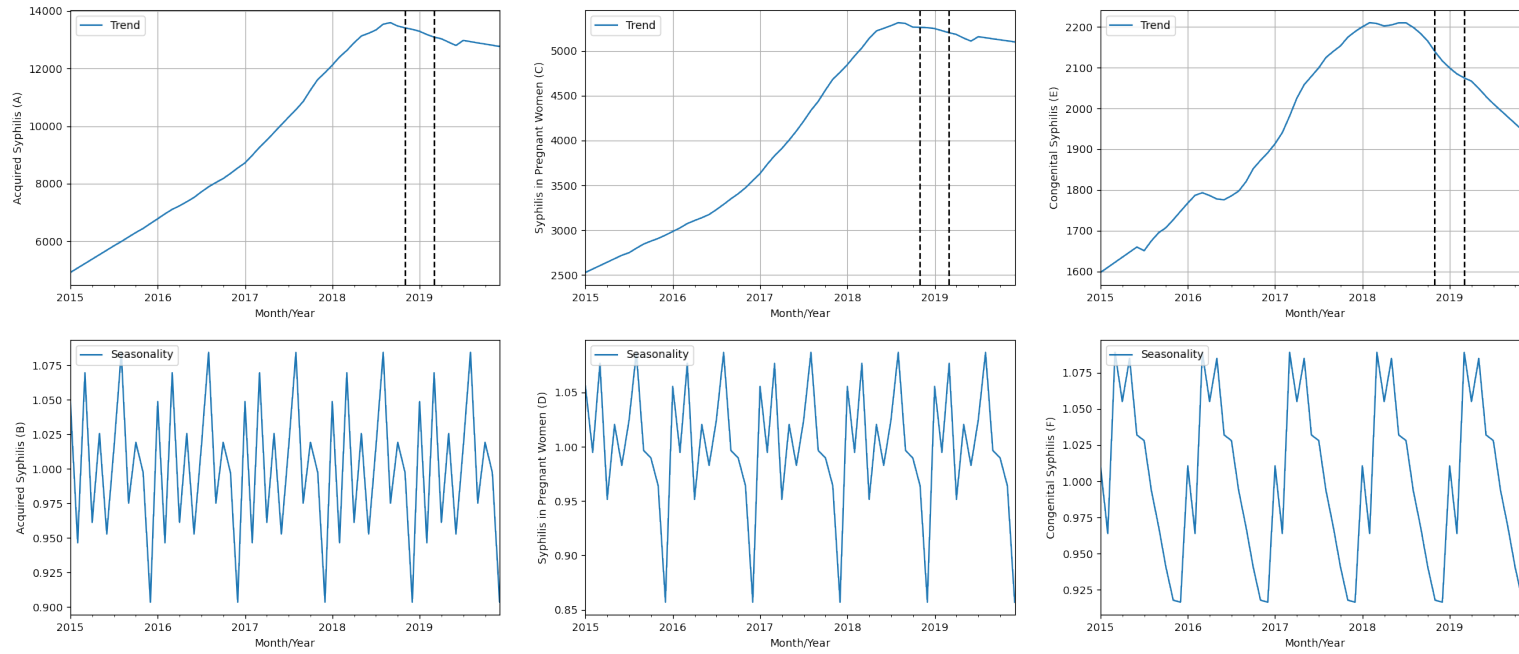


Figura 23 – Número de casos notificados de sífilis adquirida, sífilis em gestante e sífilis congênita por mês/ano decompostos em tendência e sazonalidade.

6.3 Discussão

Para analisar o alcance de uma campanha de saúde pública, é necessário observar seus dados de forma integrada por meio de diferentes dimensões. Por meio da conscientização da população sobre o problema de saúde, é possível mensurar o interesse pelo assunto através de diversas variáveis relacionadas, o que implicará em mudança de comportamento em relação à testagem, tratamento e hábitos saudáveis. Além disso, o conhecimento adquirido pelos profissionais de saúde pode auxiliar no diagnóstico, que pode ter sido anteriormente negligenciado por falta de conhecimento e divulgação do problema.

Em relação à campanha “Sífilis Não!” observamos que, a partir de junho de 2018, houve uma inclinação crescente correspondente ao ‘interesse ao longo do tempo’ em relação à atividade de busca na Internet, juntamente com um aumento de notícias *online* espontâneas relacionadas à sífilis. 12 MOOCs relacionados à doença foram disponibilizados pelo AVASUS para promover a aprendizagem ao longo da vida aos profissionais de saúde, professores e alunos, bem como ao público em geral, atingindo mais de 22.000 alunos matriculados até o momento (Set/2020).

Testes sorológicos demonstram a conscientização da população sobre a importância da testagem e do tratamento com foco na cura, que era o objetivo da campanha. Em 2019, houve um aumento de 375,18% no número de exames sorológicos para sífilis quando comparado a 2015, mesmo considerando o crescimento populacional.

Embora o Brasil tenha distribuído um número elevado de doses de penicilina em 2019 (n=1.965.550), é difícil mensurar o alcance com esses dados, pois houve escassez de penicilina em anos anteriores no Brasil e no mundo. Um dado mais preciso para a análise do tratamento da sífilis é o uso do medicamento nas unidades básicas de saúde ao invés da distribuição do medicamento, mas esse dado não está disponível.

Por fim, os casos de notificação de sífilis apresentam uma curva decrescente desde meados de 2018. Os casos de sífilis adquirida e de sífilis em gestantes ainda apresentaram um leve decréscimo na comparação entre 2018 e 2019. No entanto, é importante notar que esta é a primeira vez que houve redução nos casos notificados da doença nos últimos anos, mesmo considerando o aumento significativo de testagem no mesmo período. Além disso, os casos de sífilis congênita apresentaram queda de 31,79% no último ano (2019).

É importante destacar que nos anos de 2015 a 2017, o Brasil teve programas de fortalecimento da atenção primária à saúde. No entanto, a sífilis era uma infecção sexualmente transmissível negligenciada no país. Assim, não houve nesse período ações de comunicação para o enfrentamento da doença, principalmente no que se refere à população e aos profissionais de saúde. Nessa perspectiva, há uma influência positiva do Projeto “Sífilis Não!” como indutor da política de combate à sífilis no Brasil, pois conseguiu implantar uma agenda de testagem, tratamento e cura. Essa inferência é feita justamente porque nos anos anteriores a 2018, a campanha ainda não existia.

O *framework* de análise multidimensional aqui apresentado e implementado no

ecossistema Hermes é importante nesse contexto, pois consegue capturar os dados disponíveis, transformá-los em séries temporais e refleti-los de forma a possibilitar a análise das ações de saúde pública, indo além dos dados da vigilância epidemiológica. Vale ressaltar que os dados epidemiológicos observados isoladamente podem não explicar suficientemente os fenômenos apontados nesta análise.

O *framework* pode permitir uma avaliação mais abrangente do alcance das campanhas de saúde pública, permitindo assim que os formuladores de políticas re-analisem as estratégias de conscientização desenvolvidas para alertar as pessoas sobre os cuidados de saúde e mudanças comportamentais visando hábitos saudáveis. Além disso, o *framework* pode ser generalizado para uso com outras doenças no contexto de intervenções em saúde pública, uma vez que ele prevê a entrada de dados heterogêneos e que as análises descritivas e de correlação entre variáveis ao longo do tempo são úteis para outras intervenções em saúde.

7 Um Estudo de Séries Temporais para Analisar a Eficácia de uma Intervenção em Saúde Pública

Este capítulo apresenta um estudo de série temporal realizado com o objetivo de analisar dados epidemiológicos antes e após a implantação do Projeto “Sífilis Não!” (PSN). Foi importante analisar os dados nessa perspectiva (dimensão epidemiológica) para validar estatisticamente a resposta da saúde pública à sífilis, bem como informar aos *stakeholders* se a estratégia adotada no projeto foi adequada.

O objetivo deste estudo foi utilizar séries temporais interrompidas para estimar os efeitos da intervenção ao longo do tempo comparando as taxas de sífilis congênita por 1.000 nascidos vivos nos municípios prioritários (n=100) cobertos pelo PSN e nos municípios não prioritários (n=5.470), considerando as cinco macrorregiões do Brasil: Norte, Nordeste, Sul, Sudeste e Centro-Oeste.

Assim, este caso de estudo ajudou a compreender a dimensão epidemiológica e sua importância, a complexidade para analisar uma variável de interesse diretamente relacionada às políticas de saúde em um país continental que é social e culturalmente diverso e tem diferenças consideráveis nas desigualdades socioeconômicas, recursos e capacidades de serviços de saúde em cada região com vulnerabilidades socioeconômicas significativas que influenciam a resposta às doenças infecciosas.

Além disso, o Hermes foi validado nesta dimensão mostrando aos tomadores de decisão do PSN como as tendências da sífilis congênita mudaram após a intervenção em saúde. Embora este estudo não demonstre qual variável de interesse foi mais importante ou influenciou outras, ele mostra a influência positiva do Projeto na população e pode ser considerado um caso de sucesso a ser implementado novamente para combater outras doenças.

O texto contido neste capítulo foi organizado como um artigo intitulado “Use of Interrupted Time Series Analysis in Understanding the Course of the Congenital Syphilis Epidemic in Brazil” e foi publicado na *The Lancet Regional Health - Americas* (PINTO et al., 2022).

7.1 Métodos

O Hermes foi usado para coletar dados de sífilis congênita do Sistema de Informação de Agravos de Notificação (SINAN), fornecido pelo MS. Esses dados representam o total de casos notificados por município, por mês e ano. Os registros vitais de nascidos vivos por

município para cada ano foram recuperados do Departamento de Informática em Saúde do Ministério da Saúde do Brasil (DATASUS) e utilizados para cálculo de média mensal de nascidos vivos por município. O Hermes definiu as taxas de casos de sífilis congênita dividindo os casos mensais pelos nascidos vivos por município.

Além disso, taxas mensais por município foram separadas em municípios prioritários e não prioritários, agrupados por macrorregiões.

Um desenho quase experimental foi aplicado para analisar os dados de séries temporais usando um modelo de regressão linear segmentado no software R, adaptado para trocar dados com Hermes para analisar o impacto imediato e a mudança na tendência das taxas nacionais de sífilis congênita pré e pós intervenção do PSN. Além disso, o mês de maio de 2018 foi considerado o período inicial da intervenção para a implementação do Projeto “Sífilis Não!”.

Os dados foram coletados regularmente ao longo do tempo e em intervalos igualmente espaçados para realizar a análise de regressão segmentada (WAGNER et al., 2002; CARROLL, 2008). Para evitar qualquer viés relacionado à pandemia de Covid-19, foi definido dezembro de 2019 como data final da coleta de dados, totalizando 20 meses pré-intervenção (set/2016 a abr/2018) e 20 meses pós-intervenção (maio/ 2018 a dez/2019).

Foi assumida uma relação linear entre o tempo e o desfecho dentro de cada segmento, considerando que os modelos de regressão segmentada se ajustam a uma linha de regressão de mínimos quadrados para cada segmento da variável independente (tempo), $p < 0,05$ foi considerado estatisticamente significativo. Assim, o próximo passo na análise foi estimar a magnitude da intervenção e testar a significância estatística do impacto imediato e das taxas de tendência. O modelo de regressão usado para ajustar esses dados foi definido conforme a seguinte fórmula:

$$\begin{aligned} outcome_t = & \beta_0 + \beta_1 * t + \beta_2 * priority_t + \beta_3 * priority_t * i + \\ & \beta_4 * post-intervention_t + \beta_5 * (t - T_I) * post-intervention_t + \\ & \beta_6 * post-intervention_t * priority_t + \beta_7 * post-intervention_t * t * priority_t + \varepsilon_t \end{aligned}$$

Onde β_0 é o resultado da linha de base para o grupo não prioritário; β_1 é a tendência preexistente no resultado de interesse para o grupo não prioritário; β_2 é a diferença básica entre os grupos não prioritários e prioritários; β_3 é a diferença de tendência entre o grupo não prioritário e prioritário antes da intervenção; β_4 é o impacto imediato no grupo prioritário; β_5 é a mudança de tendência no grupo prioritário; β_6 é a diferença na mudança de impacto imediato entre o grupo não prioritário e prioritário após a intervenção, e β_7 é a diferença na mudança de tendência entre o grupo não prioritário e prioritário após a intervenção. A variável de prioridade é codificada com valor 1 para o grupo prioritário e 0 para o grupo não prioritário. A variável t é apenas um incremento que conta de 1 a 40 durante todo o período. A variável pós-intervenção foi definida como 0 antes da

intervenção e 1 depois. TI é a constante que representa o momento de intervenção, neste caso é igual a 20. A Figura 24 representa este modelo, incluindo o impacto imediato e a mudança de tendência. As linhas contínuas representam os valores observados e as linhas pontilhadas indicam o contrafactual para o período pós-intervenção.

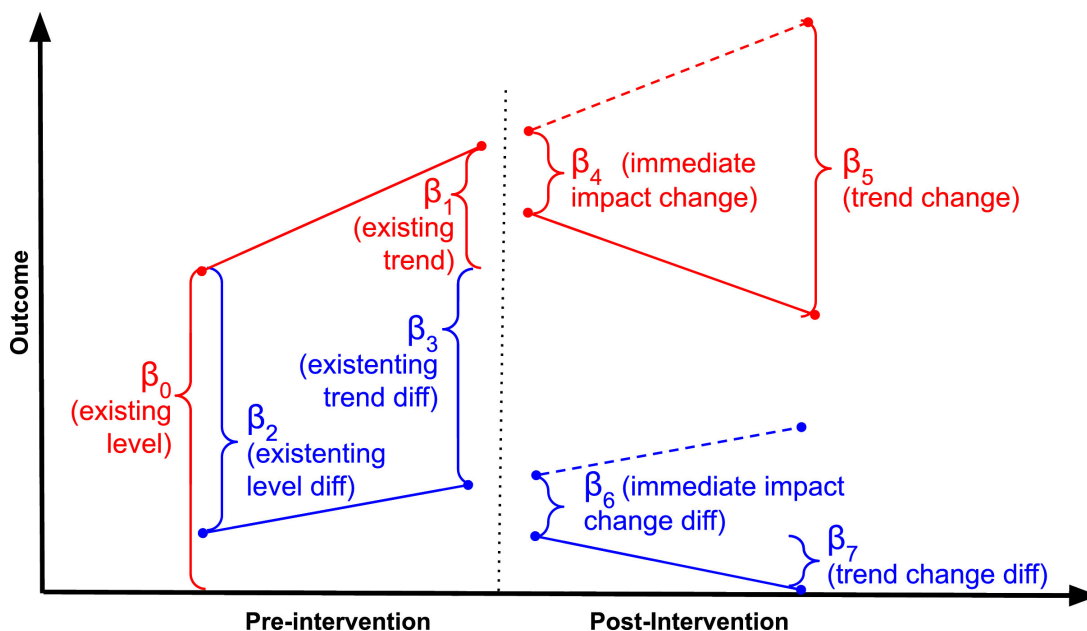


Figura 24 – Série temporal interrompida parametrizada como um modelo de regressão segmentado, incluindo o impacto imediato e a mudança de tendência.

A análise de regressão de Mínimos Quadrados Ordinários (OLS) assume que os termos de erro associados a cada observação não são correlacionados. A inspeção correta dos termos de autocorrelação pode evitar erros padrão subestimados e significância superestimada dos efeitos de uma intervenção (CARROLL, 2008). Neste estudo, um teste de Durbin-Watson foi usado para investigar a presença de autocorrelação, além de gráficos de resíduos, gráficos de Função de autocorrelação (ACF) e gráficos de ACF parcial. A autocorrelação não foi detectada nos testes estatísticos.

R-quadrado é uma medida estatística de quão próximos os dados estão da linha de regressão ajustada. Também é conhecido como coeficiente de determinação ou coeficiente de determinação múltipla para regressão múltipla. Ele foi utilizado no modelo desta pesquisa para determinar quão próximos os dados estão do modelo e como eles são apresentados nos resultados. R-quadrado é um valor entre 0 e 1, onde 0 indica que o modelo não explica nenhuma variabilidade dos dados de resposta em torno de sua média e 1 indica que o modelo explica toda a variabilidade dos dados de resposta em torno de sua média.

Outliers são valores extremos que parecem não se encaixar na série temporal. Se o outlier refletir um efeito histórico antecipatório ou de curto prazo, o ponto de dados pode ser modelado explicitamente. Alternativamente, se o outlier for uma consequência clara e inequívoca da intervenção, ele pode ser tratado como um ponto de dados regular para avaliar seu impacto. Uma análise preliminar foi realizada e um único outlier foi identificado

na região Centro-Oeste após a intervenção (julho de 2018). O método da Média Móvel (MA) foi usado para suavizar os dados com o objetivo de explicar as tendências temporais. A MA utilizada foi de ordem três, o que proporciona apenas uma suavização com os dois meses (antes e depois) da observação calculada. Após esse processo, nenhum outlier permaneceu.

Os modelos de regressão obtidos para cada grupo (municípios prioritários e não prioritários), utilizando dados dos períodos pré e pós-intervenção (maio de 2018), foram comparados com o contrafactual para o período pós-intervenção considerando a tendência observada antes da intervenção. O contrafactual permite uma comparação entre a mudança observada e o que teria acontecido se a intervenção não tivesse ocorrido. Os parâmetros do modelo de regressão foram obtidos por estimação pontual e intervalar com 95% de confiança.

7.2 Resultados

Os resultados da análise de séries temporais comparando as taxas de sífilis congênita por 1.000 nascidos vivos nos municípios prioritários e não prioritários são apresentados na Tabela 11 e Figuras 25 a 30. Os pontos vermelhos representam as taxas mensais de casos nos municípios prioritários e os pontos azuis nos não prioritários. As linhas contínuas indicam os modelos de regressão obtidos para cada grupo (prioritário e não prioritário), utilizando dados dos períodos pré e pós-intervenção (maio de 2018). Além disso, as linhas pontilhadas indicam o contrafactual para o período pós-intervenção usando a tendência dos dados observados antes da intervenção. A tabela 11 apresenta esta comparação considerando o último mês do período analisado (20º mês, dezembro de 2019).

Esses resultados descrevem detalhadamente os efeitos da intervenção para cada região, separadamente. Isso porque capta a influência regionalizada da governança do Sistema Único de Saúde — que é interfederativa, com atribuições claras nos níveis municipal, estadual e federal. Nesse sentido, como a Atenção Primária à Saúde (APS) é de responsabilidade dos municípios, é possível capturar evidências de que os efeitos do PSN podem variar entre as regiões.

Para a região Norte do Brasil, na Figura 25 e Tabela 11 (coluna 2), tanto os municípios prioritários quanto os não prioritários, apresentaram redução na taxa mensal de casos de sífilis congênita no período pós-intervenção. A redução média mensal foi de -1,77 casos por 1.000 nascidos vivos nos municípios prioritários e -0,09 por 1.000 nascidos vivos nos municípios não prioritários, conforme tabela 11. Além disso, os modelos de regressão linear estimados para cada situação mostram mudanças nas tendências médias mensais, com redução de -0,05 nos municípios prioritários e redução de -0,07 nos não prioritários.

O contrafactual apresentado na Tabela 12 mostra que para os municípios prioritários o modelo estimou 18,08 casos por 1.000 nascidos vivos ao final do 20º mês sem

intervenção. Em comparação, considerando a intervenção, foram observados 13,73 casos para o mesmo tempo, indicando um declínio de 24,09%. Nos municípios não prioritários, o modelo estimou 5,11 casos mensais por 1.000 nascidos vivos no 20º mês. O valor observado com a intervenção foi de 3,63 por 1.000 nascidos vivos, o que representa uma queda de 28,97%.

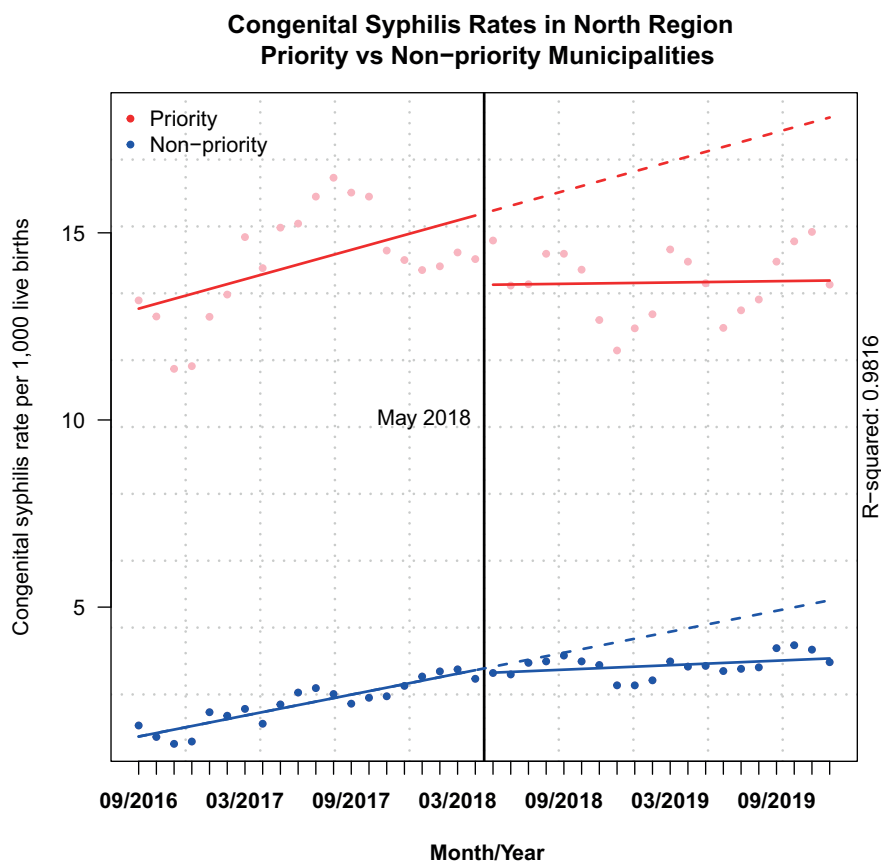


Figura 25 – Alterações nas taxas de sífilis congênita na região Norte, 20 meses antes e após a intervenção.

