

Metadados da tese de Vinícius de Araújo Oliveira

Título

Título em português:

Vigilância digital da covid-19: predizer, comparar e desvendar – relato de três soluções em ciência de dados em saúde desenvolvidas no enfrentamento da pandemia

Title in english:

Digital Health Surveillance on COVID-19: predict, assess and unravel – multiple case study of three data science solutions developed in response to the pandemic in Brazil.

Resumo e Abstract

Resumo:

VIGILÂNCIA DIGITAL DA COVID-19: PREDIZER, COMPARAR E DESVENDAR – RELATO DE TRÊS SOLUÇÕES EM CIÊNCIA DE DADOS EM SAÚDE DESENVOLVIDAS NO ENFRENTAMENTO DA PANDEMIA. **Introdução:** A pandemia de covid-19 evidenciou a necessidade crítica de dados em tempo real para orientar a saúde pública. Este trabalho investiga a "vigilância digital em saúde", que combina a vigilância em saúde com o uso de tecnologias digitais, como categoria de análise na pandemia covid-19 para aprimorar a capacidade de prever, investigar e responder a emergências sanitárias. **Objetivo central:** analisar e desenvolver um conjunto de tecnologias e processos, utilizando dados do mundo real, gerados pelos serviços de saúde, para apoiar a resposta à pandemia de covid-19 em suas diversas fases. As frentes de pesquisa buscaram prever surtos, avaliar a efetividade de intervenções como vacinas e desvendar aspectos clínicos da doença, como a covid longa. **Metodologia:** a pesquisa utilizou dados secundários de diversos sistemas de informação do Sistema Único de Saúde (SUS). Foram analisados quatro projetos: **Vigicovid**, que explorou prontuários eletrônicos; **VigiVac**, que avaliou a efetividade das vacinas com dados da Rede Nacional de Dados em Saúde (RNDS); **Aesop**, que desenvolveu um sistema de alerta precoce com dados da Atenção Primária à Saúde (SISAB); e um estudo sobre **covid longa** com mineração de texto de prontuários hospitalares. As estratégias de proteção de dados variaram entre pseudonimização na fonte, anonimização supervisionada e anonimização pelo controlador. As análises envolveram estudos de coorte, desenhos caso-controle com teste negativo e técnicas de mineração de texto. **Resultados:** Foi demonstrado o potencial dos dados estudados para prever surtos de vírus respiratórios a partir de dados da atenção primária; desenvolvida uma coorte digital em tempo real que confirmou a alta efetividade das vacinas contra a covid-19; especialmente com doses de reforço e em grupos específicos, como gestantes e indígenas e desenvolver um método capaz de "desvendar" prontuários eletrônicos e anotações em texto livre. **Conclusões:** soluções efetivas para vigilância digital em saúde foram desenvolvidas durante a pandemia nesta pesquisa, indicando que o Brasil estará mais preparado para futuras pandemias. Contudo, persistem desafios significativos relacionados à padronização, acesso e governança de dados em saúde, cuja superação é fundamental para avançar no uso de DMR para pesquisa e vigilância em saúde no SUS.

Abstract:

DIGITAL SURVEILLANCE OF COVID-19: PREDICTING, COMPARING, AND UNVEILING – A REPORT ON THREE DATA SCIENCE SOLUTIONS IN HEALTH DEVELOPED IN RESPONSE TO THE PANDEMIC. Introduction: The COVID-19 pandemic highlighted the critical need for real-time data to guide public health responses. This thesis investigates "digital health surveillance," which combines health surveillance with the use of digital technologies, as a framework for solutions in the COVID-19 pandemic to enhance the ability to predict, investigate, and respond to health emergencies. Objective: To analyze and develop a set of technologies and processes using real-world data generated by health services to support the response to the COVID-19 pandemic across its various phases. The research fronts aimed to predict outbreaks, evaluate the effectiveness of interventions such as vaccines, and unravel clinical aspects of the disease, such as long COVID. Methodology: The research utilized secondary data from various information systems of the Brazilian Unified Health System (SUS). Four projects were analyzed: Vigicovid, which explored electronic health records; VigiVac, which assessed vaccine effectiveness using data from the National Health Data Network (RNDS); Aesop, which developed an early warning system with data from Primary Health Care (SISAB); and a study on long COVID employing text mining of hospital records. Data protection strategies varied, including source pseudonymization, supervised anonymization, and controller-led anonymization. The analyses involved cohort studies, test-negative design case-control studies, and text mining techniques. Results: The potential of the studied data and methods to predict respiratory virus outbreaks from primary care data was demonstrated. A real-time digital cohort was developed that confirmed the high effectiveness of COVID-19 vaccines, particularly with booster doses and in specific groups such as pregnant women and indigenous populations. Furthermore, a method capable of "unveiling" electronic health records and free-text annotations was developed. Conclusions: Effective solutions for digital health surveillance were developed during the pandemic in this research, indicating that Brazil will be better prepared for future pandemics. However, significant challenges related to the standardization, access, and governance of health data persist. Overcoming these challenges is fundamental to advancing the use of real-world data for research and health surveillance within the SUS.

