Spatio-Temporal Analysis of STI Notifications in Brazilian Municipalities and the Effect of the COVID-19 Pandemic

Bruna I. S. Pereira[1], **Ana Luisa Lemos Bezerra**[2], **Anderson B. S. Lavinscky**[1], **Elucir Gir**[2], **Mirela T. Cazzolato**[1]

[1] Institute of Mathematics and Computer Science - University of São Paulo (ICMC-USP)
Av. Trabalhador São-carlense, 400 - Centro, 13566-590 – São Carlos, SP – Brazil

[2] School of Nursing at Ribeirão Preto - University of São Paulo (EERP-USP)
R. Prof. Hélio Lourenço, 3900 - Vila Monte Alegre, 14040-902, Ribeirão Preto, SP – Brazil

{brunaizabel, analuisalemos}@usp.br, egir@eerp.usp.br, mirela@icmc.usp.br

Abstract. *STI (Sexually Transmitted Infection) monitoring in Brazil has been an ongoing effort since the 1980s. With the establishment of DATASUS in 1991, it became possible to track annual variations in notifications across Brazilian municipalities. However, data from 2020 and 2021 reveal a significant drop in reported cases, coinciding with the COVID-19 pandemic. This study conducts a spatio-temporal analysis of STI notification time series covering Brazilian state capitals and neighboring municipalities, investigating whether this decline occurred in a systematic and geographically distributed manner. Experimental results uncover a consistent similarity pattern across states, pointing to a structural reduction in notifications during the pandemic period.*

1. Introduction

Sexually transmitted infections (STI) are a significant public health issue worldwide. These infections are caused by a wide variety of microorganisms, including viruses, bacteria, and parasites, which can be transmitted from one person to another primarily through unprotected sexual contact. It is estimated that there are over 30 agents responsible for Sexually Transmitted Infections (STI), with some of these pathogens being associated with the highest incidence rates of these infections [Sousa et al. 2024]. According to the World Health Organization (WHO)[1], more than 1 million curable STI are acquired every day. In 2020, the WHO estimated 374 million new infections worldwide.

Concurrently, in late 2019, a pneumonia caused by the novel coronavirus SARS-CoV-2 was identified in Wuhan, China [Wang et al. 2020]), and the World Health Organization (WHO) officially named the disease ‘COVID-19’, triggered a global pandemic. This unprecedented crisis severely affected healthcare systems and epidemiological surveillance services worldwide, potentially contributing to the reduction of STI notifications observed during 2020 and 2021. In Catalonia, for example, the number of reported STI cases was 51% lower than expected, reaching an average reduction of 56% during lockdown [Sentís et al. 2021]. In Brazil, the first cases of COVID-19 were confirmed in February 2020, and many measures were implemented with the aim of containing and mitigating the spread of the disease [Cavalcante et al. 2020].

[1]WHO - Sexually transmitted infections (STIs): *https://www.who.int/news-room/fact-sheets/detail/sexually-transmitted-infections-(stis)*

The pandemic also impacted national STI notifications. The number of reported HIV cases declined by 20.2%, *i.e.*, 7,726 fewer cases when compared to 2019 and 2020 [Brazil. Ministry of Health 2023]. As HIV notifications declined, the detection rate of Acquired Syphilis also experienced a significant reduction of 23.9% in 2020 compared to 2019. The rate dropped to 59.7 cases per 100,000 inhabitants, temporarily interrupting a historical growth trend. This decline was mainly attributed to the pandemic, specifically the reduced diagnostic capacity and restricted access to healthcare services during that period [Brazil. Ministry of Health 2024].

While the overall decline in STI notifications highlights a systemic disruption in healthcare and epidemiological surveillance, it is crucial to investigate how these reporting gaps manifested within specific demographic groups. With the goal of identifying the most important STI research areas to address the global public health need, in 2022 the WHO initiated a research prioritization process whose results provide a call to action to address the global STI epidemic [Gottlieb et al. 2024]. Without research and investment, STI will continue to harm the sexual and reproductive health of people worldwide, especially people of female sex and newborns, adolescents and young people, people living in low- and middle-income countries, and marginalized populations. This WHO agenda should encourage researchers to focus on priority STI research areas.