Tabela 11 – Resultados da análise de séries temporais comparando as taxas de sífilis congênita por 1.000 nascidos vivos entre municípios prioritários e não prioritários.

	Norte		Nordeste		Sul		Sudeste		Centro-oeste		Brasil	
	valor (CI 95%)	p-valor	valor (CI 95%)	p-valor	valor (CI 95%)	p-valor	valor (CI 95%)	p-valor	valor (CI 95%)	p-valor	valor (CI 95%)	p-valor
Pre-intervenção												
Intercept	1.45 (0.71 até 2.19)	<0.001	2.45 (1.51 até 3.39)	<0.001	4.65 (3.49 até 5.82)	<0.001	4.45 (3.61 até 5.30)	<0.001	2.54 (1.71 até 3.37)	<0.001	3.44 (2.71 até 4.17)	<0.001
Diferencial entre grupos	11.39 (10.35 até 12.44)	<0.001	15.58 (14.25 até 16.91)	<0.001	15.04 (13.39 até 16.68)	<0.001	8.38 (7.18 até 9.58)	<0.001	6.87 (5.70 até 8.04)	<0.001	10.96 (9.92 até 12.00)	<0.001
Tendência no grupo prioritário	0.04 (-0.05 até 0.13)	0.393	0.18 (0.07 até 0.29)	0.002	0.13 (-0.01 até 0.27)	0.062	0.09 (-0.01 até 0.19)	0.077	0.09 (-0.01 até 0.19)	0.063	0.11 (0.03 até 0.20)	0.011
Tendência no grupo não prioritário	0.09 (0.03 até 0.16)	0.004	0.07 (0.00 até 0.15)	0.062	0.03 (-0.07 até 0.13)	0.539	0.11 (0.04 até 0.18)	0.003	-0.02 (-0.08 até 0.05)	0.665	0.08 (0.01 até 0.14)	0.017
Pós-intervenção												
Impacto imediato no grupo prioritário	-1.77 (-3.20 até -0.34)	0.016	2.23 (0.42 até 4.05)	0.016	-1.76 (-4.00 até 0.49)	0.123	-1.11 (-2.75 até 0.53)	0.181	-0.47 (-2.07 até 1.13)	0.560	-0.36 (-1.77 até 1.06)	0.618
Impacto imediato no grupo não prioritário	-0.09 (-1.10 até 0.92)	0.865	0.34 (-0.94 até 1.62)	0.598	0.48 (-1.11 até 2.07)	0.548	0.13 (-1.03 até 1.29)	0.822	0.87 (-0.26 até 2.00)	0.131	0.28 (-0.72 até 1.28)	0.583
Mudança na tendência no grupo prioritário	-0.05 (-0.18 até 0.07)	0.407	-0.64 (-0.80 até -0.48)	<0.001	-0.28 (-0.48 até -0.09)	0.005	-0.02 (-0.16 até 0.13)	0.827	-0.07 (-0.21 até 0.06)	0.291	-0.21 (-0.33 até -0.09)	0.001
Mudança na tendência no grupo não prioritário	-0.07 (-0.16 até 0.01)	0.099	-0.11 (-0.22 até 0.00)	0.058	-0.04 (-0.18 até 0.10)	0.579	-0.16 (-0.26 até -0.06)	0.003	0.00 (-0.09 até 0.10)	0.949	-0.10 (-0.19 até -0.02)	0.021
R ²	0.9816		0.9877		0.9767		0.9691		0.9590		0.9848	

Tabela 12 – Resultados da análise de séries temporais comparando as taxas de sífilis congênita por 1.000 nascidos vivos entre municípios prioritários e não prioritários.

Região	Municípios prioritários			Municípios não prioritários		
	Contrafactual	Observado	Diferença (%)	Contrafactual	Observado	Diferença (%)
Norte	18.08	13.73	24.09	5.11	3.63	28.97
Nordeste	28.11	15.75	43.95	5.33	3.63	31.79
Sul	26.13	18.43	29.48	5.82	5.57	4.27
Sudeste	20.84	16.42	21.21	8.70	5.85	32.69
Centro-oeste	12.52	11.50	8.11	1.94	2.86	-47.86
Brasil	21.93	15.61	28.79	6.34	4.68	26.20

A Figura 26 e a Tabela 11 (coluna 3) apresentam os resultados para a região Nordeste, onde pode ser observado um aumento de 2,23 casos por 1.000 nascidos vivos no impacto imediato pós-intervenção em locais prioritários. Também houve redução de 0,34 nos casos para os municípios não prioritários. Uma mudança na tendência pode ser observada para ambos os grupos. Os municípios não prioritários reverteram a tendência ascendente com uma inclinação mensal de -0,11 casos por 1.000 nascidos vivos. Enquanto isso, os municípios prioritários apresentaram uma inclinação de -0,64 casos por mil nascidos vivos pós-intervenção.

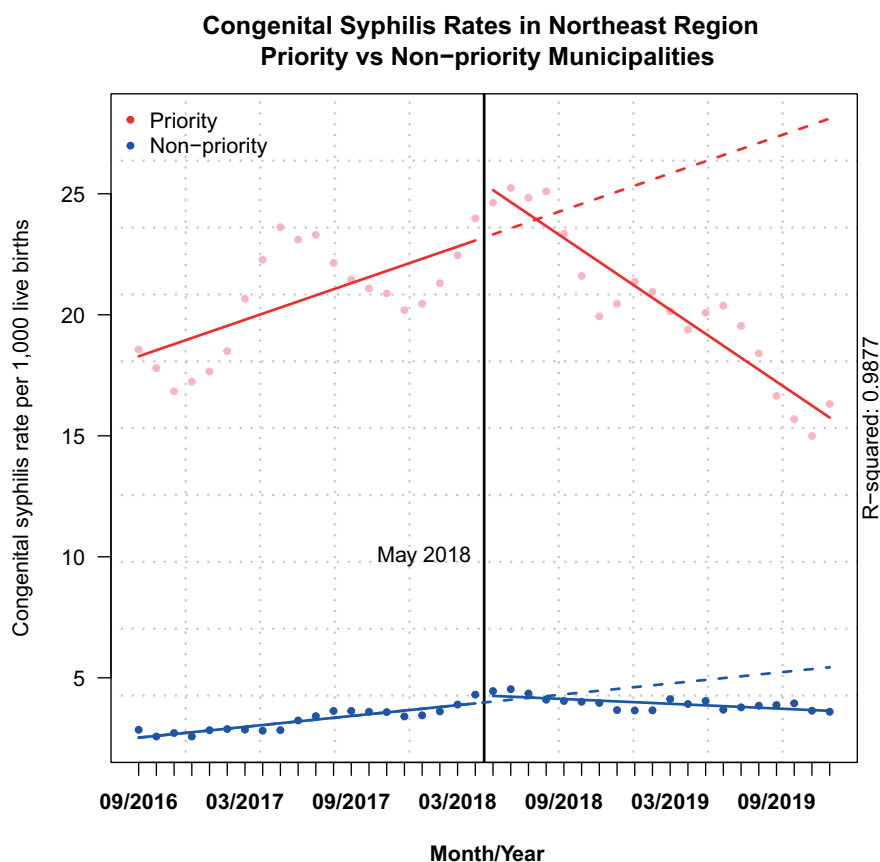


Figura 26 – Alterações nas taxas de sífilis congênita na região Nordeste, 20 meses antes e após a intervenção.

Em relação ao contrafactual (Tabela 12) para essa região, ao final do intervalo analisado nos municípios prioritários e sem intervenção, o modelo estimou 28,11 casos por 1.000 nascidos vivos, contra 15,75 casos observados após a intervenção (queda de 43,95% para este grupo). Nos municípios não prioritários, o modelo estimou 5,33 casos no último mês de coleta de dados, enquanto o valor real com a intervenção foi de 3,63 por 1.000 nascidos vivos, uma queda de 31,79%.

Conforme representado na Figura 27 e Tabela 11 (coluna 4), os municípios prioritários da região Sul apresentaram queda acentuada para -1,76 casos mensais por 1.000 nascidos vivos, e uma mudança de tendência de -0,28 casos mensais por 1.000 nascidos vivos. Por outro lado, nos municípios não prioritários houve um aumento de 0,48 casos mensais por 1.000 nascidos vivos e uma pequena mudança de tendência de -0,04 casos mensais por 1.000 nascidos vivos.

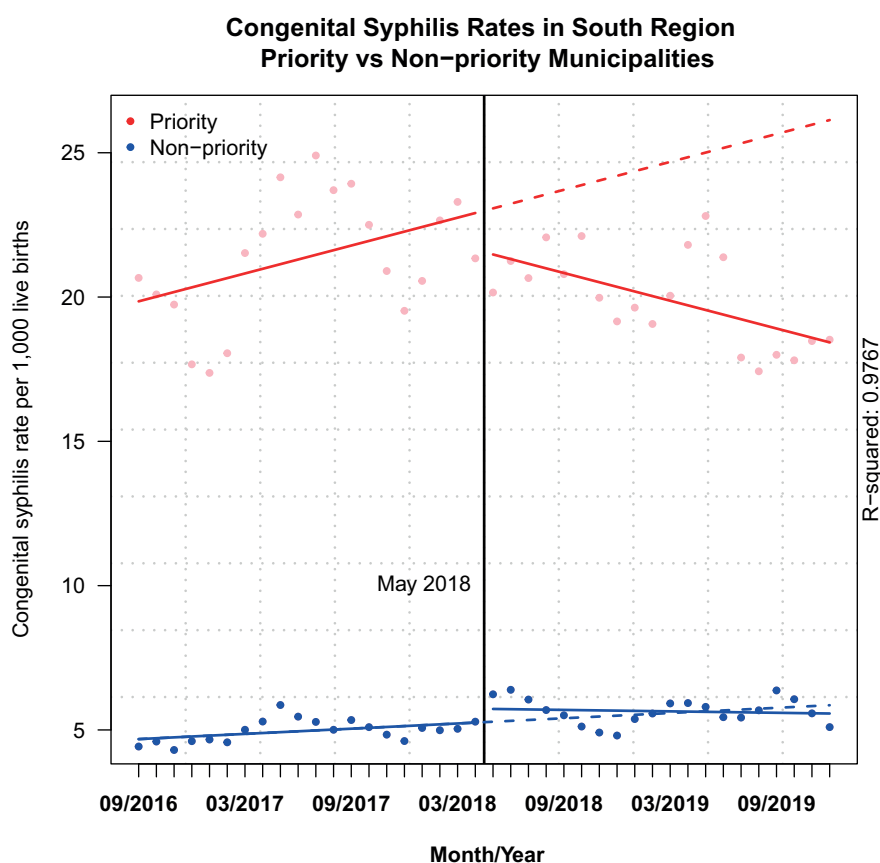


Figura 27 – Alterações nas taxas de sífilis congênita na região Sul, 20 meses antes e após a intervenção.

O modelo estimou 26,13 casos por 1.000 nascidos vivos se nenhuma intervenção tivesse ocorrido nos municípios prioritários da região Sul. O valor real com intervenção, foi de 18,42 casos por 1.000 nascidos vivos, representando uma queda de 29,48%. Uma ligeira queda de 4,27% foi observada para os municípios não prioritários com estimativa de 5,82 casos contrafactuais e 5,57 casos reais por 1.000 nascidos vivos.

Os municípios não prioritários da região Sudeste (Figura 28 e Tabela 11, coluna 5) passaram por um aumento de 0,13 casos mensais por 1.000 nascidos vivos e uma mudança de tendência com queda de -0,16 casos mensais. Já os municípios prioritários tiveram redução de -1,11 casos mensais e tendência de leve queda de -0,02 casos mensais.

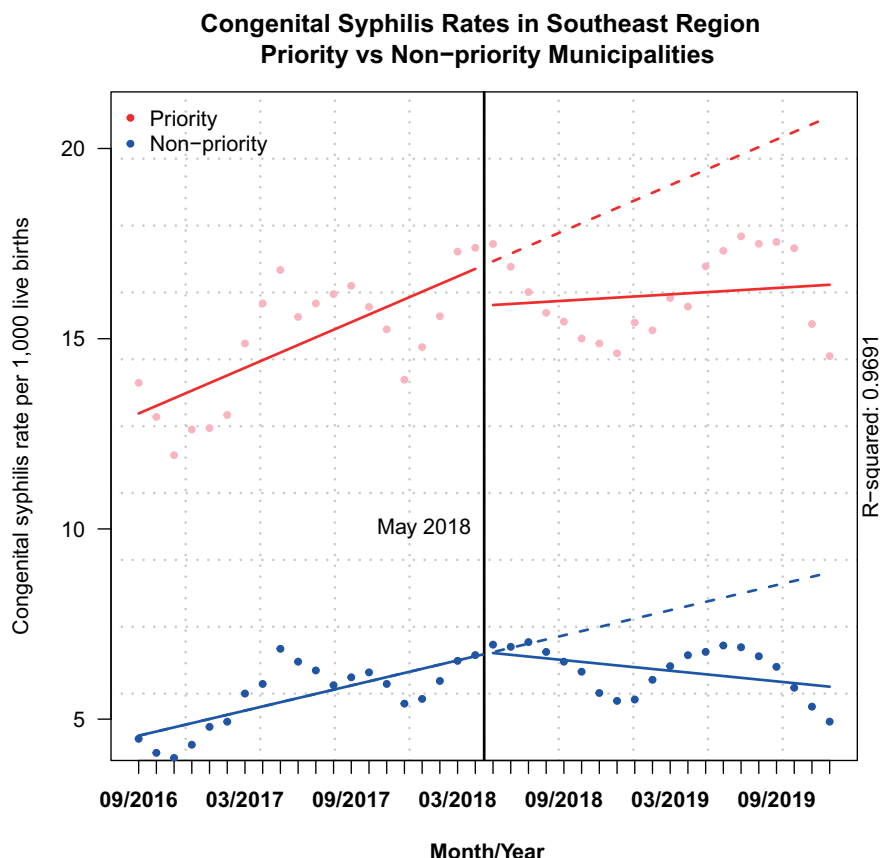


Figura 28 – Alterações nas taxas de sífilis congênita na região Sudeste, 20 meses antes e após a intervenção.

O contrafactual para esta região projetou uma estimativa de 20,84 casos por 1.000 nascidos vivos em municípios prioritários (Tabela 12) na ausência de intervenção. No entanto, observou-se uma redução de 21,21% para 16,42 casos por 1.000 nascidos vivos. Nos municípios não prioritários da região Sudeste, o contrafactual foi de 8,70 casos contra 5,85 casos reais por 1.000 nascidos vivos, com redução de 32,69% no período pós-intervenção.

Na região Centro-Oeste (Figura 29 e Tabela 11, coluna 6), houve redução de -0,47 casos mensais nos municípios prioritários após a intervenção, mas um aumento de 0,87 casos mensais nos municípios não prioritários. Houve uma reversão da tendência nos municípios prioritários de 0,09 casos mensais por 1.000 nascidos vivos pré-intervenção para -0,07 pós-intervenção. Por outro lado, em locais não prioritários, houve uma inversão da tendência. Dessa forma, a tendência de queda de -0,02 casos mensais antes da intervenção mudou para 0,00 no período pós-intervenção.

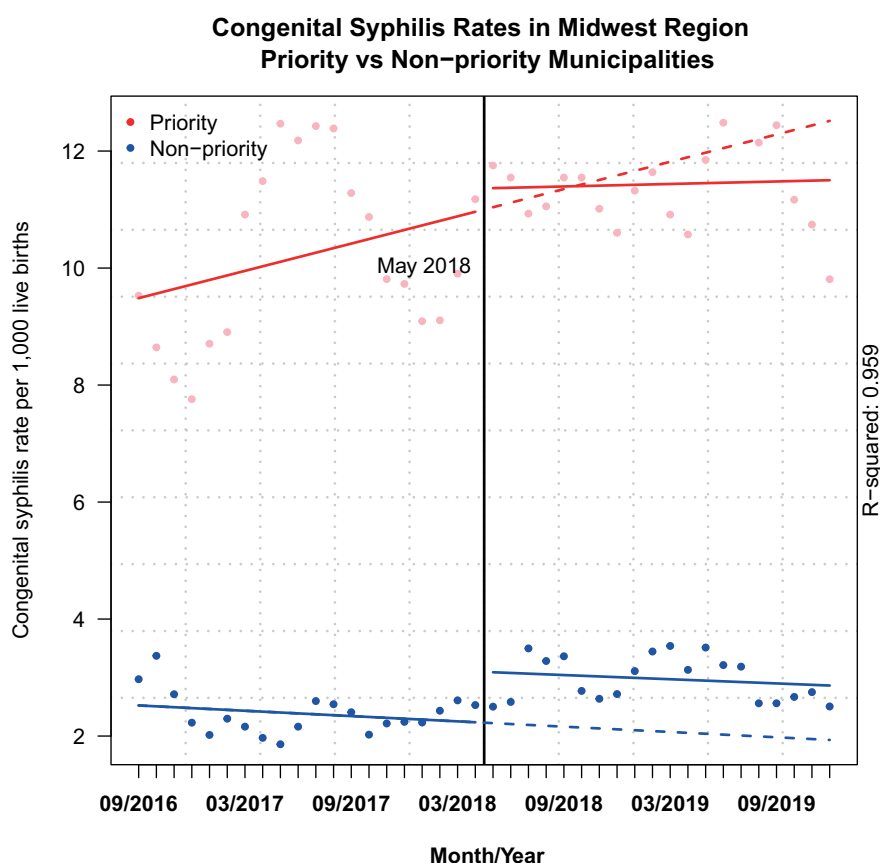


Figura 29 – Alterações nas taxas de sífilis congênita na região Centro-Oeste, 20 meses antes e após a intervenção.

O contrafactual para o período pós-intervenção nos municípios prioritários foi de 12,52 casos por 1.000 nascidos vivos no final do período (Tabela 12), contra 16,42 casos reais, o que representa uma queda de 21,21%. Em contrapartida, a comparação entre os casos contrafactuais (1,94) e reais (2,86) por 1.000 nascidos vivos para os municípios não prioritários revelou um aumento de 47,86%.

A análise conjunta dos dados nacionais revela um declínio no impacto imediato da sífilis congênita pós-intervenção, com redução média mensal de -0,36 casos por 1.000 nascidos vivos nos municípios prioritários. Por outro lado, houve um aumento de 0,28 casos por 1.000 nascidos vivos nos municípios não prioritários. No modelo de regressão linear, a verificação das tendências para ambos os grupos após a intervenção revela uma diminuição no crescimento do número médio de casos por 1.000 nascidos vivos. Assim, houve redução de -0,21 nos municípios prioritários e de -0,10 nos não prioritários, conforme ilustrado na Figura 30 e Tabela 11 (coluna 7).

Para os municípios prioritários, a análise combinada também indica uma redução de 28,79% no número de casos no último mês do período analisado (Tabela 12) quando comparado o contrafactual de 21,93 com o número real de 15,61 casos por 1.000 nascidos vivos. Nos municípios não prioritários, houve queda de 26,20%. Assim, passou de 6,34

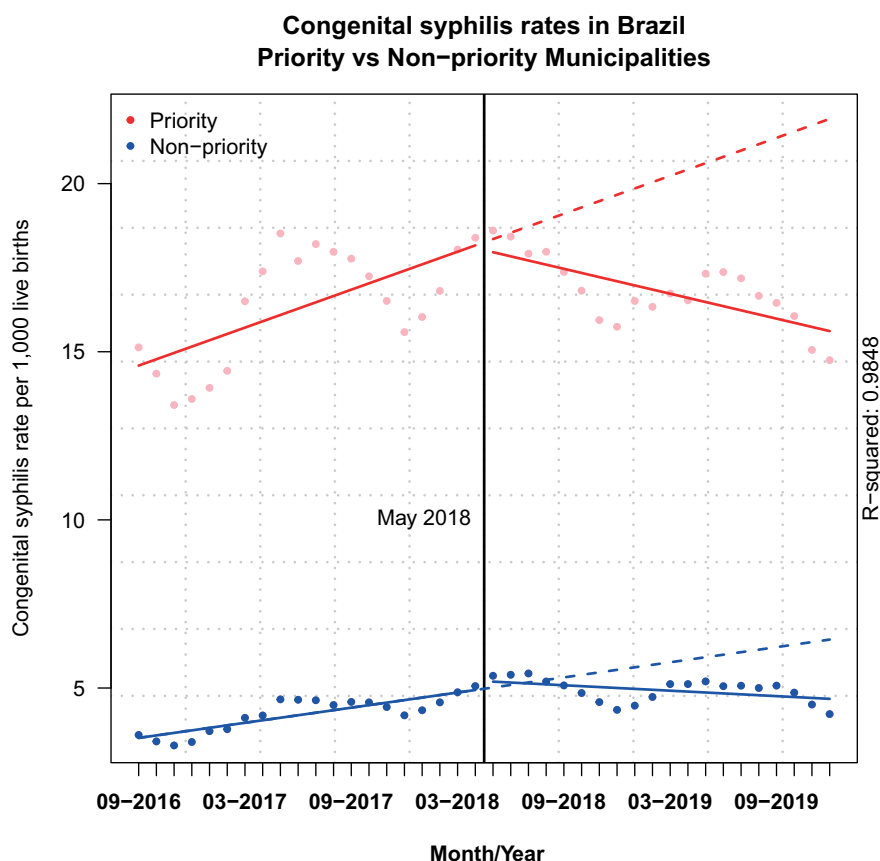


Figura 30 – Alterações nas taxas de sífilis congênita no Brasil, 20 meses antes e após a intervenção.

casos contrafactuais para 4,68 casos reais por 1.000 nascidos vivos.