Palavras-chave e Keywords

Palavras-chave DeCS:

COVID-19, Vigilância em Saúde Pública, Ciência de Dados, Pandemias, Registros Eletrônicos de Saúde, Vigilância Sindrômica, Mineração de Texto, Vacinas contra a COVID-19, Eficácia de Vacinas, Síndrome Pós-COVID-19 Aguda, Brasil

MeSH Keywords:

COVID-19; Public Health Surveillance; Data Science; Pandemics; Electronic Health Records; Syndromic Surveillance; Text Mining; COVID-19 Vaccines; Vaccine Efficacy; Post-Acute COVID-19 Syndrome; Brazil

Tabelas com principais resultados

O principal resultado da Tese é um conjunto da produção técnica resultante da implantação dos ambientes e métodos de ciência de dados em saúde implantados neste doutoramento. O Quadro 1 apresenta os métodos utilizados em cada projeto analisado:

Quadro 1 - Sistemas de informação, formas de acesso e proteção de dados pessoais

Projeto	Sistemas de Informação	Tipos de bancos de dados	Forma de acesso	Proteção de dados pessoais
Vigicovid	AGHU	PostgreSQL	SQL	Anonimização supervisionada na fonte
VigiVac	RNDS-RIA SIVEP-Gripe e-SUS VE	RedShift	SQL	Pseudonimização na fonte, seguida de anonimização
Aesop	SISAB	Oracle e arquivos CSV	SQL e transferência de arquivos	Anonimização específica pelo controlador
Long Covid HUs	Cadastro, evolução e exames do HSP	Oracle, MongoDB, Excel	Transferência de arquivos	Anonimização supervisionada na fonte

Já o Quadro 2 apresenta a produção científica resultante do trabalho em equipe.

Quadro 2 - Produção acadêmica da pesquisa

Número / Projeto	Título	Publicação	Ano
1 Vigicovid	An ontology based natural language processing pipeline for brazilian covid-19 emr (Gritz <i>et al.</i> , 2021)	Brazilian e-Science Workshop (BreSci) https://doi.org/10.5753/bresci.2021.15794	2021
2 VigiVac	Métodos para avaliação da efetividade de vacinas para COVID-19 com ênfase em abordagens quase-experimentais (Pescarini <i>et al.</i> , 2021)	Ciência & Saúde Coletiva https://doi.org/10.1590/1413-812320212611.18622021	2021
3 VigiVac	CoronaVac vaccine is effective in preventing symptomatic and severe COVID-19 in pregnant women in Brazil: a test-negative case-control study (Paixao <i>et al.</i> , 2022)	BMC medicine https://doi.org/10.1186/s12916-022-02353-w	2022
4 VigiVac	Duration of protection of CoronaVac plus heterologous BNT162b2 booster in the Omicron period in Brazil (Cerqueira-Silva <i>et al.</i> , 2022c)	Nature communications https://doi.org/10.1038/s41467-022-31839-7	2022
5 VigiVac	Effectiveness of BNT162b2 booster after CoronaVac primary regimen in pregnant people during omicron period in Brazil (Florentino <i>et al.</i> , 2022c)	The Lancet Infectious Diseases https://doi.org/10.1016/S1473-3099(22)00728-9	2022
6 VigiVac	Effectiveness of CoronaVac, ChAdOx1 nCoV-19, BNT162b2, and Ad26. COV2. S among individuals with previous SARS-CoV-2 infection in Brazil: a test-negative, case-control study (Cerqueira-Silva <i>et al.</i> , 2022)	The Lancet Infectious Diseases https://doi.org/10.1016/S1473-3099(22)00140-2	2022
7 VigiVac	Influence of age on the effectiveness and duration of protection of Vaxzevria and CoronaVac vaccines: A population-based study (Cerqueira-Silva <i>et al.</i>, 2022a)	The Lancet Regional Health–Americas https://doi.org/10.1016/j.lana.2021.100154	2022
8 VigiVac	Two-dose ChAdOx1 nCoV-19 vaccine protection against COVID-19 hospital admissions and deaths over time: a retrospective, population-based cohort study in Scotland and Brazil (Katikireddi <i>et al.</i>, 2022)	The Lancet https://doi.org/10.1016/S0140-6736(21)02754-9	2022
9 VigiVac	Vaccination plus previous infection: protection during the omicron wave in Brazil (Cerqueira-Silva <i>et al.</i> , 2022b)	The Lancet Infectious Diseases https://doi.org/10.1016/S1473-3099(22)00288-2	2022
10 VigiVac	Vaccine effectiveness of CoronaVac against COVID-19 among children in Brazil during the Omicron period (Florentino <i>et al.</i> , 2022a)	Nature communications https://doi.org/10.1038/s41467-022-32524-5	2022
11 VigiVac	Vaccine effectiveness of heterologous CoronaVac plus BNT162b2 in Brazil (Cerqueira-Silva <i>et al.</i>, 2022c)	Nature Medicine https://doi.org/10.1038/s41591-022-01701-w	2022
12 VigiVac	Vaccine effectiveness of two-dose BNT162b2 against symptomatic and severe COVID-19 among adolescents in Brazil and Scotland over time: a test-negative case-control study (Florentino <i>et al.</i> , 2022b)	The Lancet infectious diseases https://doi.org/10.1016/S1473-3099(22)00451-0	2022