This study focuses on the female population aged over 15 years. Because the national health databases provided by DATASUS, specifically the Information System for Notifiable Conditions (SINAN), the Mortality Information System (SIM), the Laboratory Research Control System (SISCEL), and the Medication Logistics Control System (SICLOM), record only biological sex rather than gender identity, the term "female sex" was adopted to characterize the analyzed population. Within this demographic, this analysis targets two STIs of compulsory notification in Brazil: HIV/AIDS, a viral infection that progressively compromises the immune system[2], and Acquired Syphilis, a bacterial infection caused by *Treponema pallidum* associated with high morbidity if untreated[3].

We perform a spatio-temporal analysis of STI notification time series across Brazilian states, focusing strictly on the female population. By leveraging national health databases consolidated via DATASUS, including SINAN, SIM, SISCEL, and SILCLOM, we investigate whether the pandemic-related decline occurred in a systematic and geographically distributed manner. The source code is available at GitHub[4].

2. Background and related work

Several studies show that the COVID-19 pandemic severely impacted global epidemiological surveillance, resulting in significant underreporting of STIs. In [Sentís et al. 2021] and [Yan et al. 2021], the authors observed drops of approximately 50% in expected notifications in Catalonia and China, respectively. This gap in healthcare compromises long-term disease control, as demonstrated by [Aribe-Jr. et al. 2022], who used Artificial Neural Networks to forecast the deficit in the HIV response in the Philippines. The authors warn that pandemic-related disruptions threaten the goal of ending the epidemic by 2030, projecting over 145,000 cumulative cases.

[2]WHO (HIV/AIDS): *https://www.who.int/news-room/fact-sheets/detail/hiv-aids*
[3]WHO (Syphilis): *https://www.who.int/health-topics/syphilis*
[4]Source-code: *https://github.com/bruIzabel/-Spatio-Temporal-Analysis-of-STI-Notifications-in-Brazil*

In the Brazilian context, the health crisis profoundly destabilized epidemiological surveillance. In [Seara-Morais et al. 2023], the authors demonstrated that mobility restrictions caused abrupt declines in outpatient HIV and syphilis testing. During periods of strict isolation, syphilis positivity rates increased inversely to the testing volume, indicating that diagnoses were largely restricted to symptomatic cases. This diagnostic deficit is corroborated by [Sousa et al. 2024], who found that nationwide detection rates for HIV/AIDS, Hepatitis B, and Hepatitis C fell significantly below expected levels during the pandemic. Rather than indicating an actual decline in disease incidence, these apparent drops reflect severe disruptions in routine screening, suggesting a potential rise in untreated infections.

To quantify the deviation between observed and expected notification counts during the pandemic, we employ Ordinary Least Squares (OLS) linear regression [Gray 2002]. OLS estimates the underlying temporal trend from pre-pandemic data by minimizing the sum of squared residuals, resulting in a fitted line. For time series data, OLS provides a straightforward counterfactual baseline: the number of cases that would have been expected had the pandemic trend continued uninterrupted (by extrapolating the line to subsequent periods).

To uncover spatial similarity patterns in notification series across Brazilian states, we employ Hierarchical Agglomerative clustering with Ward's linkage [Ward-Jr. 1963]. This method minimizes the total within-cluster variance at each merge step. All time series are z-normalized so that grouping reflect similarity in temporal shape rather than absolute notification numbers. Selecting the number of clusters can be guided by the largest-gap criterion applied ot the Ward linkage distances (also known as the elbow method). At each step of the agglomerative process, the algorithm records the distance (or dissimilarity) at which two groups are merged. A large jump between consecutive merge distances indicates that the algorithm is forced to combine groups that are substantially different. Thus, the point before this "big jump" represents a natural partitioning in the data.

3. Methodology

The analysis methodology has four main steps, as depicted in Figure 1 and described next.