Quanto à sífilis congênita nos municípios prioritários e não prioritários, as diferenças mais marcantes foram encontradas nas regiões Nordeste (15,58 por 1.000 nascidos vivos) e Sul (15,04 por 1.000 nascidos vivos), seguidas do Norte (11,39 casos por 1.000 nascidos vivos), Sudeste (8,38 casos por 1.000 nascidos vivos) e Centro-Oeste (6,87 casos por 1.000 nascidos vivos).

7.3 Discussão

O impacto do projeto na mudança do curso das tendências da sífilis congênita foi mais evidente para os municípios prioritários. Diferentemente das outras quatro regiões brasileiras, a mudança mais significativa na tendência de crescimento foi observada no Nordeste, que apresentou queda acentuada. Além disso, tendências decrescentes pós-intervenção foram observadas nas regiões Nordeste e Sul. Enquanto isso, houve estabilização da tendência no Norte. Por fim, não foram encontradas mudanças nas tendências para as regiões Sul e Centro-Oeste.

O controle da sífilis congênita depende, em parte, da capacidade dos sistemas de

atenção primária à saúde de acolher o maior número de gestantes infectadas e implementar efetivamente os protocolos de tratamento (RAC; REVELL; EPPES, 2017). Municípios com taxas crescentes de sífilis em gestantes podem indicar maior capacidade de controle da epidemia de sífilis congênita. Embora plausível, essa relação não pode ser diretamente correlacionada. Ou seja, muitos fatores ainda podem impactar no acolhimento e tratamento adequado de uma gestante com sífilis (por exemplo, a administração de penicilina na APS) (MOLINE; JR, 2016).

Nas últimas avaliações do programa de qualidade da APS no Brasil, identificou-se que a região Nordeste detinha o menor percentual na administração de penicilina no país. A administração de medicamentos não ultrapassou 40% dos casos diagnosticados de sífilis (FIGUEIREDO et al., 2020). Em seguida, o uso insuficiente da penicilina, possivelmente devido à sua escassez de oferta (CHAVES et al., 2019), foi normalizado em 2017. Além disso, há também a hipótese de que a resistência dos trabalhadores de saúde em administrar a penicilina em gestantes durante o pré-natal na APS contribuiu para uma administração reduzida.

Quanto ao eixo de atenção integral do SUS, uma das ações do projeto foi atuar vigorosamente na implantação do protocolo de tratamento da sífilis em gestantes. Isso incluiu ações de intervenção em municípios prioritários para promover ações educativas para o uso da penicilina na APS. Assim, a mudança mais acentuada nas tendências da sífilis congênita no Nordeste pode ser resultado da melhoria do tratamento das gestantes com sífilis por meio do uso da penicilina na APS.

O Nordeste do Brasil, onde o diagnóstico e acompanhamento da sífilis não seguiu o protocolo de diagnóstico, tratamento e notificação do Protocolo Clínico e Diretrizes Terapêuticas (PCDT) (TRANSMISSÍVEIS, 2015), foi uma das poucas regiões onde as taxas de sífilis congênita foram maiores que os de sífilis em gestantes (SANTOS et al., 2020). Esse fato pode ter contribuído para a intensificação das intervenções de Apoiadores de Pesquisa e Intervenção (RIS) que fortalecem as ações de formação continuada em saúde, com foco no PCDT. Essas hipóteses ainda precisam ser testadas, mas os quatro eixos de intervenção do projeto provavelmente tiveram efeitos diferentes nas diferentes redes de serviços de saúde.

Em geral, a epidemia de sífilis congênita tem sido identificada em várias partes do mundo, principalmente em países de baixa e média renda (KORENROMP et al., 2019; NEWMAN et al., 2013). No Brasil, um estudo de Santos e colaboradores (SANTOS et al., 2020), com o objetivo de avaliar as tendências da sífilis congênita no Brasil nos principais municípios de 2007 a 2017, encontrou tendência de crescimento significativo, com Variação Percentual Média Anual (AAP) de 15,75%. Nossos resultados concordam com esses achados e apontam para um aumento nos níveis de sífilis congênita até 2018, quando o Projeto “Sífilis Não!” foi efetivamente implementado.

Um dos objetivos do Projeto foi reduzir a sífilis congênita por meio do plano

estratégico de capacitação da gestão local e por meio das atividades realizadas pelos RIS nas áreas técnicas, cujas ações incluem os eixos de i) gestão e governança, ii) vigilância, iii) atenção integral, e iv) fortalecimento da educação e conscientização em municípios prioritários (FIGUEIREDO et al., 2020). As ações específicas realizadas pelos RIS foram i) a avaliação dos planos e programas de saúde das secretarias municipais de saúde; ii) o fortalecimento dos comitês de investigação da transmissão vertical; iii) o reforço dos sistemas de informação estratégicos para a vigilância sanitária e triagem de notificações; e iv) a operacionalização da linha de cuidado à criança exposta à sífilis (TRANSMISSÍVEIS, 2017).

Os resultados deste estudo sugerem que as ações do Projeto influenciaram na reversão e redução significativa das tendências da sífilis congênita no Brasil a partir do ano de 2018. Essa redução foi verificada em nível hospitalar por Andrade e colaboradores (ANDRADE; VALENTIM; OLIVEIRA, 2020). Os pesquisadores realizaram um estudo que avaliou o impacto do Projeto “Sífilis Não!” na redução de internações por sífilis congênita em municípios prioritários. Seus achados revelaram que o Projeto reduziu significativamente as internações por sífilis congênita, corroborando as mudanças nas tendências de novos casos observadas em nosso estudo.

Os resultados contrafactuais revelaram que o projeto foi capaz de reverter as projeções esperadas de casos de sífilis congênita. Embora a análise da série temporal interrompida sugira maiores reduções nas áreas prioritárias do que nas áreas não prioritárias, é importante notar que as reduções percentuais foram semelhantes para esses dois grupos de áreas. Além disso, os resultados deste estudo reforçam que a redução da sífilis congênita não ocorreu de forma equitativa em todas as cinco regiões do Brasil, embora o projeto tenha sido implementado em nível nacional. O Brasil é um país social e culturalmente diversificado com dimensões continentais. Embora tenha um sistema de saúde como o SUS, com cobertura universal de saúde, há diferenças consideráveis nas desigualdades socioeconômicas (MASSUDA et al., 2018; CASTRO et al., 2019), e recursos e capacidades do serviço de saúde em cada região (OLIVEIRA et al., 2017; CECILIO; REIS, 2018) com vulnerabilidades socioeconômicas significativas que influenciam a resposta a doenças infecciosas (ROCHA et al., 2021).

As mudanças no nível de casos e tendências nos municípios não prioritários que não receberam os RIS ou ações específicas de conscientização também podem ser atribuídas aos efeitos de transbordamento do Projeto “Sífilis Não!”, que foi um projeto de destaque nacional e bastante conhecido no Sistema Único de Saúde. Os resultados também revelam mudanças nas tendências de redução ou estabilização em quase todas as regiões (Norte, Nordeste, Sul e Sudeste). Além disso, ao investigar o impacto das ações educacionais do PSN em um estudo anterior (PINTO et al., 2021), com destaque para o lema “testar, tratar e curar”, concluímos que houve um aumento consistente nas pesquisas sobre o tema da sífilis nos mecanismos de busca da Internet, bem como o aumento da utilização do

sistema de aprendizagem *online* (AVASUS) pelos profissionais de saúde. Diante disso, pode-se inferir que as ações globais do projeto no eixo educacional foram capazes de modificar, em menor grau, as tendências da sífilis congênita em locais não prioritários.

Não foi observado impacto das ações universais do projeto em municípios não prioritários da região Centro-Oeste. Apenas ações universais do projeto foram oferecidas a esses grupos. Foram elas: campanhas de conscientização em massa para incentivo à prevenção, diagnóstico e tratamento, além da distribuição universal de recursos de testagem e tratamento, instrumentação de salas de situação e desenvolvimento de estudos e pesquisas aplicadas (TRANSMISSÍVEIS, 2017). Provavelmente, as ações relacionadas à educação não atingiram seu objetivo primordial nesta região. Consequentemente, essas atividades não têm obtido o engajamento ideal dos gestores locais e das equipes de saúde também.

Nosso estudo tem várias limitações. Uma delas é a ausência de mensuração sobre os efeitos de cada eixo de intervenção do Projeto “Sífilis Não!” sobre as mudanças de tendência observadas. Embora o reconhecimento do efeito global do projeto tenha sido adequadamente verificado, os diferentes níveis de capacidade instalada de vigilância epidemiológica em saúde nos municípios avaliados podem ter dificultado a leitura e comparação dos quatro eixos do projeto. Estudos futuros que controlem o nível de desenvolvimento do sistema de vigilância em cada localidade e outros indicadores de qualidade da atenção à saúde podem indicar quais ações do projeto tiveram o impacto mais significativo. Além disso, é natural que a própria limitação do desenho do estudo não capte outros efeitos, como a influência das diferenças no desenvolvimento sociorregional, o que poderia impactar as tendências pós-intervenção.

Vale ressaltar que este é um estudo quase experimental. Portanto, não é um experimento no qual teríamos controle sobre as variáveis. A redução na tendência observada em alguns casos foi provavelmente influenciada por fatores além da influência do projeto. Além disso, como a sífilis é uma epidemia no Brasil, as taxas foram altas em algumas regiões. Esperava-se que esses valores fossem reduzidos devido ao fenômeno de “regressão à média”. No entanto, esse efeito não pôde ser medido com precisão porque o desenho do estudo não incluiu um grupo controle, típico de estudos experimentais.

Outro ponto é que em 2017 o MS alterou os critérios de definição de casos para notificação de sífilis congênita, o que geralmente pode impactar o sistema de informação e gerar subnotificação. No entanto, assumimos que a nova definição de caso não poderia alterar significativamente os modelos matemáticos apresentados neste estudo. Outra limitação deste estudo é que não foi possível obter o número exato de nascidos vivos mensais. Esses dados foram calculados com base no número de nascidos vivos anuais por município, apresentados pelo MS. No entanto, a boa quantidade de dados e o alto índice de representação dos dados em nível nacional ajudaram a reduzir o impacto dessa limitação.

Apesar dessas limitações, os achados de nosso estudo - que utilizou testes esta-

tísticos para verificar os efeitos de uma intervenção em nível nacional - indicam que a evolução da sífilis congênita no Brasil, comparando os períodos pré e pós-intervenção, revela mudanças de tendência estatisticamente significativas nas taxas mensais da doença no Brasil. Também revela variações de taxas por região. Além disso, com apenas 20 meses de implantação do projeto, observa-se uma queda nos casos notificados por 1.000 nascidos vivos na maioria das regiões do país.

Essas descobertas fornecem uma oportunidade e direção para estudos futuros sobre as ações específicas implementadas pelos RIS em cada estado e região pelo Projeto “Sífilis Não!”, e examinar como os resultados mudarão ao longo do tempo após os efeitos adversos e interrupções causados pelo Covid-19 no sistema de saúde.

8 Notícias Online e suas Influências na Resposta à Saúde Pública

Este capítulo descreve um estudo de caso que explora notícias *online* para entender essa variável de interesse na dimensão da Comunicação. O estudo de caso foi realizado com o objetivo de buscar evidências para avaliar correlações entre notícias *online* e testes de triagem, destacando a importância desse tipo de mídia para conscientizar a população a mudar hábitos, em busca da prevenção, tratamento e cura de doenças infecciosas.

Este estudo analisa as notícias *online* sobre sífilis no Brasil, mapeando os principais elementos das notícias *online* publicadas entre 2015 e 2019 e medindo sua qualidade. O estudo abrange dois períodos: (1) 2015 a 2017, referente às ações de comunicação anteriores ao Projeto “Sífilis Não!”, e (2) 2018 a 2019, referente à cobertura jornalística das ações de comunicação realizadas durante o Projeto.

Neste estudo, nossa hipótese é que a quantidade e a qualidade das notícias sobre sífilis divulgadas na forma de mídia espontânea são indicadores significativos para mensurar o impacto das ações de comunicação desenvolvidas pelo Projeto ao longo de 2018 e 2019.

A análise dos resultados reforça a importância de avaliar variáveis de interesse possivelmente negligenciadas em outros estudos para avaliação de intervenções em saúde pública. As notícias *online* podem atingir grande parte da população, além de promover uma agenda inter-mídia (MCCOMBS, 2015), mantendo o tema na pauta da mídia por um longo período.

É importante ressaltar que é necessário procurar outros atributos ocultos no conteúdo textual das notícias online. Quando exploramos atributos básicos relacionados a séries temporais em notícias *online* (quantidade de notícias por mês - considerando apenas a data de publicação), não encontramos correlação com outras variáveis de interesse. No entanto, analisando profundamente, extraímos categorias para definir a qualidade das notícias. Considerando esse novo atributo, foi possível encontrar correlações entre notícias e testes de triagem na população.

O texto contido neste capítulo foi organizado como um artigo intitulado “How Online News Affects Public Health Response to Syphilis in Brazil: Text Mining Mediated Analysis” e está em revisão no *Journal of Medical Internet Research*.

Nas próximas seções, explicaremos este estudo, seus métodos e resultados alcançados. As discussões deste estudo foram possíveis devido aos processos automatizados pelo ecossistema Hermes, que utilizou métodos de processamento de linguagem natural e aprendizado de máquina.

8.1 Métodos

Para realizar este estudo, primeiramente realizamos uma pesquisa de campo caracterizando e qualificando notícias sobre a sífilis por meio do Hermes. A ideia era identificar indicadores de qualidade da informação sobre sífilis na agenda midiática. Assim, extraímos e realizamos uma análise de conteúdo das páginas web selecionadas. O fluxo de trabalho realizado nesta seção é apresentado na Figura 31.

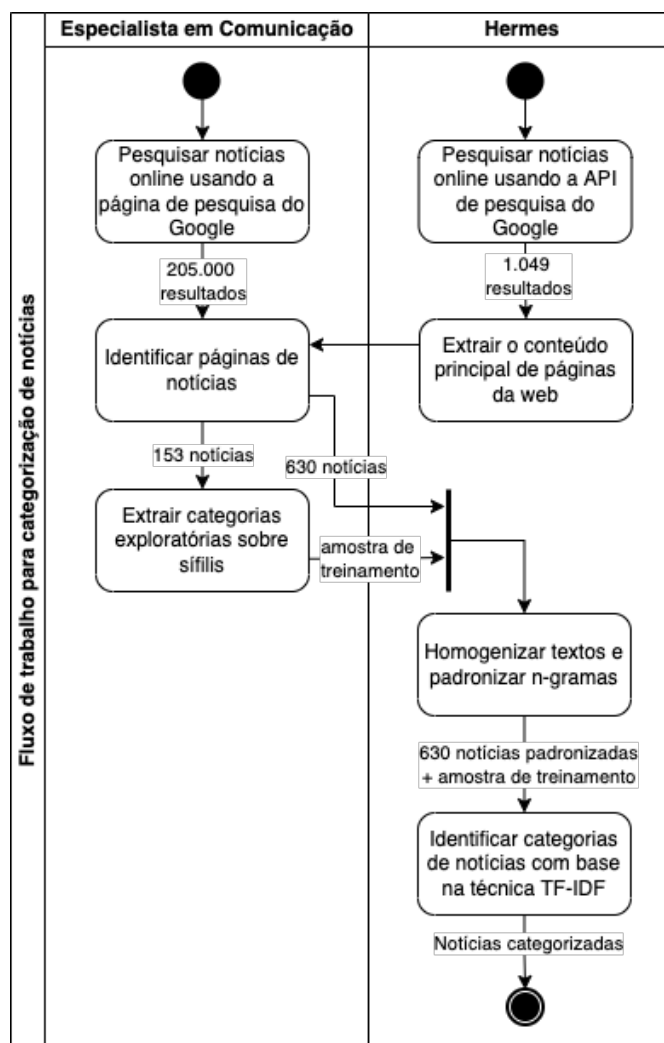


Figura 31 – Fluxo de trabalho realizado para categorização de notícias.

Utilizou-se a busca do Google para identificar as páginas de notícias da Internet, utilizando as palavras-chave: “sífilis”, “campanha sífilis”, “AIDS”, “campanha AIDS” e “doenças sexualmente transmissíveis”. Não houve filtragem por data, idioma ou geolocalização. A partir dos resultados desta pesquisa, especialistas em comunicação acessaram manualmente as páginas mais relevantes e selecionaram 153 textos, que foram caracterizadas como notícias (QING; YING; PENGZHOU, 2017), a partir de elementos que constituem uma notícia, a saber, a manchete, o lead (o parágrafo que resume o “quem, quando, onde, o que, por que e como”), e o corpo, que elabora mais sobre os elementos

mencionados no lead.

Para mapear características relevantes nas notícias identificadas, utilizamos análise temática e codificação axial para identificar temas emergentes e subtemas (OROZCO; GONZÁLEZ, 2012). Por meio da indexicalidade, ou seja, do processo de análise dos fragmentos textuais que constituíram as notícias analisadas, categorizamos, por indução, as notícias sobre sífilis nas seguintes categorias exploratórias (ou indicadores de qualidade): definições da sífilis; dados epidemiológicos sobre sífilis; como prevenir a sífilis; como ou onde fazer o teste rápido/diagnóstico da sífilis; consequências da sífilis em populações-chave (como grávidas e lactentes) e risco de letalidade devido à sífilis terciária; tratamentos eficazes para curar a sífilis; e campanhas de comunicação pública.

Essa análise resultou em sete categorias genéricas de notícias: Definição de Doenças; Indicadores Epidemiológicos; Prevenção; Teste Rápido/Diagnóstico; Consequências; Campanha e Tratamento. A Tabela 13 fornece uma visão geral das categorias exploratórias e sua distribuição quantitativa de fragmentos de texto no corpo das 153 notícias analisadas. Esses fragmentos de texto serão definidos como “amostra de treinamento” e, após homogeneização e padronização, servirão para verificar se as palavras mais importantes de cada categoria estão contidas nas notícias online.

Tabela 13 – Distribuição quantitativa das categorias exploratórias encontradas na amostra de treinamento a partir dos resultados da busca realizada pelos pesquisadores.

Categorias exploratórias	Termo genérico	n
Definições da Sífilis	Definição da Doença	90
Dados Epidemiológicos sobre Sífilis	Indicadores Epidemiológicos	96
Como Prevenir a Sífilis	Prevenção	73
Como ou Onde Fazer um Teste Rápido/Diagnóstico de Sífilis	Teste Rápido/Diagnóstico	76
Consequências da sífilis em populações-chave e o risco de letalidade da sífilis terciária	Consequências	75
Campanhas de Comunicação	Campanhas	65
Tratamento Eficaz para Sífilis	Tratamento	68

Definidas as categorias exploratórias e a amostra de treinamento, procedeu-se à segunda etapa do fluxo de trabalho, que consistiu em identificar o volume e a qualidade das notícias sobre sífilis através do Hermes.

Assim, de janeiro de 2015 a dezembro de 2019, o Hermes coletou páginas online, indexadas pelo Google, com o termo “sífilis” em português do Brasil. O Hermes executou essa tarefa usando a API de pesquisa do Google ¹ que permite recuperar, em formato JSON, um conjunto de sites indexados por palavra-chave (GAVALI, 2015). Os campos retornados pelos resultados da pesquisa automatizada incluíam: a data em que a notícia

¹ <<https://developers.google.com/custom-search/v1/overview>>

foi indexada, URL, título e um trecho da notícia. 1.049 páginas *online* foram encontradas neste período usando esses parâmetros.

O submódulo de notícias *online* do Hermes extraiu os seguintes campos de cada URL coletada: título, data de publicação, palavras-chave, conteúdo textual e resumo. Nesse ponto, foi preciso identificar se essas 1.049 páginas eram, de fato, notícias. Assim, foi realizada uma identificação manual e 630 páginas foram definidas como notícias (Tabela 14).

Tabela 14 – Páginas da Web categorizadas após a extração do conteúdo do texto principal.

Tipo	Justificativa	Quant.
Página não encontrada/Conteúdo não disponível	A página não existe mais ou o conteúdo é exclusivo para membros.	42
Resultados de pesquisa	Resultados de conteúdo de página de uma pesquisa no site.	69
Conteúdo não textual	Página não inclui conteúdo de texto (por exemplo, podcasts, imagens, vídeos, arquivos PDF ou PPT).	123
Pesquisa científica	O conteúdo principal é parte de um resumo de pesquisa científica, como um artigo, pôster, tese ou dissertação.	185
Notícias	O conteúdo principal tem características de um texto jornalístico.	630

Para fundamentar e construir a extração automatizada de características encontradas nas notícias *online* através do Hermes, utilizamos a técnica de mineração de texto, também conhecida como Text Data Mining, que calcula o peso de uma palavra-chave (escore) em um modelo vetorial (matriz) (FELDMAN; SANGER et al., 2007).

Verificamos unigramas, bigramas e trigramas por meio de uma análise comparativa para verificar qual combinação de palavras-chave poderia ser a mais adequada para extrair as características das em notícias. Neste trabalho, a melhor combinação foi a de bigramas.

Os procedimentos para homogeneizar e padronizar os bigramas são os mesmos descritos no Capítulo 3, seção 3.2. Assim, as frases são convertidas em bigramas que compõem um dicionário de palavras representando cada grupo (a amostra de treinamento e as notícias *online* sobre sífilis). A Tabela 15 fornece exemplos dos principais bigramas encontrados para cada categoria exploratória na amostra de treinamento.

