Estimating the real burden of gestational syphilis in Brazil, 2007–2018: a Bayesian modeling study

Guilherme Lopes de Oliveira,[a,b,*] Andrêa J. F. Ferreira,[a,c] Carlos Antônio de S. S. Teles,[a] Enny S. Paixao,[a,d] Rosemeire Fiaccone,[a,e] Raquel Lana,[a,f] Rosana Aquino,[a,g] Andrey Moreira Cardoso,[h] Maria Auxiliadora Soares,[a,g] Idália Oliveira dos Santos,[a,g] Marcos Pereira,[a,g] Maurício L. Barreto,[a,g] and Maria Yury Ichihara[a]

[a]Centre of Data and Knowledge Integration for Health (CIDACS), Instituto Gonçalo Moniz, Fiocruz, Salvador, Bahia, Brazil
[b]Department of Computing, Federal Centre of Technological Education of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil
[c]The Ubuntu Center on Racism, Global Movement, Population and Equity, School of Public Health, Drexel University, Pennsylvania, USA
[d]London School of Hygiene and Tropical Medicine, London, UK
[e]Statistics Department, Institute of Mathematics, Federal University of Bahia, Salvador, Bahia, Brazil
[f]Barcelona Supercomputing Center, Catalonia, Spain
[g]Institute of Collective Health, Federal University of Bahia, Salvador, Bahia, Brazil
[h]National School of Public Health, Fiocruz, Rio de Janeiro, Rio de Janeiro, Brazil

The Lancet Regional Health - Americas 2023;25: 100564

Published Online xxx https://doi.org/10.1016/j.lana.2023.100564

Summary

Background Although several studies have estimated gestational syphilis (GS) incidence in several countries, underreporting correction is rarely considered. This study aimed to estimate the level of under-registration and correct the GS incidence rates in the 557 Brazilian microregions.

Methods Brazilian GS notifications between 2007 and 2018 were obtained from the SINAN-Syphilis system. A cluster analysis was performed to group microregions according to the quality of GS notification. A Bayesian hierarchical Poisson regression model was applied to estimate the reporting probabilities among the clusters and to correct the associated incidence rates.

Findings We estimate that 45,196 (90%-HPD: 13,299; 79,310) GS cases were underreported in Brazil from 2007 to 2018, representing a coverage of 87.12% (90%-HPD: 79.40%; 95.83%) of registered cases, where HPD stands for the Bayesian highest posterior density credible interval. Underreporting levels differ across the country, with microregions in North and Northeast regions presenting the highest percentage of missed cases. After underreporting correction, Brazil's estimated GS incidence rate increased from 8.74 to 10.02 per 1000 live births in the same period.

Interpretation Our findings highlight disparities in the registration level and incidence rate of GS in Brazil, reflecting regional heterogeneity in the quality of syphilis surveillance, access to prenatal care, and childbirth assistance services. This study provides robust evidence to enhance national surveillance systems, guide specific policies for GS detection disease control, and potentially mitigate the harmful consequences of mother-to-child transmission. The methodology might be applied in other regions to correct disease underreporting.

Funding National Council for Scientific and Technological Development; The Bill Melinda Gates Foundation and Wellcome Trust.

Copyright © 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Keywords: Sexual transmitted infection; Syphilis; Syphilis during pregnancy; Under-registration; Underreporting

Introduction

Syphilis is a sexually transmitted infection caused by *Treponema pallidum*, first identified in 1905.[1] Despite that, it is still a persistent public health issue in many countries worldwide since the rate of syphilis has increased dramatically in recent decades,[1] becoming both a burden and damage to public health,[2–4] particularly, during pregnancy. Syphilis diagnosed during pregnancy (also called Gestational Syphilis (GS) and hereafter referred to as such) is the second leading cause of stillbirth globally and results in severe adverse outcomes in more than 50% of cases, such as prematurity,

*Corresponding author. Avenida Amazonas, 7675 Nova Gameleira, Belo Horizonte, Minas Gerais, Brazil.
E-mail addresses: guilhermeoliveira@cefetmg.br, guilopes2110@gmail.com (G. Lopes de Oliveira).

Research in context

Evidence before this study

The World Health Organization (WHO) estimates that syphilis infection affects one million pregnancies yearly worldwide, with important consequences for pregnant women and newborns. Untreated gestational syphilis (GS) has been linked to a significantly higher incidence of stillbirth, neonatal death, prematurity, and low birth weight. Notifying GS is compulsory for all WHO member countries; it is mandatory to communicate all cases to health authorities and investigate them. However, the lack of quality in the case detection and data recording processes may contribute to under-registration of GS cases on the health information systems, which underestimates the real magnitude of the disease, especially in middle-income countries such as Brazil.