Quadro 2 – Produção acadêmica da pesquisa (continuação) – 2023 a 2025

Número / Projeto	Título	Publicação	Ano
13 Aesop	Early detection of respiratory disease outbreaks through primary healthcare data (Cerqueira-Silva <i>et al.</i> , 2023a)	Journal of Global Health https://doi.org/10.7189/jogh.13.04124	2023
14 VigiVac	Effectiveness of mRNA boosters after homologous primary series with BNT162b2 or ChAdOx1 against symptomatic infection and severe COVID-19 in Brazil and Scotland: A test-negative design case-control study (Cerqueira-Silva <i>et al.</i> , 2023b)	PLoS medicine https://doi.org/10.1371/journal.pmed.1004156	2023
15 VigiVac	Factors associated with COVID-19 vaccination among pregnant women in Rio De Janeiro City, Brazil (Borges <i>et al.</i> , 2023)	Nature Scientific Reports https://doi.org/10.1038/s41598-023-44370-6	2023
16 VigiVac	Integrating real-world data from Brazil and Pakistan into the OMOP common data model and standardized health analytics framework to characterize COVID-19 in the Global South (Junior <i>et al.</i> , 2022)	Journal of the American Medical Informatics Association https://doi.org/10.1093/jamia/ocac180	2023
17 VigiVac	Safety of BNT162b2 and CoronaVac during pregnancy on birth outcomes and neonatal mortality: a cohort study from Brazil (Tavares Veras Florentino <i>et al.</i> , 2023)	International Journal of Epidemiology https://doi.org/10.1093/ije/dyad120	2023
18 Aesop	Syndromic detection of upper respiratory infections in primary healthcare as a candidate for COVID-19 early warning in Brazil: a retrospective ecological study (De Araújo Oliveira <i>et al.</i>, 2023)	SSRN Preprints 4364869 https://dx.doi.org/10.2139/ssrn.4364869	2023
19 VigiVac	Vaccine coverage and effectiveness against laboratory-confirmed symptomatic and severe Covid-19 in indigenous people in Brazil: a cohort study (Pescarini <i>et al.</i> , 2023)	BMC Public Health https://doi.org/10.1186/s12889-023-16196-4	2023
20 Aesop	Combining digital and molecular approaches using health and alternate data sources in a next-generation surveillance system for anticipating outbreaks of pandemic potential (Ramos <i>et al.</i> , 2024)	JMIR public health and surveillance https://doi.org/10.2196/47673	2024
21 Aesop	Sistema de alerta precoce usando dados de saúde primária na era pós-pandemia de COVID-19: estudo de caso do Brasil (Cerqueira-Silva <i>et al.</i> , 2024)	Cadernos de Saúde Pública https://doi.org/10.1590/0102-311XEN010024	2024
22 Long Covid	Text mining method to unravel long COVID's clinical condition in hospitalized (Veras Florentino <i>et al.</i>, 2024)	Nature Cell Death & Disease https://doi.org/10.1038/s41419-024-07043-4	2024
23 Aesop	Impact of Primary Health Care Data Quality on Infectious Disease Surveillance in Brazil: Case Study (Florentino <i>et al.</i> , 2025)	JMIR Public Health and Surveillance https://doi.org/10.2196/67050	2025
24 Aesop	Primary health care data-based early warning system for dengue outbreaks: a nationwide case study in Brazil (Santos-Silva <i>et al.</i> , 2025)	The Lancet Regional Health-Americas https://doi.org/10.1016/j.lana.2025.101165	2025

E o Quadro 3 apresenta os artigos que integram a tese

Quadro 3 - Artigos que integram a tese

	Título	Projeto
1	Syndromic detection of upper respiratory infections in Primary Healthcare as a candidate for Covid-19 early warning in Brazil: A retrospective ecological study	Aesop
2	Influence of age on the effectiveness and duration of protection of Vaxzevria and CoronaVac vaccines: A population-based study	VigiVac
3	Two-dose ChAdOx1 nCoV-19 vaccine protection against COVID-19 hospital admissions and deaths over time: a retrospective, population-based cohort study in Scotland and Brazil	VigiVac
4	Vaccine effectiveness of heterologous CoronaVac plus BNT162b2 in Brazil	VigiVac
5	Text mining method to unravel long COVID's clinical condition in hospitalized patients	Long-Covid HU