Esse processo gerou uma matriz, com cada linha correspondendo a uma notícia e cada coluna aos bigramas. As células referem-se ao escore de relevância do bigrama nas notícias, e foram estimadas através de Term Frequency – Inverse Document Frequency (TF-IDF) (SALTON; BUCKLEY, 1988). Essa técnica determina estatisticamente a importância de uma palavra em um corpus de documento em relação a outros textos dentro do mesmo banco de dados. O peso de uma palavra para tal documento depende do número de vezes que aparece nele, mas é compensado pelas frequências das palavras nos outros documentos dentro do mesmo banco de dados (HONG et al., 2013). Portanto,

Tabela 15 – Principais bigramas encontrados para cada categoria exploratória na amostra de treinamento.

Categoria	Bigramas
Definição da Doença	Bactéria cura; sexualmente transmitida; transmitida causada; curavel tratamento; tratamento simples; infeccao apresentar; doenca sexualmente; chamada sifilis
Índices Epidemiológicos	ano idade; saude aumento; novas infeccoes; casos tipo; congenita periodo; tipo sifilis; sifilis pais; mil habitantes; ultimo boletim;
Prevenção	distribuir preservativos; preservativo melhor; atraves uso; prevencao usar; medicos reforcam; necessidade preservativos; reforcam necessidade; usar preservativo;
Teste Rápido/Diagnóstico	exame sorologico; rapido exame; sifilis resultado; detectada teste; facilmente detectada; rapida usada; usada diagnostico; resultado minutos;
Consequências	caso nao; formacao bebe; causar complicacoes; levar morte; feridas genitalia; genitalia manchas; causar aborto
Campanha	campanha sifilis; sifilis nao; campanha municipal; atividades integram; encerramento campanha; campanha destinada
Tratamento	meio penicilina; tratamento penicilina; base antibioticos; tratamento base; meio antibioticos; sifilis tratada;

palavras que são comuns em todos os documentos, como “este”, “que” e “se”, têm uma classificação baixa, embora possam aparecer muitas vezes, pois não significam muito para esse documento em particular. O TF-IDF usado para analisar esses dados foi definido da seguinte forma:

$$TF-IDF(t, d, D) = TF(t, d) * IDF(t, d, D)$$

Onde o Termo-Frequência (TF) mede quantas vezes (freq) a palavra t existe no documento d . Essa frequência é calculada da seguinte forma:

$$TF(t, d) = freq(t, d)$$

Por exemplo, neste conjunto de textos: “este é o primeiro documento.”, “este documento é o segundo documento.” e “Este é o primeiro documento?”, o TF da palavra documento é 1 para o primeiro documento, 2 para o segundo documento e 1 novamente para o terceiro documento.

O IDF é usado para medir o score t a partir da frequência em d , no conjunto de documentos. O IDF é definido pelo log entre o total de documentos (N) e a frequência de documentos d , onde ocorre o termo t (df_t). Essa frequência é calculada da seguinte forma:

$$IDF(t, d, D) = \log \left(\frac{N}{df_t} \right)$$

Usando o mesmo exemplo, o IDF do documento t é $\log(3/3)$, $\log(1)$, logo o IDF do documento t é 0. Portanto, a pontuação final de t é o peso resultante da frequência do t em d (TF) e a frequência de documento inversa (IDF).

Posteriormente, foram somados todos os escores de bigramas encontrados dentro das características das notícias, e então foi definido um escore final. Por fim, considerando os resultados, o ecossistema Hermes, em um processo automatizado, determinou o limite de aceitação para identificação de categorias para cada uma das sete categorias identificadas, com base na pontuação média de cada categoria.

O limite de aceitação é responsável por garantir que uma notícia tenha a pontuação mínima para que a presença das características sejam identificadas. Assim, verificamos se a pontuação aplicada é maior ou igual ao limite de aceitação. Se sim, a categoria identificada está presente na notícia. Esse procedimento foi utilizado para todas as 630 notícias extraídas e para todas as sete categorias que definimos, nas quais uma notícia pode incluir nenhuma, uma ou mais categorias exploratórias.

Os parâmetros de qualidade das notícias foram determinados com base na presença e combinação de categorias nelas, sendo: baixa qualidade (0–2 categorias), média qualidade (3–4), alta qualidade (5 ou mais).

Usando o Coeficiente de Correlação de Spearman (KUMAR; ABIRAMI, 2018), examinamos a associação entre o número de notícias no período de janeiro de 2015 a dezembro de 2019, agrupadas por seus respectivos níveis de qualidade (baixo, médio e alto), e o número de testes de triagem para diagnóstico de sífilis realizado na atenção primária à saúde no Brasil. Os dados de número de exames foram obtidos por meio do Sistema de Informação Ambulatorial (SIA) do Sistema Único de Saúde (SUS). Também repetimos a análise usando os coeficientes de Pearson e Kendall, mas o método do Coeficiente de Correlação de Spearman (também conhecido como RHO de Spearman) revelou melhores resultados. Talvez porque (i) os dados geralmente não são distribuídos entre as duas variáveis, (ii) há uma relação monotônica entre os dados, e (iii) ambas as variáveis são ordinais (HAUKE; KOSSOWSKI, 2011).

8.2 Resultados

O mapeamento das 630 notícias relacionadas à sífilis indexadas pelo Google, entre 2015 e 2019, revela um número crescente em 2018 e 2019.

A Tabela 16 fornece o número total de notícias mapeadas por ano, juntamente com as categorias exploratórias encontradas. O número de notícias aumentou gradativamente de 18 em 2015 para 26 em 2016 e 42 em 2017, depois subiu substancialmente para 107 em 2018. Em 2019 o número de notícias aumentou para 437 – número superior à soma do número de notícias nos quatro anos anteriores.

A Figura 32 mostra o mapeamento das categorias por trimestre de 2015 a 2019.

Tabela 16 – Distribuição das categorias exploratórias encontradas nas notícias analisadas sobre sífilis por ano, entre 2015 e 2019.

Categoria Exploratória	2015	2016	2017	2018	2019
Número anual de notícias <i>online</i> sobre sífilis	18 (2,86%)	26 (4,15%)	42 (6,67%)	107 (16,98%)	437 (69,37%)
Campanha	11	16	32	78	308
Consequências	15	21	41	91	352
Definição da Doença	11	14	26	67	241
Índices Epidemiológicos	13	11	29	55	243
Prevenção	15	21	35	86	360
Teste Rápido/Diagnóstico	13	25	39	92	382
Tratamento	16	22	34	88	348
Nenhuma das categorias	1	1	0	11	22
Número total de notícias	630				

Cada célula representa o número de notícias com as respectivas categorias. Em 2015, o maior número de notícias foi no segundo trimestre (abril a junho), citando os termos “prevenção”, “consequências” e “tratamento”. Em 2016 e 2017 o maior número de notícias foi no quarto trimestre (outubro-dezembro), citando os termos “campanha”, “consequências”, “teste/diagnóstico” e “tratamento”, coincidindo com o Dia Nacional de Combate à Sífilis e Sífilis Congênita em outubro. No resto do ano, essas categorias quase nunca ou nunca foram mencionadas nas notícias.

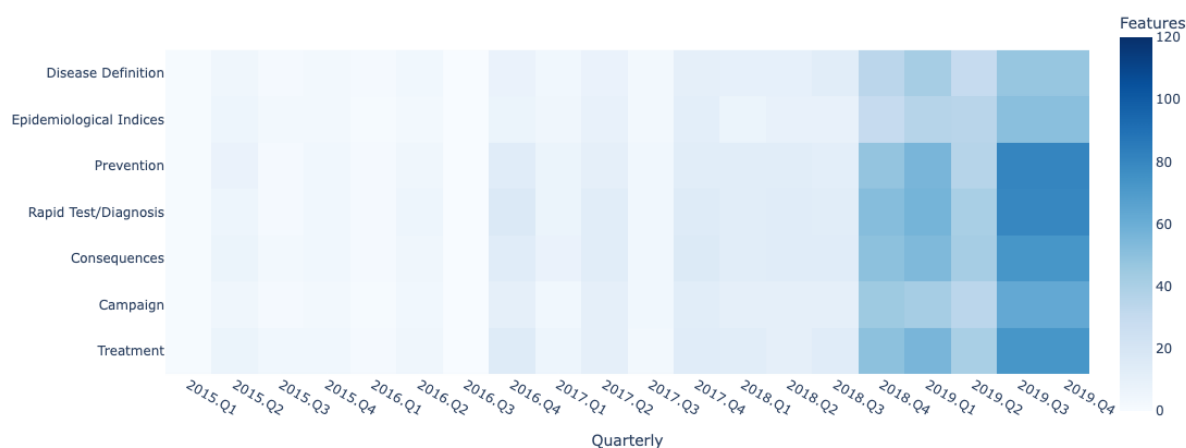


Figura 32 – Mapa de calor de categorias por trimestre entre 2015 e 2019.

Em 2018 e 2019, embora houvesse uma preponderância de notícias no 4º trimestre de cada ano, e em particular em outubro, houve notícias ao longo de todo o ano. A maioria das notícias mencionou “campanha”, “consequências”, “teste/diagnóstico rápido”, “tratamento” e “prevenção”.

A Tabela 17 mostra o número de notícias categorizadas por qualidade. Embora o volume de notícias tenha aumentado substancialmente em 2018 e 2019, a proporção relativa de notícias de alta qualidade permaneceu consistentemente alta (77,6% e 70,5%, respectivamente) e em linha com anos semelhantes.

Tabela 17 – Número de notícias a cada ano categorizadas por qualidade.

Nível de qualidade	2015	2016	2017	2018	2019
	n (% do total)	n (% do total)	n (% do total)	n (% do total)	n (% do total)
Alta	13 (72,2%)	19 (73,1%)	33 (78,6%)	83 (77,6%)	308 (70,5%)
Média	2	2	6	10	71
Baixa	3	5	3	14	58
Total	18	26	42	107	437

Houve correlação entre a qualidade das notícias sobre sífilis e o número de testes de triagem realizados na atenção primária à saúde para diagnóstico de sífilis conforme Tabela 18. Os testes rápidos para sífilis em gestantes ou pais/companheiros apresentaram correlação de 65,33% com notícias de alta qualidade, enquanto os testes rápidos na população geral tiveram correlação de 58,98%. Os testes não treponêmicos para sífilis mostraram uma correlação de 69,15% com as notícias de alta qualidade e os testes treponêmicos, (testes confirmatórios para sífilis) mostraram 82,32%, 78,13% e 73,20% no mesmo período.

Tabela 18 – Correlação entre o número de testes e o número de notícias de acordo com sua qualidade.

Tipos de Teste	Qualidade da Notícia		
	Alta	Média	Baixa
Teste treponêmico para detecção de sífilis	73.20%	67.15%	62.78%
Teste não treponêmico para detecção de sífilis	69.15%	62.04%	58.66%
Teste FTA-ABS IgG para diagnóstico da sífilis	78.13%	62.12%	64.58%
Teste FTA-ABS IgM para diagnóstico da sífilis	82.32%	68.84%	68.31%
Teste não treponêmico para detectar sífilis em gestantes	64,92%	58,76%	50,85%
Teste rápido de sífilis	58.98%	46.31%	42.64%
Teste rápido de sífilis para detectar a infecção em mulheres grávidas ou pais/parceiros	65.33%	56.27%	52.07%

8.3 Discussão

O estudo revela a importância de campanhas de mídia para apoiar a adoção e difusão de intervenções em saúde pública. O número de notícias sobre sífilis aumentou rapidamente após a implementação da estratégia de comunicação como parte do Projeto “Sífilis Não!”. Embora tenha havido um total acumulado de 86 notícias sobre sífilis em 2015 a 2017, o número de notícias subiu para 107 em 2018 (aumento de 154,8% em relação a 2017) e 308 em 2019 (aumento de 308,4% em relação a 2018).

Em 2010, a sífilis tornou-se uma doença de notificação compulsória no Brasil e foi declarada emergência de saúde pública em 2016 (SAÚDE., 2019; TRANSMISSÍVEIS, 2017). No entanto, essas grandes políticas de saúde pública não levaram a uma mudança no número de notícias ou aumento de testagem e diagnóstico. Em maio de 2015, o MS realizou uma campanha sobre transmissão vertical (MTCT) para o Dia das Mães e em 2016 e 2017 realizou campanhas de comunicação no Dia Nacional de Combate à Sífilis e Sífilis Congênita. Entretanto, estes foram eventos pontuais (MACHADO; NOSCHANG, 2020) e não tiveram efeito significativo nas notícias e testes.

O lançamento das ações de comunicação do PSN em fevereiro de 2018 coincidindo com o Carnaval brasileiro foi um divisor de águas. Seguiu-se uma campanha nas redes sociais com iniciativas locais e estaduais. Em março de 2018, o PSN lançou agendas de intervenção em todo o país por meio do Ministério da Saúde, com ênfase em 100 municípios prioritários (PINTO et al., 2022; PINTO et al., 2021; ANDRADE; VALENTIM; OLIVEIRA, 2020). Em novembro de 2018, foi lançada a campanha nacional “Sífilis Não!”(2018–2019). A campanha de comunicação fez com que as ações deixassem de se limitar a eventos anuais pontuais, como o Dia Nacional de Combate à Sífilis e Sífilis Congênita, mas se espalhassem ao longo do ano (MACHADO; NOSCHANG, 2020) envolvendo uma infinidade de produtos e ações de comunicação, inclusive em ‘mídias pagas e próprias’ que ajudaram a aumentar as notícias na agenda midiática.

A campanha “Sífilis Não!”envolveu Apoiadores de Pesquisa e Intervenção cujas ações ajudaram a melhorar a coordenação dos esforços de comunicação e mensagens entre as secretarias estaduais e municipais de saúde e a imprensa local, atuando como um “líder de opinião”, para estabelecer um “duplo-fluxo da comunicação” (SHARMA; NAHAK, 2021) para garantir notícias de alta qualidade. As ações de comunicação lideradas pelo PSN produziram maior “ressonância”, com a imprensa local e nacional (PESSONI; CARMO, 2014).

McCombs (MCCOMBS, 2015) argumenta que o processo de definição da agenda depende não apenas do período de exposição na mídia que um tópico tem, mas também da potencial correspondência com a necessidade de orientação do público sobre o tópico. Os resultados revelam um aumento substancial de notícias na mídia e em testes de triagem, com correlação entre os dois, sugerindo que uma campanha de comunicação sustentada pode ser uma ferramenta poderosa para promover políticas públicas de enfrentamento às crises de saúde.

O Ministério da Saúde registrou o maior número de testes de sífilis realizados por 1.000 habitantes em 2018 (2,1 milhões) e 2019 (2,5 milhões), comparado a 1,4 milhão em 2017. Enquanto em 2010–2018, o número de casos de sífilis adquirida, sífilis durante a gravidez e sífilis congênita aumentaram substancialmente, mas caíram em 2019 (PINTO et al., 2022; PINTO et al., 2021; VALENTIM et al., 2022; ANDRADE; VALENTIM; OLIVEIRA, 2020).

Existem potenciais limitações do nosso estudo. A primeira diz respeito à completude da coleção de notícias utilizada como conjunto de dados, ameaça externa mitigada pela escolha de um dos maiores indexadores de conteúdo existentes, o Google Search. No entanto, seria um erro supor que o Hermes pode recuperar todas as notícias *online* existentes por meio da API de pesquisa do Google. Além disso, o Hermes utilizou os seguintes filtros: i) idioma, que restringe a busca a documentos escritos em português do Brasil, e ii) geolocalização, que limita os resultados da busca a documentos originados no Brasil. Isso pode constituir uma barreira na medida em que o parâmetro de geolocalização verifica o domínio (URL) e a localização geográfica do endereço IP do servidor Web. A segunda limitação está relacionada aos resultados de pesquisa relacionados às 42 páginas ausentes indexadas pelo Google. Se tivessem sido incorporados à análise, provavelmente aumentariam o número de notícias resultantes. No entanto, ao analisar a data de notícias de páginas não encontradas, observamos que elas tiveram a mesma proporcionalidade de resultados: 2015 (n=1), 2016 (n=3), 2017 (n=1), 2018 (n=15) e 2019 (n=22).

Apesar das limitações, o estudo revela a aplicação efetiva de um ecossistema de saúde digital que incorpora todos os elementos de uma complexa intervenção de saúde pública, previamente identificada pelo nosso framework, que incluiu uma campanha de comunicação, educação, intervenções no sistema de saúde, treinamento, ampliação do acesso a testes e tratamento.

Este estudo do Brasil, um país de renda média alta, liderado por um grupo multidisciplinar de pesquisadores envolvendo especialistas em saúde pública, médicos, especialistas em ciência da computação, cientistas de dados, especialistas em comunicação e especialistas em marketing revela o sucesso da implementação de ações de saúde pública com uma campanha de comunicação que levou a grandes aumentos nas notícias *online* relacionadas à sífilis, à mudança no cenário da mídia e na resposta da saúde pública após as intervenções do Projeto “Sífilis Não!”.

O ecossistema Hermes foi capaz de capturar efetivamente o número e a frequência de notícias antes e depois do Projeto “Sífilis Não!”, classificar efetivamente as notícias de acordo com as categorias exploratórias, identificar notícias de alta, média e baixa qualidade e examinar a relação entre a campanha de comunicação e resultados de saúde pública para orientar políticas e intervenções de saúde mais efetivas e direcionadas para o manejo da epidemia de sífilis no Brasil.

O estudo revela a utilidade de sistemas integrados de informação em saúde digital, baseados em um *framework* multidimensional, na orientação de políticas e ações para garantir respostas efetivas aos desafios de saúde pública em todo o país.

9 Estudo sobre Hepatites Virais

Com o objetivo de demonstrar a eficácia do *framework* proposto, utilizando um estudo de caso fora do escopo do Projeto “Sífilis Não!”, realizamos um estudo com dados sobre hepatites no Brasil. O objetivo geral deste estudo é compreender como se comportaram os dados epidemiológicos e de comunicação sobre hepatites ao longo do tempo, a fim de observar possíveis mudanças de tendências e correlações com variáveis de interesse em diferentes dimensões.

Embora o Brasil não apresente epidemia de hepatites virais, como no caso da sífilis, nem tenha havido intervenção nacional no combate a essa doença, optamos por esse agravo uma vez que os dados estavam disponíveis na *Internet* e no portal do Ministério da Saúde.

9.1 Métodos

Para realizar este estudo, analisamos os dados do SIA/SUS sobre hepatites virais, ou seja, hepatites A, B e C. O ecossistema Hermes coletou os dados disponíveis conforme tabela 19. Nessa etapa, foi possível compreender um pouco do contexto das hepatites, identificando diferentes tipos de testes de triagem e tratamentos implementados pelo SUS.

Uma vez identificados esses dados, analisamos a viabilidade de processá-los individualmente ou em conjunto. Identificamos que a quantidade de tipos de dados epidemiológicos era extensa (23 tipos diferentes de exames de detecção de hepatites e oito tipos diferentes de tratamentos) e que não tínhamos informações detalhadas sobre a função específica de cada um deles, ou seja, (i) para que serve, (ii) para quem serve, (iii) se o procedimento foi banido pelo SUS ao longo do período analisado, (iv) se o procedimento iniciou-se/foi encontrado meses antes ou após o período analisado. Assim, resolvemos unificá-los e considerá-los como ‘Hepatites Virais’.

Outro aspecto desta análise diz respeito à dimensão da comunicação, onde analisamos (i) dados do Google Trends a fim de observar o interesse da população em buscar pelo termo ‘hepatite’ no Google e (ii) dados das notícias online indexadas pelo Google com o termo ‘hepatite’.

É importante ressaltar que embora o Google retorne, através de sua API, páginas *online* com o termo procurado, nem todas as páginas podem ser classificadas como notícias. Isso porque o Google retorna, dentre estas notícias, documentos, artigos científicos, páginas que estão fora do ar, resultado de buscas, dentre outros. Atualmente o processo de identificação de notícias *online* é feito manualmente no Hermes, analisando cada registro dentro da interface do Hermes.

Tabela 19 – Dados obtidos pelo ecossistema Hermes para avaliar a hepatite no Brasil, ao longo do tempo.

Teste de Triagem
Histopatologia para Identificacao da Hepatite A
Histoquímica para Identificação do Vírus da Hepatite A
Pesquisa de Anticorpos IGG Contra o Vírus da Hepatite A (Hav-Igg)
Pesquisa de Anticorpos IGM Contra o Vírus da Hepatite A (Hav-Igg)
Histopatologia para Identificação de Hepatite B
Identificação do Vírus da Hepatite B Por Pcr (Quantitativo)
Imunohistoquímica para Identificação do Vírus da Hepatite B
Pesquisa de Anticorpos contra Antígeno de Superfície do Virus da Hepatite B (Anti-Hbs)
Pesquisa de Anticorpos contra Antígeno e do Vírus da Hepatite B (Anti-Hbe)
Pesquisa de Anticorpos IGG e IGM contra Antígeno Central do Vírus da Hepatite B (Anti-Hbc-Total)
Pesquisa de Anticorpos IGM contra Antígeno Central do Vírus da Hepatite B (Anti-Hbc-Igm)
Pesquisa De Antígeno De Superfície do Vírus da Hepatite B (HBSAG)
Pesquisa de Antígeno e do Vírus da Hepatite B (HBEAG)
Detecção de RNA do Vírus da Hepatite C (Qualitativo)
Genotipagem de Vírus da Hepatite C
Histopatologia para Identificação de Hepatite C
Histoquímica para Identificação de Hepatite C
Pesquisa de Anticorpos Contra o Vírus da Hepatite C (Anti-HCV)
Quantificação De RNA do Vírus da Hepatite C
Teste Rápido para Detecção de Hepatite C
Tratamento
Hemodiálise em Paciente com Sorologia HIV Positiva e/ou Hepatite B e/ou Hepatite C (Tratamento Excepcional)
Hemodiálise em Paciente com Sorologia HIV Positiva e/ou Hepatite B e/ou Hepatite C (Máximo 3)
Imunoglobulina Anti-Hepatite B 100 UI Injetável (Por Frasco)
Imunoglobulina Anti-Hepatite B 1000 UI Injetável (Por Frasco)
Imunoglobulina Anti-Hepatite B 500 UI Injetável (Por Frasco)
Imunoglobulina Anti-Hepatite B 600 UI Injetável (Por Frasco)
Imunoglobulina Humana Anti-hepatite B (por dose de 500 UI)
Tratamento de Hepatites Virais
Dados de Comunicação
Notícias Online sobre Hepatite
Interesse ao longo do tempo (procura pelo termo 'hepatite' na Internet)

Os dados identificados nestas 4 variáveis de interesse (testes de triagem, tratamento para hepatite, notícias *online* e interesse ao da população pelo termo 'hepatite') foram agrupados mensalmente, no período de janeiro de 2015 a dezembro de 2019.