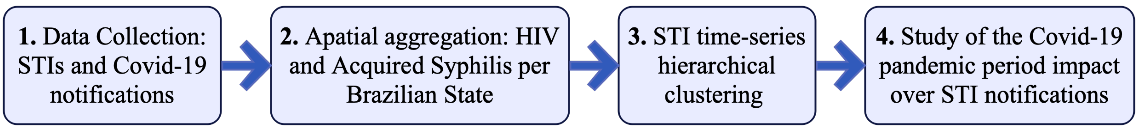


Figure 1. Pipeline to analyze the effect of COVID-19 pandemic on STI cases.

Step 1: Data collection. We collected STI data from DataSUS and Coronavirus Brazil[5], open-data platforms maintained by the Brazilian government. We focused on HIV and Acquired Syphilis, as STI of compulsory notification in Brazil, selecting notifications from 2015 to 2021, *specifically from female subjects*. We also collected COVID-19 case notifications from years 2020 and 2021. Both datasets were joined by municipality and state, totalizing 5,570 Brazilian cities.

Step 2: Spatial aggregation. We spatially aggregated the STI time series at the state level. Since each state is composed of multiple municipalities, we summed the annual

[5]Coronavirus Brazil: *https://covid.saude.gov.br/*, last access in April 22, 2026

notification counts of all municipalities belonging to the same state, year by year. This procedure aggregated municipal-level notifications into a single macroscopic time series for each state, resulting in 27 state-level entries (including the Federal District, DF). Consequently, each state was represented by a single time series describing the total annual burden of the specific STI within its territory. The same aggregation procedure was applied to COVID-19 case counts, enabling a consistent assessment of state-level correlations between pandemic burden and deviations in STI notifications.

Step 3: Time-series clustering. Hierarchical clustering was applied to the z-score normalized time series on STI notifications using Ward's linkage method, minimizing the total within-cluster variance at each merge step. With this method, we obtained a dendogram for each STI, displaying the hierarchical structure of similarity among states: those joined at low-heights are highly similar in their temporal trajectory, and those merged at greater heights represent the fusion of increasingly dissimilar groups. Each cluster groups states that share a similar shape in their trajectory in the study period, regardless of absolute volume (since series were z-score normalized prior to clustering).

Step 4: Observed versus projected STI notifications. For this step, we fit a linear model using the Ordinary Least Squares (OLS) on the STI notification counts prior to 2020. This regression aims to capture the underlying temporal trend in each STI reporting data before the pandemic. Then, the fitted model was extrapolated to the pandemic years (2020 and 2021) to produce the projected notification count, representing the number of cases that would have been expected had the pre-pandemic trend continued uninterrupted.

The pandemic notification deviation (Δ_c) is given by the difference between the observed (o_c) and the projected (p_c) counts of a given STI for each year $\Delta_c = o_c - p_c$ by state. States where the pre-pandemic series had zero variance were assigned a deviation of zero, since no trend could be estimated. A positive deviation indicates that notifications exceeded the pre-pandemic trend (more cases reported than expected). A negative deviation suggests a drop in reporting relative to the trend, which may reflect disruptions in health-seeking behavior, reduced access to testing, or underreporting due to the pandemic context and possible overload of the health system. The analysis uses absolute notification counts, so states with larger populations naturally have larger absolute deviations.

4. Results

In this section we present the experimental results obtained with open data.

4.1. Time-series analysis of HIV notifications

Figure 2(a) presents the hierarchical clustering dendogram for HIV notification time series aggregated by Brazilian state. The clustering resulted in 9 groups of states presenting similar temporal trajectories over the 2015-2021 period. States merged at low hights, such as PI, PE, MA, RS, SP, MG and RJ, share highly similar notification shapes, suggesting comparable epidemiological dynamics. Most clusters exhibit a drop pattern during the pandemic years, with magnitude varying across groups.

Figure 2(b) contrasts observed (2020+2021) versus projected HIV notifications per state, where projections are based on the 2015-2019 linear trend. In the plot, the diagonal dashed line ($y = x$) represents the no-deviation baseline: states above the line reported more cases than expected, while those below experienced a notification drop.