9.2 Resultados

Após coletar e processar os dados, podemos analisar os resultados individualmente a partir de gráficos de tendência e sazonalidade. A Figura 33 apresenta os resultados relacionados aos exames para detecção de hepatites virais. Neste gráfico, foi possível observar uma queda no número de exames no período analisado, embora tenha sido possível observar um leve aumento no segundo semestre de 2017. Em relação à sazonalidade, é possível observar um pico de exames em março e novembro e uma queda acentuada em fevereiro.

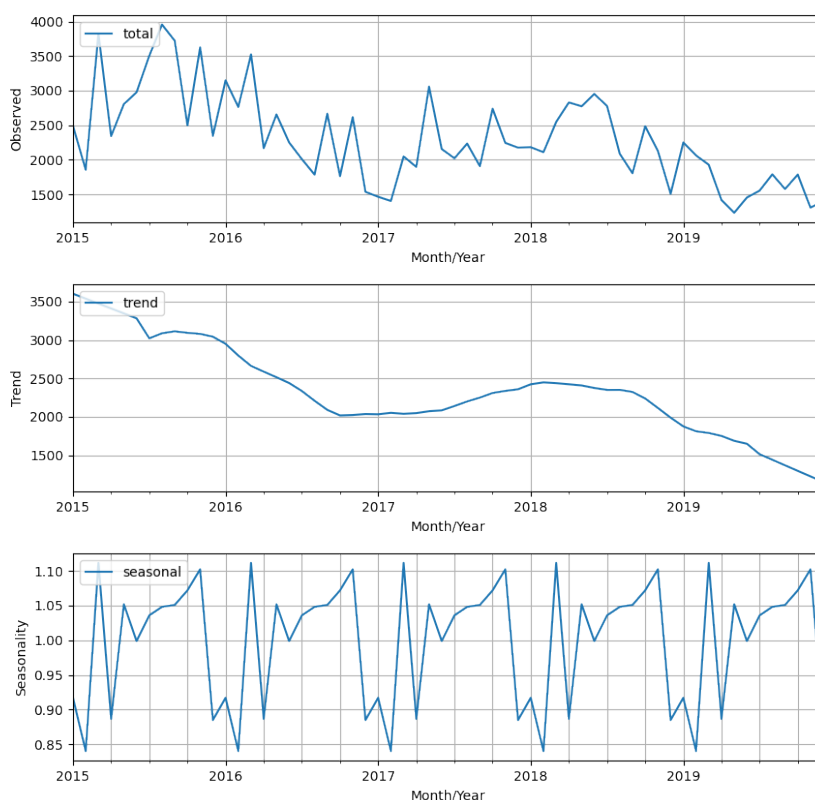


Figura 33 – Gráfico de tendência e sazonalidade dos testes de detecção de hepatite viral entre 2015 e 2019.

Em relação aos tratamentos para Hepatites Virais realizados pelo SUS, foi possível observar (Figura 34) um crescimento abrupto em 2015 e estabilização entre 2016 e 2019. A sazonalidade desses dados tem pico em julho e agosto e queda em fevereiro.

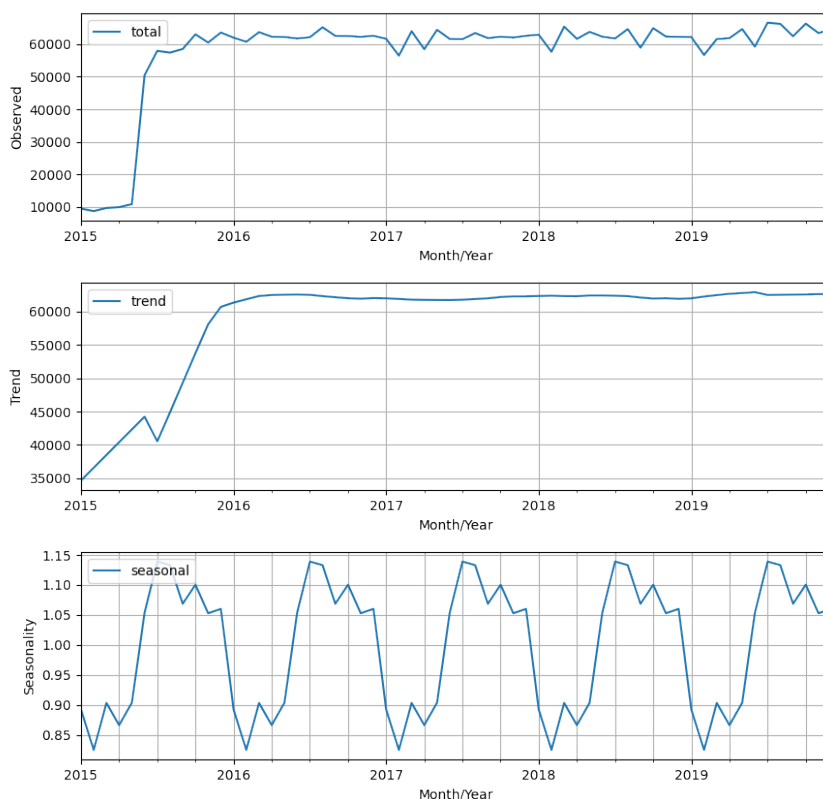


Figura 34 – Gráfico de tendências e sazonalidade de tratamentos para Hepatites Virais entre 2015 e 2019.

Um dos requisitos para usar o algoritmo de decomposição de séries temporais é que cada registro da série histórica seja composto por um valor diferente de zero/nulo ao longo do tempo. Portanto, esse algoritmo não conseguiu processar dados de notícias online, pois alguns meses não apresentavam notícias.

Assim, analisamos a série usando a Tabela 20. É possível observar o aumento gradativo do número de notícias sobre Hepatite ao longo dos anos. A taxa de crescimento foi maior em 2017 (crescimento de 85,71% em relação a 2016), e os anos seguintes apresentaram taxas de crescimento menores, mas o volume de notícias aumentou. Em 2018, foram identificadas 54 notícias (taxa de crescimento de 38,46%) e 60 notícias em 2019 (taxa de crescimento de 11,11%).

É importante notar que, embora o Google tenha encontrado 525 páginas da web com o termo 'hepatite' no período analisado, uma minuciosa análise manual teve como objetivo identificar quais eram de fato notícias. Assim, as páginas foram classificadas da seguinte forma: 194 notícias, 137 produções científicas, 124 mídias não textuais (Podcast/Vídeo/PDF/PPT/DOCX), 54 páginas do tipo FAQ (Frequently Asked Questions), 15 páginas de 'resultados de pesquisa', e 1 página não encontrada.

Tabela 20 – Número de notícias sobre hepatite indexadas pelo Google entre 2015 e 2019.
*Taxa de crescimento em relação ao ano anterior.

Mês	2015	2016	2017	2018	2019
Janeiro	-	2	3	9	1
Fevereiro	1	-	4	3	3
Março	5	2	3	6	3
Abril	-	-	4	1	6
Mai	2	2	-	2	6
Junho	1	-	5	1	1
Julho	3	7	4	10	17
Agosto	1	3	3	6	6
Setembro	1	-	5	1	3
Outubro	4	2	4	3	4
Novembro	2	-	3	4	7
Dezembro	-	3	1	8	3
Total	20	21	39	54	60
Taxa de crescimento*	-	5%	85.71%	38.46%	11.11%

Os dados do Google Trends (Figura 35) agrupados semanalmente apresentaram uma tendência de queda entre o início de 2015 até o final de 2016, tendência de alta acentuada em 2017, uma queda em 2018 e novamente alta em 2019. A sazonalidade desta variável de interesse apresentou um pico no início do segundo semestre e queda no final do segundo semestre. Nesta análise, não especificamos o mês, pois os dados dessa série são detalhados por semana, diferente das demais que são mensais.

Por fim, os resultados dos testes de correlação estatística entre Notícias Online, Tratamentos e Testes de Triagem usando três métodos diferentes (Pearson, Kendall, Spearman), retornaram valores inconsistentes com possíveis correlações.

A tabela 21 apresenta esses resultados pela variável de interesse. A análise dessa distribuição foi feita mês a mês, portanto, os dados do Google Trends não foram avaliados. Seus valores são atribuídos com base em uma faixa de -1 a +1, onde valores negativos indicam uma correlação contrária, ou seja, quando uma variável cresce a outra diminui. Valores positivos indicam uma correlação diretamente proporcional, ou seja, quando uma variável cresce, a outra segue o mesmo caminho. Finalmente, um valor igual a zero indica que os conjuntos de dados não estão correlacionados monotonicamente.

9.3 Discussão

Hepatites virais é um problema de saúde pública, assim como a sífilis. A cada ano, 1,4 milhão de pessoas morrem de doenças relacionadas a hepatites virais, como cirrose e câncer de fígado (JEFFERIES et al., 2018). No Brasil, as consequências das hepatites virais atingem diferentes segmentos da população, inclusive os mais vulneráveis, o que pode causar uma sobrecarga no Sistema Único de Saúde (STRAUSS, 2018; SCARPONI

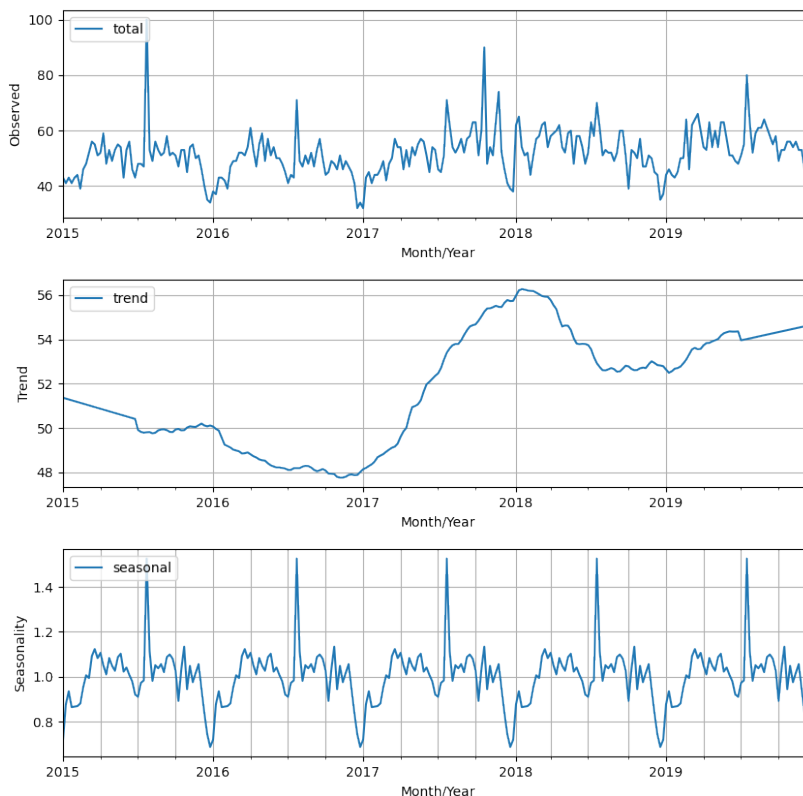


Figura 35 – Gráfico de tendência e sazonalidade do interesse da população pelo termo 'hepatite' usando o mecanismo de pesquisa do Google.

Tabela 21 – Resultado do processamento de correlação entre variáveis de interesse usando três métodos diferentes.

	Notícias Online	Tratamento	Testes de Triagem
Pearson			
Notícias Online	1.000000	0.220108	-0.338344
Tratamento	0.220108	1.000000	-0.204530
Testes de Triagem	-0.338344	-0.204530	1.000000
Kendall			
Notícias Online	1.000000	0.203390	-0.298305
Tratamento	0.203390	1.000000	-0.154802
Testes de Triagem	-0.298305	-0.154802	1.000000
Spearman			
Notícias Online	1.000000	0.296733	-0.431628
Tratamento	0.296733	1.000000	-0.237642
Testes de Triagem	-0.431628	-0.237642	1.000000

et al., 2022).

Conforme observado na Tabela 19 o governo brasileiro oferece tratamento contra as Hepatites A, B e C gratuitamente, disponível para os pacientes do SUS. Nesse sentido, podemos observar estabilidade nos tratamentos a partir de 2016.

Esse movimento de estabilidade no tratamento das hepatites, aliado à queda no número de exames, pode ser decorrente da crescente disponibilidade de vacinas no SUS que vem ocorrendo nos últimos anos (NEVES et al., 2022). A partir de julho de 2014, a vacina contra hepatite A foi ofertada pelo SUS.

Segundo o Sitnik et al. (2021), o caminho para a eliminação das hepatites virais depende do conhecimento do impacto dessas doenças nas diferentes regiões, da identificação precisa de novos casos dessas infecções, da interrupção da cadeia de transmissão da doença e finalmente, da administração de tratamento adequado para todos os pacientes infectados.

As campanhas de saúde pública desempenham um papel fundamental na condução da população para a mudança de hábitos e comportamentos por meio de mensagens de conscientização. Pinto et al. (2021) sugere diferentes estratégias como parte de uma abordagem multifacetada para reduzir a carga de infecção pelo vírus da hepatite B (HBV), considerando as diferentes realidades regionais no Brasil, tais como:

Norte Priorizar ações e abordar a transmissão sexual principalmente para jovens, bem como ampliar o acesso ao preservativo e fornecer orientação sobre sexo seguro;

Nordeste Ampliar as ações educativas para prevenção e controle da hepatite B em jovens;

Centro-Oeste Ampliar as ações educativas para prevenção e controle da hepatite B em jovens;

Sul Incentivar a adoção de boas práticas nos serviços de saúde, para reduzir o risco de transmissão viral, principalmente para maiores de 40 anos; Incentivar a vacinação da população adulta; Fortalecer as ações de redução da transmissão vertical; Montar salas de vacinação na maternidade;

Sudeste Melhorar a qualidade da Vigilância Epidemiológica, principalmente no que diz respeito ao preenchimento adequado das fichas de investigação epidemiológica;

As notícias *online* também têm seu papel nesse processo de divulgação de mensagens para a população. Analisando esses dados, percebemos um crescimento no volume de notícias *online* ano a ano. No entanto, como não identificamos uma intervenção pública e nacional no combate a essa doença. Esse volume crescente e espontâneo mostrou-se muito pequeno em relação às notícias sobre sífilis (apresentadas no Capítulo 8). Apesar disso, esse fato demonstra a viabilidade da intervenção da saúde pública para divulgar o tema na mídia, principalmente na Internet.

Não foi possível processar as notícias *online* quanto à qualidade, uma vez que utilizamos um modelo supervisionado para avaliar as notícias sobre sífilis, o que requer uma amostra para treinamento. Trabalhos futuros devem aprimorar essa funcionalidade no ecossistema Hermes para possibilitar uma análise genérica da notícia de qualidade com base no conceito 5W1H, que deve identificar se uma notícia possui os seguintes elementos: Quem, Por que, Como, O quê, Onde e Quando.

Por fim, os dados do Google Trends mostraram picos e declínios ao longo dos anos. Em 2017 houve maior interesse pelo termo na internet, o que sugere que as notícias *online* publicadas no período podem ter ajudado nesse pico e impulsionado as buscas. Infelizmente, não encontramos estudos que validem essa hipótese, e os resultados dos testes estatísticos não corroboram essa afirmação. Possivelmente outra variável esteja motivando o declínio dos exames e a estabilização do tratamento.

10 Conclusões

Esta tese tem como objetivo desenvolver um *framework* para análise multidimensional do alcance de intervenções em saúde pública. Esse *framework* apresenta quais variáveis de interesse foram menos ou mais impactadas pelas ações de intervenção em saúde pública e identifica correlações entre elas, permitindo que os tomadores de decisão reavaliem novas intervenções. Além disso, desenvolvemos um ecossistema digital que explora métodos de aprendizado de máquina e processamento de linguagem natural para analisar dados de séries temporais e dar suporte ao nosso *framework*.

Nossa hipótese, apresentada na seção 1.3, foi definida da seguinte forma: *Uma análise multidimensional das intervenções em saúde pública é necessária para entender a relevância das variáveis de interesse ao longo do tempo e suas correlações, a fim de permitir que os tomadores de decisão reavaliem as ações de resposta a crises de saúde pública ao projetar intervenções futuras.*

Esta hipótese foi investigada através do *framework* desenhado e dos estudos de caso apresentados anteriormente. O *framework* proposto permitiu uma análise multidimensional de uma intervenção em saúde pública, demonstrando que variáveis de interesse não usuais, como notícias online, podem influenciar outras variáveis e ajudar a impulsionar e manter o projeto de intervenção por mais tempo.

Nosso estudo operacional sobre notícias *online* mostrou que a sífilis estava em pauta na mídia *online* desde o início do Projeto. Mesmo sem adicionar recursos financeiros (notícias pagas) a essa variável, foi possível observar a sustentação do tema na *Internet* após o Projeto. É importante ressaltar que as variáveis epidemiológicas servem como termômetro para atestar o sucesso da intervenção, uma vez que seu objetivo final está diretamente relacionado à melhoria da saúde humana. Assim, os achados do estudo de séries temporais revelaram mudanças de tendência estatisticamente significativas nas taxas mensais de sífilis congênita no Brasil após a implementação do Projeto.

As tecnologias digitais de informação e comunicação em saúde e a aplicação de métodos computacionais baseados em técnicas de Aprendizado de Máquina e Processamento de Linguagem Natural podem ser usadas para desenvolver análises preditivas, para informar a resposta em tempo real a fim de gerenciar efetivamente surtos e crises de saúde pública.

O ecossistema Hermes, desenvolvido para o Projeto, desempenhou um papel crítico na captura de dados heterogêneos e no processamento de informações relacionadas a ações de comunicação, intervenções do sistema de saúde e parâmetros epidemiológicos (número de casos de sífilis adquirida, sífilis em gestantes e sífilis). A utilização de métodos computacionais possibilitou a análise de dados de natureza heterogênea para examinar a resposta da saúde pública e seus resultados em todo o país, além de fornecer uma poderosa ferramenta no planejamento e monitoramento das intervenções em saúde, que pode ser

transferida para outros desafios da saúde pública.

Desta forma, o Hermes mostra claramente evidências de mudança de comportamento na população considerando aspectos epidemiológicos e de comunicação (notícias *online* e interesse da população pelo tema na Internet), além de analisar a correlação entre as variáveis de interesse levantadas durante esta pesquisa, fortalecendo a análise multidimensional discutida nesta tese.

Além disso, os *stakeholders* desempenham um papel fundamental na identificação dos conjuntos de dados e dos processos analíticos para atestar a evidência de mudança de comportamento na população após uma intervenção. Por fim, os cientistas de dados utilizam esse *framework* para possibilitar a aquisição e especificação desses conjuntos de dados, verificando algoritmos viáveis a serem executados uma vez que as condições que devem ser satisfeitas pela técnica precisam estar de acordo com o conjunto de dados transformado para produzir resultados significativos.

Ao longo desse período de doutoramento, contribuições relacionadas a esta pesquisa foram organizadas como artigos, listados a seguir e alguns deles já publicados. Inclusive o artigo intitulado de “Use of Interrupted Time Series Analysis in Understanding the Course of the Congenital Syphilis Epidemic in Brazil” foi capa da revista *The Lancet Regional Health – Americas*.

Artigo 1 Assessing the Impact of Public Health Campaigns Through Epidemiological, Communication and Education Indicators. Situação: **Publicado** na 20th International Conference on Advanced Learning Technologies (ICALT), IEEE, 2020 ([PINTO et al., 2020](#)).

Artigo 2 Systematic Review on Information Technology Approaches to Evaluate the Impact of Public Health Campaigns: Real Cases and Possible Directions. Situação: **Published** in *Frontiers in Public Health*, 2022 ([PINTO et al., 2021](#)).

Artigo 3 Analyzing the Reach of Public Health Campaigns Based on Multidimensional Aspects: the case of the syphilis epidemic in Brazil. Situação: **Publicado** em *BMC Public Health*, Springer, 2021 ([PINTO et al., 2021](#)).

Artigo 4 Use of Interrupted Time Series Analysis in Understanding the Course of the Congenital Syphilis Epidemic in Brazil. Situação: **Publicado** na *The Lancet Regional Health – Americas*, 2022 ([PINTO et al., 2022](#)).

Artigo 5 How Online News Affects Public Health Response to Syphilis in Brazil: Text Mining Mediated Analysis. Situação: **Em Revisão** no *Journal of Medical Internet Research*.

Artigo 6 Similarity Analysis in Understanding the Potential of Intermedia Agenda-Setting in Online News in the Public Health Response. Situação: **Em Revisão** no *Journal of Medical Internet Research*.

Artigo 7 Sentiment Analysis in Understanding the Potential of Online News in the Public Health Crisis Response. Situação: **Em Revisão** em *Frontiers in Public Health*.

Artigo 8 HERMES: A Digital Ecosystem For Monitoring and Assessing Public Health Interventions. Situação: **Escrevendo**.

Nossas publicações se complementam. O artigo 1 foi o primeiro desenho de nossa pesquisa em que focamos na dimensão da comunicação e exploramos duas variáveis de interesse. O artigo 2 foi uma revisão sistemática que resumiu as variáveis de interesse, técnicas e ferramentas empregadas para avaliar campanhas de saúde pública, apresentadas no Capítulo 2. O artigo 3 foi um estudo analítico onde foi possível utilizar o Hermes como solução digital e analisar dados heterogêneos na perspectiva de séries temporais, apresentado no Capítulo 6. O artigo 4 foi realizado com o objetivo de avaliar a dimensão epidemiológica antes e após a implantação do Projeto “Sífilis Não!”, conforme relatado no Capítulo 7. O artigo 5, apresentado no Capítulo 8, foi realizado com o objetivo de buscar evidências para avaliar correlações entre notícias *online* e testes de triagem, destacando a importância desse tipo de mídia para a conscientização da população a mudar hábitos, buscar testes, tratamentos, e curas para doenças infecciosas.

O conteúdo relacionado aos artigos 6 e 7 não foi incluído nesta tese, uma vez que o autor desta tese não é o primeiro autor desses artigos. No entanto, o desenho do estudo, a coleta de dados, a visualização dos dados e os achados fazem parte desta pesquisa, possibilitando uma discussão final sobre a intervenção do Projeto “Sífilis Não!”. Além disso, o código-fonte utilizado para processar os dados foi integrado ao ecossistema Hermes, possibilitando pesquisas futuras. Esses dois artigos se baseiam em notícias on-line, explorando a análise de sentimentos e o agrupamento de conteúdo textual (corpo das notícias) relacionados à sífilis.

Finalmente, o artigo 8 é nosso próximo estudo, onde pretendemos explorar o potencial do ecossistema Hermes como ferramenta para monitorar e avaliar intervenções de saúde pública.

10.1 Trabalhos Relacionados

Frameworks para analisar intervenções de saúde pública surgiram nos últimos anos. A seguir, comparamos os principais trabalhos relacionados a esta tese, destacando as principais diferenças.

Pitt, Goldhaber-Fiebert e Brandeau (2020) fornecem um *framework* gráfico para categorizar e comparar intervenções de saúde pública que examinam a distribuição de benefícios e danos entre e dentro de subgrupos populacionais para uma única intervenção. O grupo utilizou como caso de estudo um aumento hipotético no preço da carne, pelo qual reduziu o consumo de carne nos indivíduos em resposta a essa política de preços, e

projetou a redução resultante no consumo calórico e no IMC de cada indivíduo ao longo do tempo. O ponto focal deste trabalho é analisar a correlação entre dano e benefício em subgrupos dentro da população. Os autores apresentam uma dimensão instigante: o dano causado por uma intervenção (política). Embora essa dimensão seja de difícil avaliação no contexto das doenças sexualmente transmissíveis (objeto de estudo desta tese), trabalhos futuros poderão explorar mais a fundo essa dimensão.

[Brakefield, Ammar e Shaban-Nejad \(2021\)](#) apresentam o Urban Population Health Observatory (UPHO), uma plataforma de análise de dados para integrar um grande volume de vigilância ao CoVID-19 através de dados multidimensionais e heterogêneos, incluindo determinantes sociais de indicadores de saúde, dados clínicos e de saúde da população. Neste estudo, embora o grupo proponha uma análise multidimensional com fontes de dados heterogêneas, o trabalho apresentado apenas menciona as dimensões epidemiológica e sociodemográfica, deixando de lado a dimensão Comunicação, essencial para a compreensão da agenda midiática e do interesse da população na Internet. Além disso, o grupo destaca que a UPHO é um trabalho em andamento e não apresentou um estudo de caso real.

[El-Kassabi et al. \(2021\)](#) propõem um *framework* para analisar a enorme quantidade de reportagens públicas sobre Covid-19 na Internet. O *framework* engloba um fluxo de trabalho automatizado, incluindo módulos de pré-processamento de texto e extração de informações de notícias. Este trabalho tem uma intersecção interessante com a nossa tese. A dimensão 'Cloud' definida pelo grupo, ou seja, Comunicação nesta tese, é explorada, com o objetivo de extrair informações relevantes do conteúdo textual para prever e compreender tendências, conteúdos e conexões significativas para lidar com a disseminação, controle e intervenções da pandemia. No entanto, o grupo não menciona correlação com outras dimensões, como vigilância epidemiológica (por exemplo, exames de triagem e número de casos) e dados sociodemográficos. Este segundo é um desafio para avaliar quais regiões ou populações estão mais expostas às notícias. A correlação entre diferentes dimensões nos ajuda a melhorar intervenções futuras.

[Sohn et al. \(2020\)](#) desenvolveram um *framework* conceitual para avaliar os custos de implementação de intervenções de saúde pública e ilustrou seu uso com exemplos da literatura envolvendo intervenções para tuberculose e HIV/AIDS. O estudo concentrou-se na identificação e estimativa dos custos e benefícios da implementação da intervenção baseada em evidências, que poderiam ser usados para orientar as avaliações econômicas. No entanto, não apresenta uma avaliação de impacto nas intervenções nem incorpora uma solução digital para melhoria do processo.

[Lamos et al. \(2015\)](#) apresentaram um *framework* estatístico para estimar a prevalência de uma campanha de intervenção na população a partir de dados da Internet. O estudo de caso se concentra no programa de vacinação contra influenza lançado na Inglaterra durante a temporada 2013/14. As observações consistem em milhões de consultas de pesquisa geolocalizadas no mecanismo de pesquisa Bing e postagens no Twitter. Esta

pesquisa apresenta os conceitos básicos discutidos nesta tese para avaliar intervenções em saúde. No entanto, não se concentrou em variáveis multidimensionais, como notícias online, questões sociodemográficas e dados de vigilância epidemiológica.

A tabela 22 resume os principais trabalhos relacionados a esta tese.

Tabela 22 – Comparação de trabalhos relacionados com esta tese.

Autor	Dimensões	Como foi validado?	Incorpora uma solução tecnológica?	Apresenta uma solução gráfica para análise dos resultados?	Apresenta uma abordagem de séries temporais?	Apresenta uma análise de impacto das intervenções em saúde?	Apresenta correlações entre as variáveis de interesse?
Pitt, Goldhaber-Fiebert e Brandeau (2020)	Saúde da População; Sociodemografia	Estudo de caso do aumento hipotético no preço da carne	Não	Sim	Sim	Sim	Sim
Brakefield, Ammar e Shaban-Nejad (2021)	Vigilância Epidemiológica; Sociodemografia	Não fornecido	Não	Sim	Sim	Não	Sim
El-Kassabi et al. (2021)	Comunicação	Classificação de Notícias sobre Covid	Sim	Sim	Sim	Não	Não
Sohn et al. (2020)	Finanças	Exemplos da literatura	Não	Não	Não	Não	Não
Lamos et al. (2015)	Comunicação	Estudo de caso real sobre vacinação contra influenza na Inglaterra	Não	Sim	Sim	Sim	Sim
Esta tese	Campanha; Comunicação; Educação; Vigilância Epidemiológica; Finança; e Sociodemografia	Estudo de casos reais	Sim	Sim	Sim	Sim	Sim

10.2 Trabalhos Futuros

A pesquisa apresentada nesta tese pode ser estendida em várias direções, entre as quais se destacam:

Aplicar o *framework* proposto em novos estudos de caso, fora do contexto da sífilis, mas dentro de uma intervenção de saúde pública. Raciocínio: Nesta tese, realizamos diversos estudos no contexto do Projeto “Sífilis Não!”. Este Projeto contemplou uma complexa intervenção em saúde pública em nível nacional. Também realizamos um estudo

sobre as hepatites, um problema de saúde pública, já que 1,4 milhão de pessoas morrem no mundo por doenças relacionadas às hepatites virais, como cirrose e câncer de fígado. No entanto, este segundo estudo não possui uma intervenção de saúde pública no Brasil. A ideia é que novos estudos de caso melhorem o *framework* proposto, levantando novas variáveis de interesse e corroborando com as já discutidas nesta tese.

Realizar uma avaliação do *framework* por especialistas em intervenções de saúde pública para identificar possíveis contribuições na atuação desses profissionais. Justificativa: Este *framework* atende principalmente especialistas e prospectores de intervenções de saúde pública, ou seja, gestores interessados em modificar futuras intervenções para alcançar melhores resultados com base na análise de ações anteriores. Dessa forma, tais especialistas devem avaliar esse *framework* para listar, em sua opinião, os pontos fortes e possíveis melhorias.

Explorar variáveis sociodemográficas. Justificativa: Nos estudos realizados durante esta tese, o volume de dados e o surgimento de variáveis de interesse não exploradas na literatura (como notícias online) levaram o estudo a uma análise exploratória, desconsiderando, em alguns casos, contextos sociodemográficos. As análises de novos estudos devem incluir essa dimensão como agrupamento de geolocalização e considerar fatores socioeconômicos, para melhor direcionar a discussão do impacto da intervenção.

Definir e testar uma nova variável de interesse na dimensão Comunicação: palavras de notícias relevantes ao longo do tempo. Raciocínio: Ao extrair palavras relevantes em notícias, é possível identificar o que mais se destacou ao longo do tempo e buscar correlações com outras variáveis. Esse insight ajudará os *stakeholders* a identificar mudanças históricas que podem passar despercebidas diante do volume de notícias. Por exemplo, as notícias sobre hepatite por um tempo destacaram palavras como “consequências” e “aumento de casos”. Após uma data específica, as notícias destacaram a palavra “vacina”; simultaneamente, o número de casos diminuiu. Assim, cria-se a hipótese de que a vacina foi introduzida após um determinado período, reduzindo o número de casos da doença.

Melhorar o ecossistema Hermes para otimizar tarefas e facilitar a análise de informações. Justificativa: O ecossistema Hermes, amplamente discutido no Capítulo 4, necessita de alguns ajustes para otimizar suas tarefas, como (i) identificação automática de notícias; (ii) criação de *dashboard* dinâmico para facilitar a navegação dos especialistas; e (iii) implementação de novos gráficos para comparação de variáveis de interesse em um único gráfico;

Referências

- AHMED, N. et al. Attitudes to chlamydia screening elicited using the social networking site facebook for subject recruitment. *Sexual health*, CSIRO, v. 10, n. 3, p. 224–228, 2013. [28](#), [38](#), [40](#), [41](#), [139](#)
- ALLEM, J.-P. et al. Campaigns and counter campaigns: reactions on twitter to e-cigarette education. *Tobacco control*, BMJ Publishing Group Ltd, v. 26, n. 2, p. 226–229, 2017. [28](#), [38](#), [40](#), [41](#), [138](#)
- ALMEIDA, R. J. A. *LeIA - Léxico para Inferência Adaptada*. [S.l.]: GitHub, 2018. <<https://github.com/rafjaa/LeIA>>. [62](#)
- ANDRADE, I. G. M. d.; VALENTIM, R. A. d. M.; OLIVEIRA, C. A. P. d. The influence of the no syphilis project on congenital syphilis admissions between 2018 and 2019. *Brazilian Journal of Sexually Transmitted Diseases*, p. 1–6, 2020. [97](#), [108](#)
- ANSARI, F. et al. Outcomes of an intervention to improve hospital antibiotic prescribing: interrupted time series with segmented regression analysis. *Journal of Antimicrobial Chemotherapy*, Oxford University Press, v. 52, n. 5, p. 842–848, 2003. [31](#), [63](#)
- ARAÚJO, K. et al. The production of open educational resources as an alternative for training volunteer health workers in rural communities in tanzania. <https://wcol2019.ie>, p. 39, 2019. [51](#)
- ARDÈVOL-ABREU, A.; ZÚÑIGA, H. G. de; MCCOMBS, M. E. Orígenes y desarrollo de la teoría de la agenda setting en comunicación. tendencias en españa (2014-2019). *El profesional de la información (EPI)*, v. 29, n. 4, 2020. [30](#)
- BAAS, J. et al. Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quantitative Science Studies*, MIT Press, v. 1, n. 1, p. 377–386, 2020. [33](#)
- BERMAN, F. et al. Realizing the potential of data science. *Commun. ACM*, Association for Computing Machinery, New York, NY, USA, v. 61, n. 4, p. 67–72, mar. 2018. ISSN 0001-0782. Disponível em: <<https://doi.org/10.1145/3188721>>. [45](#), [46](#)
- BETANCOURT, R.; CHEN, S. pandas library. In: *Python for SAS Users*. [S.l.]: Springer, 2019. p. 65–109. [53](#)
- BLOKH, I.; ALEXANDROV, V. News clustering based on similarity analysis. *Procedia computer science*, Elsevier, v. 122, p. 715–719, 2017. [30](#)
- BOYD, S. Trend: a system for generating intelligent descriptions of time series data. In: CITESEER. *Proceedings of the IEEE international conference on intelligent processing systems (ICIPS-1998)*. [S.l.], 1998. [31](#)
- BRAKEFIELD, W. S.; AMMAR, N.; SHABAN-NEJAD, A. Upho: Leveraging an explainable multimodal big data analytics framework for covid-19 surveillance and research. In: IEEE. *2021 IEEE International Conference on Big Data (Big Data)*. [S.l.], 2021. p. 5854–5858. [121](#), [122](#)

- BRISS, P. A. et al. Developing an evidence-based guide to community preventive services—methods. *American journal of preventive medicine*, Elsevier, v. 18, n. 1, p. 35–43, 2000. [31](#)
- BUCCHI, M.; TRENCH, B. Health campaign research: Enduring challenges and new developments. In: *Routledge handbook of public communication of science and technology*. [S.l.]: Routledge, 2014. p. 214–229. [28](#)
- BULL, S. S. et al. Social media–delivered sexual health intervention: a cluster randomized controlled trial. *American journal of preventive medicine*, Elsevier, v. 43, n. 5, p. 467–474, 2012. [28](#), [38](#), [40](#), [41](#), [137](#)
- CAMP, W. Formulating and evaluating theoretical frameworks for career and technical education research. *Journal of Vocational Education Research*, Association for Career and Technical Education Research, v. 26, n. 1, p. 4–25, 2001. [44](#)
- CARROLL, N. Application of segmented regression analysis to the kaiser permanente colorado critical drug interaction program. In: *Proceedings of the Western Users of SAS Software 2008 Conference*. [S.l.: s.n.], 2008. p. 5–7. [49](#), [86](#), [87](#)
- CASTILLO, J. A. G. D. et al. Social networks as tools for the prevention and promotion of health among youth. *Psicologia: Reflexão e Crítica*, Springer, v. 33, n. 1, p. 1–9, 2020. [51](#)
- CASTRO, M. C. et al. Brazil’s unified health system: the first 30 years and prospects for the future. *The lancet*, Elsevier, v. 394, n. 10195, p. 345–356, 2019. [97](#)
- CECILIO, L. C. d. O.; REIS, A. A. C. d. Notes on persistent challenges for basic health care in brazil. *Cadernos de Saúde Pública*, SciELO Brasil, v. 34, 2018. [97](#)
- CHAVES, L. A. et al. Desabastecimento de medicamentos na literatura científica da saúde: uma revisão narrativa. *PHYSIS: revista de saúde coletiva*, SciELO Brasil, v. 29, 2019. [96](#)
- COMPUTING, R. F. F. S. et al. R: A language and environment for statistical computing. *Vienna: R Core Team*, 2013. Disponível em: <http://www.r-project.org>. [53](#)
- CRAIG, P. et al. Natural experiments: an overview of methods, approaches, and contributions to public health intervention research. *Annual review of public health*, Annual Reviews, v. 38, p. 39–56, 2017. [17](#)
- DEERWESTER, S. et al. Indexing by latent semantic analysis. *Journal of the American society for information science*, Wiley Online Library, v. 41, n. 6, p. 391–407, 1990. [60](#)
- DORFMAN, L.; ERVICE, J.; WOODRUFF, K. Voices for change: A taxonomy of public communications campaigns and their evaluation challenges. *Communications Consortium Media Center, Media Evaluation Project, Washington, DC*, 2002. [17](#), [28](#), [33](#)
- DOWSHEN, N. et al. Iknowshould2: feasibility of a youth-driven social media campaign to promote sti and hiv testing among adolescents in philadelphia. *AIDS and Behavior*, Springer, v. 19, n. 2, p. 106–111, 2015. [17](#), [28](#), [38](#), [40](#), [41](#), [138](#)
- DREHER, P. C. et al. Use of google trends to track online behavior and interest in kidney stone surgery. *Urology*, Elsevier, v. 121, p. 74–78, 2018. [42](#)

- EL-KASSABI, H. T. et al. A cloud-based framework for covid-19 media classification, information extraction, and trends analysis. In: IEEE. *2021 IEEE Cloud Summit (Cloud Summit)*. [S.l.], 2021. p. 7–12. [121](#), [122](#)
- EMERY, S. L. et al. Are you scared yet? evaluating fear appeal messages in tweets about the tips campaign. *Journal of Communication*, Oxford University Press, v. 64, n. 2, p. 278–295, 2014. [28](#), [38](#), [40](#), [41](#), [137](#)
- ESTER, M. et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In: *kdd*. [S.l.: s.n.], 1996. v. 96, n. 34, p. 226–231. [60](#)
- EWUSIE, J. E. et al. Methods, applications and challenges in the analysis of interrupted time series data: A scoping review. *Journal of Multidisciplinary Healthcare*, Dove Press, v. 13, p. 411, 2020. [18](#)
- FEDERAL, S. *Lei cria o Dia Nacional de Combate à Sífilis e à Sífilis Congênita*. 2017. Disponível em: <https://www12.senado.leg.br/noticias/materias/2017/04/03/lei-cria-o-dia-nacional-de-combate-a-sifilis-e-a-sifilis-congenita>. [74](#)
- FELDMAN, R.; SANGER, J. et al. *The text mining handbook: advanced approaches in analyzing unstructured data*. [S.l.]: Cambridge university press, 2007. [103](#)
- FIGUEIREDO, D. C. M. M. d. et al. Relationship between the supply of syphilis diagnosis and treatment in primary care and incidence of gestational and congenital syphilis. *Cadernos de saude publica*, SciELO Brasil, v. 36, 2020. [96](#)
- FORRESTEL, A. K.; KOVARIK, C. L.; KATZ, K. A. Sexually acquired syphilis: Laboratory diagnosis, management, and prevention. *Journal of the American Academy of Dermatology*, Elsevier, v. 82, n. 1, p. 17–28, 2020. [80](#)
- GABARRON, E. et al. Impact of facebook ads for sexual health promotion via an educational web app: a case study. In: *Healthcare Policy and Reform: Concepts, Methodologies, Tools, and Applications*. [S.l.]: IGI Global, 2019. p. 990–1003. [28](#), [37](#), [40](#), [41](#), [135](#)
- GAMAGE, D. G. et al. Advertising sexual health services that provide sexually transmissible infection screening for rural young people—what works and what doesn't. *Sexual health*, CSIRO, v. 8, n. 3, p. 407–411, 2011. [28](#), [39](#), [40](#), [139](#)
- GAVALI, R. Discovery service for engineering and technology literature through google custom search: A case study. *DESIDOC Journal of Library & Information Technology*, v. 35, n. 6, p. 417–421, 2015. [56](#), [102](#)
- GERVASI, R.; MASTROGIACOMO, L.; FRANCESCHINI, F. A conceptual framework to evaluate human-robot collaboration. *The International Journal of Advanced Manufacturing Technology*, Springer, v. 108, p. 841–865, 2020. [44](#)
- GLEZ-PEÑA, D. et al. Web scraping technologies in an api world. *Briefings in bioinformatics*, Oxford University Press, v. 15, n. 5, p. 788–797, 2014. [47](#)
- GLYNN, R. W. et al. The effect of breast cancer awareness month on internet search activity—a comparison with awareness campaigns for lung and prostate cancer. *BMC cancer*, Springer, v. 11, n. 1, p. 442, 2011. [42](#)

- GRANDPRE, J. et al. Adolescent reactance and anti-smoking campaigns: A theoretical approach. *Health communication*, Taylor & Francis, v. 15, n. 3, p. 349–366, 2003. [28](#), [38](#), [40](#), [42](#), [137](#)
- GREENHALGH, T.; PEACOCK, R. Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. *Bmj*, British Medical Journal Publishing Group, v. 331, n. 7524, p. 1064–1065, 2005. [33](#)
- HARZING, A.-W.; ALAKANGAS, S. Google scholar, scopus and the web of science: a longitudinal and cross-disciplinary comparison. *Scientometrics*, Springer, v. 106, n. 2, p. 787–804, 2016. [33](#)
- HAUKE, J.; KOSSOWSKI, T. Comparison of values of pearson’s and spearman’s correlation coefficients on the same sets of data. *Quaestiones geographicae*, De Gruyter Poland, v. 30, n. 2, p. 87, 2011. [30](#), [105](#)
- HILLAR, G. C. *Django RESTful Web Services: The Easiest Way to Build Python RESTful APIs and Web Services with Django*. [S.l.]: Packt Publishing Ltd, 2018. [53](#)
- HINTON, G. E.; ROWEIS, S. Stochastic neighbor embedding. *Advances in neural information processing systems*, v. 15, 2002. [60](#)
- HONG, T.-P. et al. Using tf-idf to hide sensitive itemsets. *Applied Intelligence*, Springer, v. 38, n. 4, p. 502–510, 2013. [59](#), [103](#)
- JEFFERIES, M. et al. Update on global epidemiology of viral hepatitis and preventive strategies. *World journal of clinical cases*, Baishideng Publishing Group Inc, v. 6, n. 13, p. 589, 2018. [114](#)
- KEELE, S. et al. Guidelines for performing systematic literature reviews in software engineering. In: *Technical report, Ver. 2.3 EBSE Technical Report. EBSE*. [S.l.]: sn, 2007. [32](#)
- KHARDE, V.; SONAWANE, P. et al. Sentiment analysis of twitter data: a survey of techniques. *arXiv preprint arXiv:1601.06971*, 2016. [29](#)
- KITE, J. et al. Impact of the make healthy normal mass media campaign (phase 1) on knowledge, attitudes and behaviours: a cohort study. *Australian and New Zealand journal of public health*, Wiley Online Library, v. 42, n. 3, p. 269–276, 2018. [28](#), [39](#), [40](#), [41](#), [140](#)
- KITE, J. et al. User perceptions of the make healthy normal campaign facebook page: a mixed methods study. *Social Media+ Society*, SAGE Publications Sage UK: London, England, v. 4, n. 3, p. 1–10, 2018. [28](#), [37](#), [40](#), [41](#), [136](#)
- KORENROMP, E. L. et al. Global burden of maternal and congenital syphilis and associated adverse birth outcomes—estimates for 2016 and progress since 2012. *PloS one*, Public Library of Science San Francisco, CA USA, v. 14, n. 2, p. e0211720, 2019. [96](#)
- KUMAR, A.; ABIRAMI, S. Aspect-based opinion ranking framework for product reviews using a spearman’s rank correlation coefficient method. *Information Sciences*, Elsevier, v. 460, p. 23–41, 2018. [105](#)