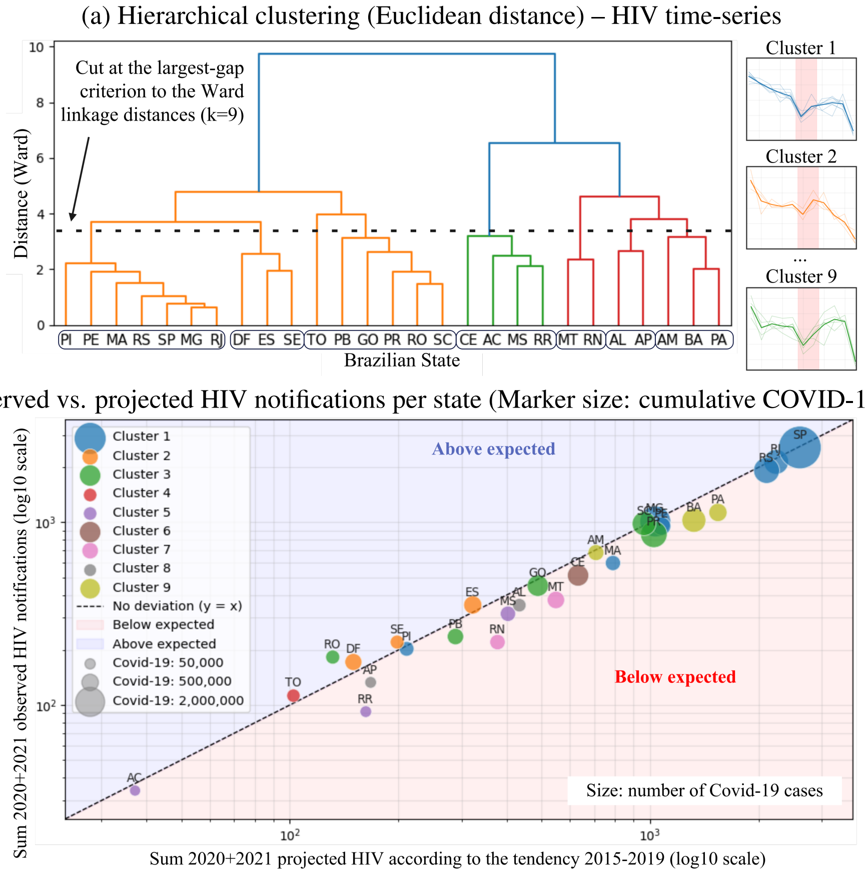


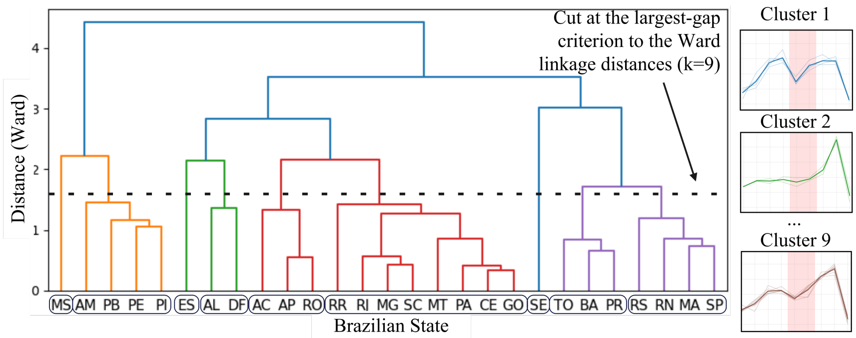
Figure 2. Clustering (a) of HIV notification time-series by state, and (b) observed versus projected HIV notifications for 2020-2021.

The majority of states fall below the diagonal, indicating a widespread suppression of HIV notifications during the pandemic. The marker size gives the cumulative COVID-19 burden in that State. We notice that states with larger caseloads, such as SP, RJ, and PR, tend to cluster near or slightly below the diagonal, suggesting that higher healthcare activation may have partially sustained testing and reporting capacity. In contrast, states like RR, RN and MT show pronounced below-diagonal positioning, which could indicate a collapse in notification infrastructure relative to their pre-pandemic trend.