- LAMPOS, V. et al. Assessing the impact of a health intervention via user-generated internet content. *Data Mining and Knowledge Discovery*, Springer, v. 29, n. 5, p. 1434–1457, 2015. 28, 37, 40, 42, 121, 122, 135
- LING, R.; LEE, J. Disease monitoring and health campaign evaluation using google search activities for hiv and aids, stroke, colorectal cancer, and marijuana use in canada: a retrospective observational study. *JMIR public health and surveillance*, JMIR Publications Inc., Toronto, Canada, v. 2, n. 2, p. e156, 2016. 42
- LOPES¹, I. et al. O uso de recursos educacionais abertos como uma alternativa para materiais autoinstrucionais na educação a distância. 2018. 77
- LUCAS, M. C. V. et al. The experience of institutional support in the project of quick response to syphilis on health care networks. *Revista Brasileira de Inovação Tecnológica em Saúde-ISSN: 2236-1103*, p. 17–17, 2019. 74
- LUO, N.; VIDAL, Y.; ACHO, L. *Advances in Industrial Control Control*. [S.l.: s.n.], 2014. ISSN 1430-9491. ISBN 978-1-4471-6346-6. 30
- MACHADO, M. B. d. C.; NOSCHANG, F. A. Diálogos entre a publicidade e a propaganda política: anúncios de oportunidade, eleitoral ou para promover a cidadania? *Encontro Nacional de Pesquisadores em Publicidade e Propaganda (10.: 2019 maio 22-24: São Paulo)*. E-book. São Paulo: USP/ECA, 2020., 2020. 108
- MANIKAS, K.; HANSEN, K. M. Software ecosystems – a systematic literature review. *Journal of Systems and Software*, Elsevier, v. 86, n. 5, p. 1294–1306, 2013. ISSN 0164-1212. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S016412121200338X>>. 53
- MASSUDA, A. et al. The brazilian health system at crossroads: progress, crisis and resilience. *BMJ global health*, BMJ Specialist Journals, v. 3, n. 4, p. e000829, 2018. 97
- MCCOMBS, M. Agenda setting, media effects on. In: WRIGHT, J. D. (Ed.). *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)*. Second edition. Oxford: Elsevier, 2015. p. 351–356. ISBN 978-0-08-097087-5. Disponível em: <<https://www.sciencedirect.com/science/article/pii/B9780080970868950074>>. 62, 100, 108
- MCKINNEY, W. et al. pandas: a foundational python library for data analysis and statistics. *Python for High Performance and Scientific Computing*, Seattle, v. 14, n. 9, p. 1–9, 2011. 53
- MESSERSCHMITT, D. G.; SZYPERSKI, C. *Software ecosystem: understanding an indispensable technology and industry*. [S.l.]: MIT press, 2003. 53
- MILONE, J. de C. Mccombs, maxwell. a teoria da agenda: a mídia e a opinião pública. petrópolis, rj: Vozes, 2009. *Revista Opinião Filosófica*, v. 3, n. 2, 2012. 30
- MOHER, D. et al. Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. *Annals of internal medicine*, American College of Physicians, v. 151, n. 4, p. 264–269, 2009. 32

- MOHER, D. et al. Preferred reporting items for systematic review and meta-analysis protocols (prisma-p) 2015 statement. *Systematic reviews*, BioMed Central, v. 4, n. 1, p. 1–9, 2015. 32
- MOLINE, H. R.; JR, J. F. S. The continuing threat of syphilis in pregnancy. *Current Opinion in Obstetrics and Gynecology*, LWW, v. 28, n. 2, p. 101–104, 2016. 96
- MONGEON, P.; PAUL-HUS, A. The journal coverage of web of science and scopus: a comparative analysis. *Scientometrics*, Springer, v. 106, n. 1, p. 213–228, 2016. 33
- NADARZYNSKI, T. et al. Targeted advertisement of chlamydia screening on social media: A mixed-methods analysis. *Digital health*, SAGE Publications Sage UK: London, England, v. 5, p. 1–10, 2019. 17, 28, 37, 40, 41, 136
- NEVES, R. G. et al. Tendência da disponibilidade de vacinas no brasil: Pmaq-ab 2012, 2014 e 2018. *Cadernos de Saúde Pública*, SciELO Public Health, v. 38, p. PT135621, 2022. 116
- NEWMAN, L. et al. Global estimates of syphilis in pregnancy and associated adverse outcomes: analysis of multinational antenatal surveillance data. *PLoS medicine*, Public Library of Science San Francisco, USA, v. 10, n. 2, p. e1001396, 2013. 96
- NUTI, S. V. et al. The use of google trends in health care research: a systematic review. *PloS one*, Public Library of Science, v. 9, n. 10, p. e109583, 2014. 73
- OLIVEIRA, A. P. C. d. et al. Challenges for ensuring availability and accessibility to health care services under brazil’s unified health system (sus). *Ciencia & saude coletiva*, SciELO Brasil, v. 22, p. 1165–1180, 2017. 97
- ORGANIZATION, W. H. *WHO | Shortages of benzathine penicillin. How big is the problem? And why it matters*. 2017. <<https://www.who.int/reproductivehealth/shortages-benzathine-penicillin/en/>>. (Accessed on 11/05/2020). 80
- OROZCO, G.; GONZÁLEZ, R. *Una coartada metodológica: abordajes cualitativos en la investigación en comunicación, medios y audiencias*. [S.l.]: Tintable, 2012. 102
- OSANLOO, A.; GRANT, C. Understanding, selecting, and integrating a theoretical framework in dissertation research: Creating the blueprint for your “house”. *Administrative issues journal: connecting education, practice, and research*, v. 4, n. 2, p. 7, 2016. 44
- ØVRETVEIT, J.; GUSTAFSON, D. Using research to inform quality programmes. *Bmj*, British Medical Journal Publishing Group, v. 326, n. 7392, p. 759–761, 2003. 18
- O’HARA, B. J. et al. Impact of the swap it, don’t stop it australian national mass media campaign on promoting small changes to lifestyle behaviors. *Journal of health communication*, Taylor & Francis, v. 21, n. 12, p. 1276–1285, 2016. 28, 39, 40, 41, 140
- PALACHY, S. *Stationarity in time series analysis*. [S.l.]: Towards Data Science, 2019. <<https://towardsdatascience.com/stationarity-in-time-series-analysis-90c94f27322>>. (Accessed on 03/04/2021). 30
- PALIT, A. K.; POPOVIC, D. *Computational intelligence in time series forecasting: theory and engineering applications*. [S.l.]: Springer Science & Business Media, 2006. 29, 74

- PEARSON, K. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, Taylor & Francis, v. 2, n. 11, p. 559–572, 1901. [60](#), [61](#)
- PESSONI, A.; CARMO, C. *Releases sobre saúde nas assessorias de imprensa das administrações públicas do ABC: produção e tendências*. Rio de Janeiro: Brazilian Congress of Communication Sciences, 2014. 37 p. [108](#)
- PETERSEN, K. et al. Systematic mapping studies in software engineering. In: *EASE*. [S.l.: s.n.], 2008. v. 8, p. 68–77. [32](#)
- PINTO, C. S. et al. Clinical, epidemiological aspects, and trends of hepatitis b in brazil from 2007 to 2018. *Scientific reports*, Nature Publishing Group, v. 11, n. 1, p. 1–10, 2021. [116](#)
- PINTO, R. et al. Assessing the impact of public health campaigns through epidemiological, communication and education indicators. In: IEEE. *2020 IEEE 20th International Conference on Advanced Learning Technologies (ICALT)*. [S.l.], 2020. p. 147–149. [44](#), [119](#)
- PINTO, R. et al. Analyzing the reach of public health campaigns based on multidimensional aspects: the case of the syphilis epidemic in brazil. *BMC Public Health*, Springer, 2021. [44](#), [71](#), [97](#), [108](#), [119](#)
- PINTO, R. et al. Use of interrupted time series analysis in understanding the course of the congenital syphilis epidemic in brazil. *The Lancet Regional Health-Americas*, Elsevier, v. 7, p. 100163, 2022. [44](#), [64](#), [85](#), [108](#), [119](#)
- PINTO, R. d. M. et al. Systematic review on information technology approaches to evaluate the impact of public health campaigns: Real cases and possible directions. *Frontiers in Public Health*, Frontiers, p. 2296, 2021. [19](#), [31](#), [119](#)
- PITT, A. L.; GOLDBERGER-FIEBERT, J. D.; BRANDEAU, M. L. Public health interventions with harms and benefits: a graphical framework for evaluating tradeoffs. *Medical Decision Making*, SAGE Publications Sage CA: Los Angeles, CA, v. 40, n. 8, p. 978–989, 2020. [120](#), [122](#)
- PLATT, T. et al. Facebook advertising across an engagement spectrum: a case example for public health communication. *JMIR public health and surveillance*, JMIR Publications Inc., Toronto, Canada, v. 2, n. 1, p. e27, 2016. [28](#), [39](#), [40](#), [41](#), [139](#)
- QING, S.; YING, Z.; PENGZHOU, Z. Research review on key techniques of topic-based news elements extraction. In: IEEE. *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*. [S.l.], 2017. p. 585–590. [101](#)
- RAC, M. W.; REVELL, P. A.; EPPES, C. S. Syphilis during pregnancy: a preventable threat to maternal-fetal health. *American journal of obstetrics and gynecology*, Elsevier, v. 216, n. 4, p. 352–363, 2017. [96](#)
- ROCHA, R. et al. Effect of socioeconomic inequalities and vulnerabilities on health-system preparedness and response to covid-19 in brazil: a comprehensive analysis. *The Lancet Global Health*, Elsevier, v. 9, n. 6, p. e782–e792, 2021. [97](#)
- SALTON, G.; BUCKLEY, C. Term-weighting approaches in automatic text retrieval. *Information processing & management*, Elsevier, v. 24, n. 5, p. 513–523, 1988. [29](#), [103](#)

- SANTOS, M. Marques dos et al. Trends of syphilis in brazil: a growth portrait of the treponemic epidemic. *Plos one*, Public Library of Science San Francisco, CA USA, v. 15, n. 4, p. e0231029, 2020. 96
- SAÚDE., M. da. Boletim epidemiológico - sífilis. *Brasília (DF)*, Ministério da Saúde - Brasília, n. Especial, p. 1–44, 2019. 107
- SAÚDE., M. da. *Indicadores e Dados Básicos da Sífilis nos Municípios Brasileiros*. [S.l.]: Ministério da Saúde - Brasília, 2020. <<http://indicadoressifilis.aids.gov.br/>>. (Accessed on 10/29/2020). 70
- SAÚDE., M. da. Resposta rápida á sífilis: uma análise situacional da estratégia de apoio institucional. *Brasília (DF)*, Ministério da Saúde - Brasília, n. 51, p. 17–21, 2020. 68
- SCARPONI, C. F. O. et al. Low eligibility for hepatitis b treatment in the brazilian public health system. *Revista da Sociedade Brasileira de Medicina Tropical*, SciELO Brasil, v. 55, 2022. 114
- SEABOLD, S.; PERKTOLD, J. Statsmodels: Econometric and statistical modeling with python. In: AUSTIN, TX. *Proceedings of the 9th Python in Science Conference*. [S.l.], 2010. v. 57, p. 61. 63
- SHARMA, N.; NAHAK, F. M. Public health communication by health workers during covid-19: A study of jalandhar district of punjab. *Journal of Contemporary Issues in Business and Government Vol*, v. 27, n. 1, 2021. 108
- SITNIK, R. et al. Study protocol: epidemiological and clinical characteristics of acute viral hepatitis in brazilian health services. *BMJ open*, British Medical Journal Publishing Group, v. 11, n. 7, p. e045852, 2021. 116
- SMITH, P. G.; MORROW, R. H.; ROSS, D. A. Field trials of health interventions. Oxford University Press, 2015. 17
- SOHN, H. et al. Costing the implementation of public health interventions in resource-limited settings: a conceptual framework. *Implementation Science*, Springer, v. 15, n. 1, p. 1–8, 2020. 121, 122
- S.PAULO, F. de. *Falta de penicilina afeta pacientes no mundo - 08/05/2017 - Equilíbrio e Saúde - Folha de S.Paulo*. 2017. <<https://www1.folha.uol.com.br/equilibrioesaude/2017/05/1881791-falta-de-penicilina-afeta-pacientes-no-mundo.shtml>>. (Accessed on 11/05/2020). 80
- STRAUSS, E. *Chronic hepatitis C is still a problem for the public health care system in brazil*. [S.l.]: SciELO Brasil, 2018. 321–323 p. 114
- SYRED, J. et al. Would you tell everyone this? facebook conversations as health promotion interventions. *Journal of medical Internet research*, JMIR Publications Inc., Toronto, Canada, v. 16, n. 4, p. e108, 2014. 28, 38, 40, 41, 138
- TALJAARD, M. et al. The use of segmented regression in analysing interrupted time series studies: an example in pre-hospital ambulance care. *Implementation Science*, Springer, v. 9, n. 1, p. 1–4, 2014. 31

- TAYEB, H. F.; KARABATAK, M.; VAROL, C. Time series database preprocessing for data mining using python. In: IEEE. *2020 8th International Symposium on Digital Forensics and Security (ISDFS)*. [S.l.], 2020. p. 1–4. 74
- TRANSMISSÍVEIS, D. de Doenças de Condições Crônicas e I. S. *Protocolo Clínico e Diretrizes Terapêuticas para Atenção Integral às Pessoas com Infecções Sexualmente Transmissíveis (IST)*. [S.l.]: Ministério da Saúde - Brasília, 2015. [Http://www.aids.gov.br/pt-br/pub/2015/protocolo-clinico-e-diretrizes-terapeuticas-para-atencao-integral-pessoas-com-infecoes](http://www.aids.gov.br/pt-br/pub/2015/protocolo-clinico-e-diretrizes-terapeuticas-para-atencao-integral-pessoas-com-infecoes). (Accessed on 14/07/2021). 96
- TRANSMISSÍVEIS, D. de Doenças de Condições Crônicas e I. S. *Strategic actions agenda for reducing syphilis in Brazil*. Brasília - DF: Ministério da Saúde - Brasília, 2017. 34 p. [Http://www.aids.gov.br/pt-br/pub/2016/agenda-de-acoes-estrategicas-para-reducao-da-sifilis-congenita](http://www.aids.gov.br/pt-br/pub/2016/agenda-de-acoes-estrategicas-para-reducao-da-sifilis-congenita). 67, 97, 98, 107
- TURNER, S. L. et al. Design characteristics and statistical methods used in interrupted time series studies evaluating public health interventions: a review. *Journal of clinical epidemiology*, Elsevier, v. 122, p. 1–11, 2020. 31
- VALENTIM, R. A. et al. Stochastic petri net model describing the relationship between reported maternal and congenital syphilis cases in brazil. *BMC medical informatics and decision making*, BioMed Central, v. 22, n. 1, p. 1–12, 2022. 108
- VALENTIM, R. A. d. M. et al. Educommunication as a strategy to face syphilis: an analysis of the open educational resources available at avasus. *DST j. bras. doenças sex. transm*, p. 1–5, 2021. 20, 68
- VARGIU, E.; URRU, M. Exploiting web scraping in a collaborative filtering-based approach to web advertising. *Artif. Intell. Research*, v. 2, n. 1, p. 44–54, 2013. 47
- WAGNER, A. K. et al. Segmented regression analysis of interrupted time series studies in medication use research. *Journal of clinical pharmacy and therapeutics*, Wiley Online Library, v. 27, n. 4, p. 299–309, 2002. 30, 31, 63, 86
- WAKEFIELD, M. A. et al. Impact of tobacco control policies and mass media campaigns on monthly adult smoking prevalence. *American journal of public health*, American Public Health Association, v. 98, n. 8, p. 1443–1450, 2008. 17, 28, 37, 40, 41, 136
- WAKEFIELD, M. A.; LOKEN, B.; HORNIK, R. C. Use of mass media campaigns to change health behaviour. *The Lancet*, Elsevier, v. 376, n. 9748, p. 1261–1271, 2010. 17
- WIERINGA, R. et al. Requirements engineering paper classification and evaluation criteria: a proposal and a discussion. *Requirements engineering*, Springer, v. 11, n. 1, p. 102–107, 2006. 34
- WOHLIN, C. et al. On the reliability of mapping studies in software engineering. *Journal of Systems and Software*, Elsevier, v. 86, n. 10, p. 2594–2610, 2013. 34
- WU, Z. et al. On the trend, detrending, and variability of nonlinear and nonstationary time series. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 104, n. 38, p. 14889–14894, 2007. 31
- ZHAN, Q. et al. Inferring social influence of anti-tobacco mass media campaign. *IEEE transactions on nanobioscience*, IEEE, v. 16, n. 5, p. 356–366, 2017. 28, 37, 40, 41, 135

ÉPOCA. *Por que o Brasil não tem penicilina - ÉPOCA | Check-Up*. 2017. <<https://epoca.globo.com/saude/check-up/noticia/2017/06/por-que-o-brasil-nao-tem-penicilina.html>>. (Accessed on 11/05/2020). 80

Apêndices

APÊNDICE A – Catálogo com dados de campanhas e propriedades extras

Neste apêndice, apresentamos um catálogo com a identificação dos estudos primários de nossa pesquisa.

[S1] Assessing the impact of a health intervention via user-generated Internet content ([LAMPOS et al., 2015](#))

- Nome da campanha: Live Attenuated Influenza Vaccine
- Tipo de anúncio: não informado
- Área de Tópico: Vacinação Infantil
- Local da campanha do país: Inglaterra
- Público-alvo: Público-alvo
- Período: temporada de gripe 2013/2014
- Nível de Organização: Nacional
- Lançado por: Não reportado
- Fontes de Dados: Bing da Microsoft; Dados do Twitter

[S2] Impact of Facebook ads for sexual health promotion via an educational web app: A case study ([GABARRON et al., 2019](#))

- Nome da campanha: Não informado
- Tipo de Anúncios: Campanha de Anúncios do Facebook; Cartazes
- Área de Tópico: Cuidados de Saúde Sexual
- Local da campanha do país: Noruega
- Público-alvo: Público-alvo
- Período: 4 de abril a 3 de maio de 2013
- Nível de Organização: Local
- Lançado por: Hospital Universitário do Norte da Noruega
- Fontes de dados: Dados do Facebook

[S3] Inferring Social Influence of Anti-Tobacco Mass Media Campaign ([ZHAN et al., 2017](#))

- Nome da Campanha: a) Tips from Former Smokers 2013; b) “Legacy Truth”
- Tipo de anúncio: anúncios de televisão
- Área de Tópico: Antifumo
- Local da campanha do país: Estados Unidos

- Público-alvo: Público-alvo
- Período: a) 1º de março a 23 de junho de 2013; b) 11 de agosto a 28 de outubro de 2013
- Nível de Organização: Nacional
- Lançado por: a) Centros de Controle de Doenças; b) Fundação do Legado Americano
- Fontes de dados: Dados de anúncios de TV; Dados do Twitter

[S4] Targeted advertisement of chlamydia screening on social media: A mixed-methods analysis ([NADARZYNSKI et al., 2019](#))

- Nome da campanha: Lets Talk About It (STI)
- Tipo de Anúncios: a) Cartazes; b) Campanha de anúncios do Facebook
- Área de Tópico: Clamídia - DST
- Local da campanha do país: Inglaterra
- Público-alvo: Público-alvo
- Período: a) Jan e Mar 2017; b) 4 de agosto a 11 de outubro de 2017
- Nível de Organização: Local
- Lançado por: a) University of Southampton b) Solent NHS Trust Governança Clínica
- Fontes de Dados: Entrevistas semiframeworkdas; Facebook

[S5] User Perceptions of the Make Healthy Normal campaign Facebook Page: A Mixed Methods Study ([KITE et al., 2018b](#))

- Nome da campanha: Make Healthy Normal (MHN)
- Tipo de anúncio: comerciais de TV apoiados por eventos da comunidade, imprensa, anúncios fora de casa, online, relações públicas, site e mídia social.
- Área de Tópico: Sobrepeso e Obesidade
- Local da campanha do país: Austrália
- Público Alvo: a) Público em geral; b) Público-alvo
- Período: a) 2015-2016; b) 2016-2018
- Nível de Organização: Estado
- Lançado por: Governo Australiano
- Fontes de dados: Pesquisa online com participantes recrutados via Facebook Ads e postagens na página do MHN; Dados do Facebook

[S6] Impact of tobacco control policies and mass media campaigns on monthly adult smoking prevalence ([WAKEFIELD et al., 2008](#))

- Nome da campanha: Campanha Nacional do Tabaco na Austrália (entre 1996 a 2001) e programas estaduais de controle do tabaco (desde 2002)
- Tipo de anúncio: comerciais de televisão
- Área de Tópico: Antifumo

- Local da campanha do país: Austrália
- Público-alvo: Público-alvo
- Período: De junho de 1995 a dezembro de 2006
- Nível de Organização: Nacional; Estado
- Lançado por: Governo Australiano
- Fontes de dados: Pesquisa, dados de anúncios de TV, custo do cigarro, vendas mensais de terapia de reposição de nicotina e bupropiona e leis antifumo.