4.2. Time-series analysis of Syphilis notifications

Figure 3(a) shows the hierarchical clustering result for Acquired Syphilis notifications. The number of clusters was also 9, but the grouping structure was different compared to HIV. The overall cut point was lower, indicating greater temporal homogeneity among states for Acquired Syphilis. Figure 3(b) presents the observed versus projected Acquired Syphilis notifications for 2020+2021. In contrast to HIV, the majority of states appear below the diagonal, indicating that Acquired Syphilis notifications were also broadly suppressed during the pandemic period relative to the pre-pandemic trend. Again, it is possible to observe that states such as SP, MG, RS, SC, and RJ remain closer to the diagonal line, even with high COVID-19 notification rates. This trend suggests that states with larger infrastructure and higher COVID-19 caseloads were better able to maintain notification capacity even under pandemic pressure over the healthcare system.

(a) Hierarchical clustering (Euclidean distance) of 'Acquired Syphilis' time-series



(b) Observed vs. projected Acquired Syphilis notifications per state (Marker size: cumulative COVID-19 cases)

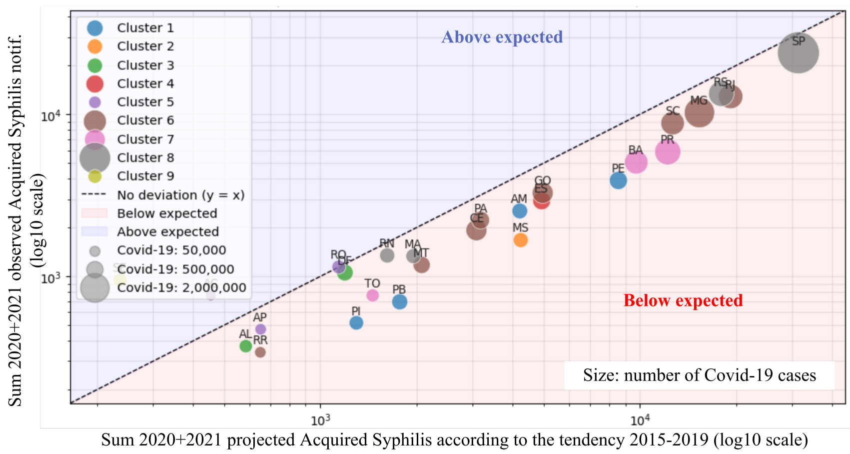


Figure 3. Clustering (a) of Acq. Syphilis notification time-series by state, and (b) observed versus projected Acq. Syphilis notifications for 2020-2021.

5. Conclusions

The results for both STI reveal a geographically heterogeneous but structurally consistent pattern of pandemic-related notification disruption. While HIV notifications declined in nearly all Brazilian states, Acquired Syphilis exhibited a more heterogeneous spatial pattern, where the majority of states notified Acquired Syphilis cases below the expected rates. In both cases, states with larger populations and higher COVID-19 cases tended to maintain notification levels closer to the expected trend, underscoring the role of healthcare infrastructure capacity as an important factor in diagnostic suppression.

With this research it was possible to contribute to WHO agenda and focus on priority STI research areas in order to end STI as a public health problem. Further analysis could evaluate the pattern of other comorbidities that affect disproportionately most vulnerable groups. One limitation of this study is the possibility that the number of cases may be underestimated due to underreporting. To circumvent this possibility, only available cases were considered, that is, the cases reported each year of the historical series.

Furthermore, DATASUS is the only government database where notifications are accessible and with which we can evaluate the effectiveness of public policies implemented to combat the problem. Therefore, it is a reliable source of data and widely used for public health studies. Thus, health authorities can use the results of this and other studies to support specific policies for their municipality and/or Federative Unit.

Acknowledgements. This work was supported by the São Paulo Research Foundation (FAPESP – grants #2026/11130-2, #2024/15430-5, #2024/13328-9, #2024/02813-3), the Coordination for Higher Education Personnel Improvement (CAPES – Finance Code 001), the National Research Council (CNPq – grants #401496/2025-2 and #402987/2025-0), and the University of São Paulo (PRPI 1032, grant #207).