[S7] Adolescent reactance and anti-smoking campaigns: A theoretical approach (GRANDPRE et al., 2003)

- Nome da campanha: Não informado
- Tipo de anúncio: Mensagens que utilizam programas de computador multimídia interativos para mesclar fotografia, música, vídeo, locução de áudio e texto
- Área de Tópico: Antifumo
- Local da campanha do país: Estados Unidos
- Público-alvo: Público-alvo
- Período: Não informado
- Nível de Organização: Local
- Lançado por: Não reportado
- Fontes de dados: pesquisa online com alunos

[S8] Social media-delivered sexual health intervention: A cluster randomized controlled trial (BULL et al., 2012)

- Nome da campanha: Just/Us (página da intervenção)
- Tipo de anúncio: campanha de anúncios do Facebook
- Área de Tópico: Cuidados de Saúde Sexual - DST
- Local da campanha do país: Estados Unidos
- Público-alvo: Público-alvo
- Período: outubro de 2010 e maio de 2011
- Nível de Organização: Local
- Lançado por: Não reportado
- Fontes de Dados: Pesquisa; Dados do Facebook

[S9] Are You Scared Yet? Evaluating Fear Appeal Messages in Tweets About the Tips Campaign (EMERY et al., 2014)

- Nome da campanha: Tips From Former Smokers
- Tipo de anúncio: outdoors, anúncios de TV, rádio, impressos & websites
- Área de Tópico: Antifumo
- Local da campanha do país: Estados Unidos

- Público-alvo: Público-alvo
- Período: 15 de março de 2012 a 9 de junho de 2012
- Nível de Organização: Nacional
- Lançado por: Centros de Controle de Doenças
- Fontes de dados: dados do Twitter

[S10] Campaigns and counter campaigns: Reactions on Twitter to e-cigarette education ([ALLEM et al., 2017](#))

- Nome da campanha: Still Blowing Smoke
- Tipo de anúncio: anúncios de TV, outdoors e um site
- Área de Tópico: Antifumo
- Local da campanha do país: Estados Unidos
- Público-alvo: N/A
- Período: 23 de março de 2015 a junho de 2015
- Nível de Organização: Estado
- Lançado por: Departamento de Saúde Pública da Califórnia
- Fontes de dados: dados do Twitter

[S11] Would you tell everyone this? Facebook conversations as health promotion interventions ([SYRED et al., 2014](#))

- Nome da campanha: Say Yes to the Test
- Tipo de anúncio: campanha publicitária do Facebook
- Área de Tópico: Clamídia - DST
- Local da campanha do país: Inglaterra
- Público-alvo: Público-alvo
- Período: janeiro e junho de 2010
- Nível de Organização: Nacional
- Lançado por: English Department of Health and the Health Protection Agency (agora Public Health England)
- Fontes de dados: Dados do Facebook

[S12] IknowUshould2: Feasibility of a Youth-Driven Social Media Campaign to Promote STI and HIV Testing Among Adolescents in Philadelphia ([DOWSHEN et al., 2015](#))

- Nome da campanha: IknowUshould2
- Tipo de anúncio: anúncios impressos, camisetas, rádio, site, Facebook, Twitter, Instagram, YouTube, eventos de campanha e divulgação da comunidade & parceria.
- Área de Tópico: Teste de DST e HIV
- Local da campanha do país: Estados Unidos

- Público-alvo: Público-alvo
- Período: setembro de 2012 - agosto de 2013
- Nível de Organização: Local
- Lançado por: Hospital Infantil da Filadélfia
- Fontes de dados: Facebook, Twitter, Instagram, página da campanha do site

[S13] Attitudes to chlamydia screening elicited using the social networking site Facebook for subject recruitment (AHMED et al., 2013)

- Nome da campanha: (componente da) Young Female Health Initiative
- Tipo de anúncio: campanha de anúncios do Facebook
- Área de Tópico: Clamídia - DST
- Local da campanha do país: Austrália
- Público-alvo: Público-alvo
- Período: maio e setembro de 2010
- Nível de Organização: Estado
- Lançado por: Young Female Health Initiative
- Fontes de dados: Dados do Facebook, dados do Censo de 2006 do Australian Bureau of Statistics; Índices Socioeconômicos por Áreas (SEIFA).

[S14] Facebook advertising across an engagement spectrum: A case example for public health communication (PLATT et al., 2016)

- Nome da campanha: Michigan BioTrust for Health (BioTrust)
- Tipo de anúncio: campanha publicitária do Facebook
- Área de Tópico: A triagem neonatal e os programas de biobanco de grande população
- Local da campanha do país: Estados Unidos
- Público-alvo: Público-alvo
- Período: Primavera de 2015, 11 semanas
- Nível de Organização: Estado
- Lançado por: Programa de Ciências da Vida e Sociedade da Universidade de Michigan
- Fontes de dados: Dados do Facebook

[S15] Advertising sexual health services that provide sexually transmissible infection screening for rural young people - What works and what doesn't (GAMAGE et al., 2011)

- Nome da campanha: programa TESTme
- Tipo de anúncio: site, página no Facebook, cartazes, panfletos, cartões de visita, pulseiras e programas regionais de conscientização sobre saúde. Jornais, revistas e boletins para jovens, anúncios em diários de estudantes em universidades e anúncios de mensagens curtas para telefones celulares.

- Área de Tópico: Cuidados de Saúde Sexual - DST
- Local da campanha do país: Austrália
- Público-alvo: Público-alvo
- Período: agosto de 2009 a junho de 2010
- Nível de Organização: Estado
- Lançado por: Melbourne Sexual Health Centre (MSHC)
- Fontes de dados: Pesquisa. Clientes que entraram em contato com o serviço após a exposição dos anúncios

[S16] Impact of the Swap It, Don't Stop It Australian National Mass Media Campaign on Promoting Small Changes to Lifestyle Behaviors (O'HARA et al., 2016)

- Nome da campanha: Swap It, Don't Stop It
- Tipo de anúncio: anúncios de TV, rádio, revistas, configurações on-line e fora de casa
- Área de Tópico: Sobrepeso e Obesidade
- Local da campanha do país: Austrália
- Público-alvo: Público-alvo
- Período: março a junho de 2011 e setembro a dezembro de 2011
- Nível de Organização: Nacional
- Lançado por: Governo Australiano
- Fontes de dados: Pesquisas telefônicas.

[S17] Impact of the make healthy normal mass media campaign (Phase 1) on knowledge, attitudes and behaviours: A cohort study (KITE et al., 2018a) (KITE et al., 2018a)

- Nome da campanha: Make Healthy Normal (MHN)
- Tipo de anúncio: comerciais de TV apoiados por eventos comunitários, imprensa, fora de casa (por exemplo, outdoors), anúncios on-line, relações públicas, site e mídia social.
- Área de Tópico: Sobrepeso e Obesidade
- Local da campanha do país: Austrália
- Público-alvo: Público-alvo
- Período: Fase 1) junho de 2015 a junho de 2016;
- Nível de Organização: Estado
- Lançado por: Governo Australiano
- Fontes de dados: pesquisas online com participantes recrutados por meio de um painel de pesquisa

APÊNDICE B – Intervenções do Projeto “Sífilis Não!” e linha do tempo

Neste apêndice, apresentamos as Intervenções do Projeto “Sífilis Não!” e sua linha do tempo.

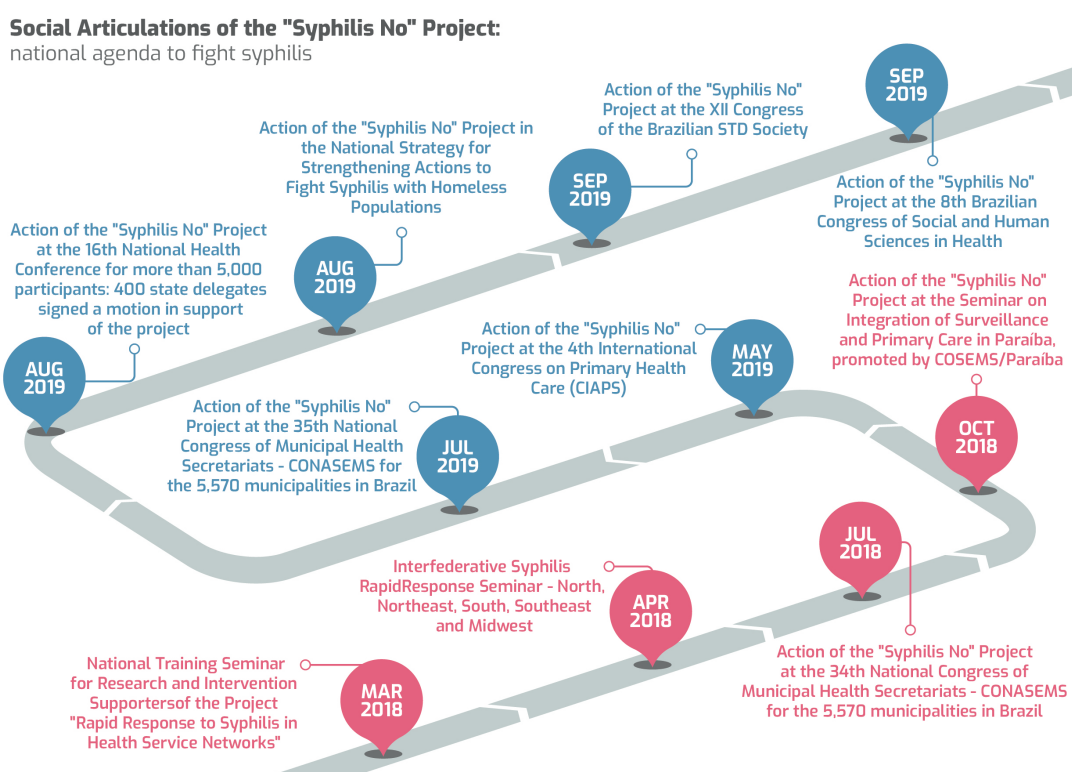


Figura 36 – Articulações Sociais do Projeto “Sífilis Não!” realizado entre março de 2018 e setembro de 2019.

Awareness Campaign on Facebook From February to November 2018



Figura 37 – Indicadores da primeira campanha de conscientização do Projeto “Sífilis Não!” no Facebook de fevereiro a novembro de 2018.

Awareness Campaign: #2

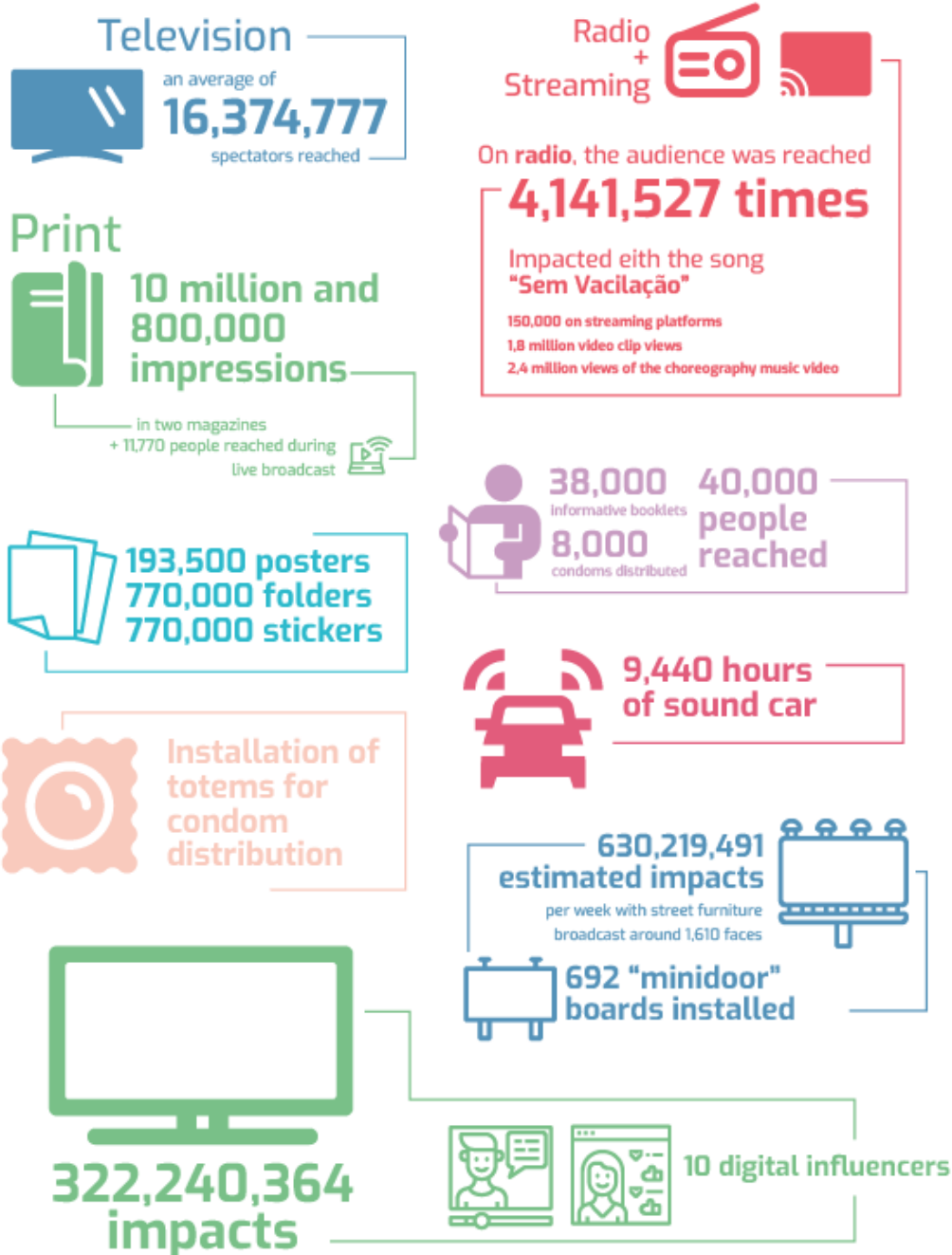


Figura 38 – Indicadores da segunda campanha de conscientização do Projeto “Sífilis Não!”.

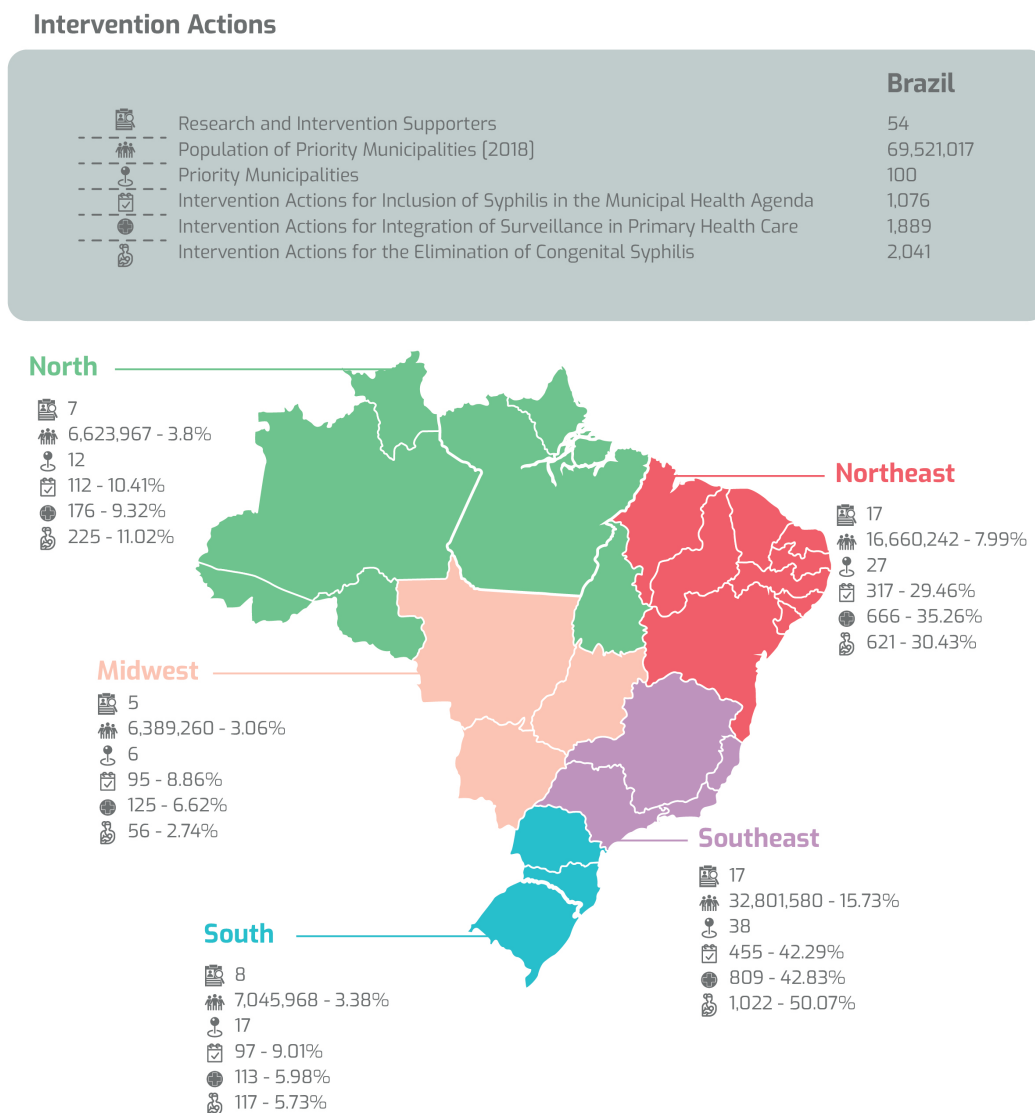


Figura 39 – Ações de intervenção realizadas por Apoiadores de Pesquisa e Intervenção, que realizaram trabalhos específicos com gestores locais de saúde.

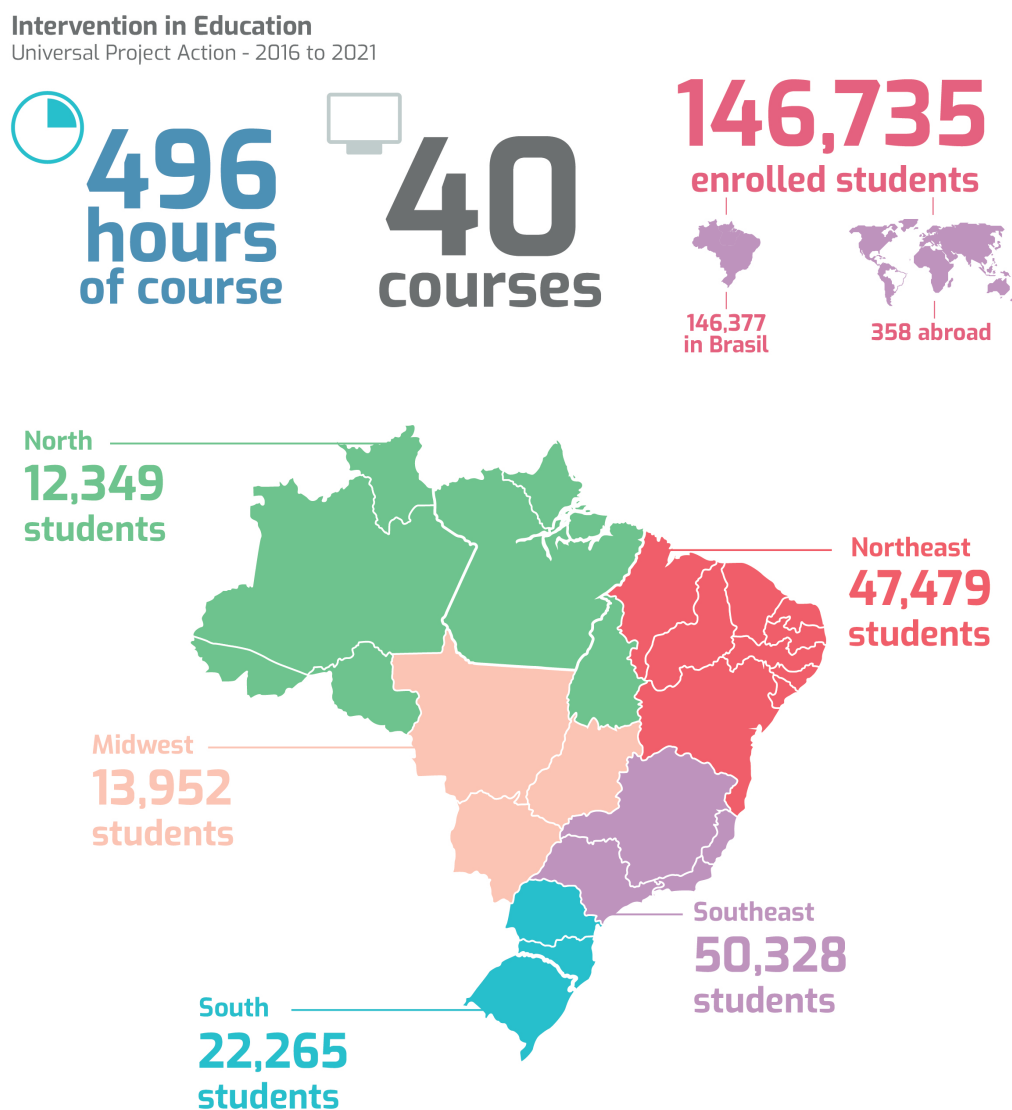


Figura 40 – Indicadores da trilha de treinamento “Sífilis e outras ISTs” que oferece 40 cursos em formato autoinstrucional na plataforma AVASUS.