Origin of the work. This paper reports results obtained from an Scientific Initiation (IC) project by the undergraduate students B.I.S.P. and A.B.S.L. under supervision of M.T.C., which focuses on data analysis to support an ongoing Doctoral project at EERP-USP by A.L.L.B. under supervision of E.G.

Generative AI usage statement. The NotebookLM tool was used to support the literature review, and Grammarly was used to support the linguistic revision of the manuscript.

References

Aribe-Jr., S. G., Gerardo, B. D., and Medina, R. P. (2022). Time series forecasting of hiv/aids in the philippines using deep learning: Does covid-19 epidemic matter? *International Journal of Emerging Technology and Advanced Engineering*, 12(9):144–157. DOI: 10.46338/ijetae0922_15.

Brazil. Ministry of Health (2023). Epidemiological bulletin: Hiv and aids 2023. Technical report, Health and Environmental Surveillance Secretariat, Department of HIV/AIDS, Tuberculosis, Viral Hepatitis and Sexually Transmitted Infections, Brasília, DF, Brazil. Special Issue, ISSN 2358-9450.

Brazil. Ministry of Health (2024). Epidemiological report – syphilis 2024. Technical report, Secretariat of Health and Environmental Surveillance, Department of HIV/AIDS, Tuberculosis, Viral Hepatitis and Sexually Transmitted Infections, Brasilia, DF, Brazil. Special Issue, ISSN 2358-9450.

Cavalcante, J. R., Cardoso-dos Santos, A. C., Bremm, J. M., Lobo, A. d. P., Macário, E. M., Oliveira, W. K. d., and França, G. V. A. d. (2020). Covid-19 no brasil: evolução da epidemia até a semana epidemiológica 20 de 2020. *Epidemiologia e Serviços de Saúde*, 29(4). DOI: 10.5123/s1679-49742020000400010.

Gottlieb, S. L., Spielman, E., et al. (2024). Who global research priorities for sexually transmitted infections. *The Lancet Global Health*, 12(9):e1544–e1551. DOI: 10.1016/s2214-109x(24)00266-3.

Gray, J. B. (2002). Introduction to linear regression analysis. *Technometrics*, 44(2):191–192. DOI: 10.1198/TECH.2002.S714.

Seara-Morais, G. J., Pousada, B. F., Escaleira, F. F., Doi, A. M., Welter, E. A. R., and Avelino-Silva, V. I. (2023). Mobility restrictions during the covid-19 pandemic and reduced outpatient hiv and syphilis testing in brazil. *Brazilian Journal of Infectious Diseases*, 27(3):102771. DOI: 10.1016/j.bjid.2023.102771.

Sentís, A., Prats-Uribe, A., López-Corbeto, E., Montoro-Fernandez, M., Nomah, D. K., de Olalla, P. G., Mercuriali, L., Borrell, N., Guadalupe-Fernández, V., Reyes-Urueña, J., and Casabona, J. (2021). The impact of the covid-19 pandemic on sexually transmitted infections surveillance data: incidence drop or artefact? *BMC Public Health*, 21(1):1637. DOI: 10.1186/s12889-021-11630-x.

Sousa, R. F. V. d., Mallet, J. R. d. S., Nascimento, E. F. d., Cardoso, D. R. F., Neves, J. M. M., and Oliveira, B. F. A. d. (2024). Análise das infecções sexualmente transmissíveis durante a pandemia da covid-19 no brasil. *Revista Eletrônica Acervo Saúde*, 24(8):e18163. DOI: 10.25248/reas.e18163.2024.

Wang, L., Wang, Y., Ye, D., and Liu, Q. (2020). Review of the 2019 novel coronavirus (sars-cov-2) based on current evidence. *International Journal of Antimicrobial Agents*, 55:105948. DOI: 10.1016/j.ijantimicag.2020.105948.

Ward-Jr., J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244. DOI: 10.1080/01621459.1963.10500845.

Yan, X., Wang, X., Zhang, X., Wang, L., Zhang, B., and Jia, Z. (2021). The epidemic of sexually transmitted diseases under the influence of covid-19 in china. *Frontiers in Public Health*, 9:737817. DOI: 10.3389/fpubh.2021.737817.