We searched PubMed/Medline, EMBASE, and LILACS using the search strategy ("underreporting of syphilis" OR "syphilis underreporting" OR "under-registration of syphilis" OR "under-registration") AND ("Pregnant women") AND (Brazil) seeking for studies on under-registration of GS across Brazil until July 2023. We supplemented the literature review with internet searches (Google scholar and Semantic scholar), analysis of references in identified papers, and the authors' own knowledge. We found 6 studies focusing on small to medium-sized Brazilian cities, which indicated: 13.74% of GS under-registration between 2010 and 2013 in the west of São Paulo State; 6.5% of GS notification in Montes Claros city, Minas Gerais State, between 2007 and 2013; 50% more cases of congenital syphilis than GS cases from 2009 to 2015 in Londrina, Paraná State; and 46% of GS underreporting in a special indigenous sanitary district in Mato Grosso State. None of these studies covered the whole country and they do not considered factors associated with underreported cases of GS or even provided corrected estimation of GS incidence. Based on such literature review, the modeling of syphilis under-registration is also rare worldwide since none study was found on this regard.

Added value of this study

This is the first study to map the quality of GS notifications at the national, regional, and small area levels in a large country such as Brazil, emphasizing that GS detection and incidence are related to socio-economic conditions and access to health services such as deprivation level and adequate prenatal care. States and microregions in the Brazilian North and Northeast regions presented the highest GS underreporting cases. Health managers and policymakers could easily use our methodology to amplify national syphilis surveillance and estimate the real incidence of GS. Also, the statistical modeling strategy may be used as a reference for application in other middle-income countries.

Implications of all the available evidence

Taken together, modeling results from this national-level study on GS burden in Brazil, along with findings from other local surveys, suggest that GS remains a significant public health problem, highlighting the need for improved surveillance. Our research reveals important differences in underreporting of GS cases across Brazilian regions, as well as the association of syphilis surveillance quality with socioeconomic conditions and access to healthcare. This underscores the urgent need to enhance syphilis surveillance and case notification within the healthcare system, particularly in the country's poorest region. Furthermore, correcting GS incidence at different area levels could provide a better understanding of the epidemiological profile and help estimate the social, economic, and health burdens associated with the disease nationwide. Additionally, our findings offer valuable insights for policymakers to enhance planning and take action to reduce the burden of GS, thereby contributing to the reduction of mother-to-child transmission and adverse effects during childbirth.

low birth weight, miscarriage, stillbirth, and neonatal death.[1]

Data from 2016 shows that over 900,000 pregnant women were diagnosed with syphilis worldwide, resulting in more than 350,000 adverse outcomes, from which over 200,000 were stillbirths or neonatal deaths.[5,6] The region of the Americas had the second-highest estimated prevalence of GS worldwide (0.86%), after Africa (1.52%).[6] In 2016, the Brazilian Ministry of Health declared gestational and congenital syphilis as an epidemic in the country due to the exponential increase in their incidence rates, with unfavorable outcomes for both women and newborns.[7]

Despite the mandatory registration of gestational syphilis and the notable increase in notifications in recent years, underreporting GS cases is still a problem in Brazil.[8–12] Failures on report GS could bias the actual scenario of the disease and compromise the knowledge of the real burden of GS in the country. Many factors can influence the underreporting (under-registration) of GS cases in Brazil, mainly the inequality of access to health services and prenatal care, failures in syphilis testing during the antenatal period, and failures to register the notification in the Health Information Systems (HIS).[13–15]

Accurate reporting of GS cases allows policymakers to design control policies and helps to reduce its consequences for pregnant individuals and their offspring. Although Brazil and other countries recognize the importance of improving the quality of syphilis detection and notification during pregnancy,[16] epidemiological studies adopting robust methodologies to analyze

underreporting of GS are rare. Few studies have addressed this issue in Brazil and most of the available literature presents analyses with small sample sizes and focuses on specific regions.[9–11] Despite those efforts, there are still gaps in the investigation of underreporting of GS across the country and correction of its incidence. This study aims to estimate the real burden of GS in Brazilian microregions from 2007 to 2018 by applying a statistical modeling strategy to correct under-registration of cases in an official data reporting system, which may be used as a reference for application in other middle-income countries. Measures of underreporting levels are provided by comparing the model estimates with the reported data.

Methods

Study design and target population

We conducted a cross-sectional ecological study design, considering GS cases notified in Brazil between 2007 and 2018. Brazil, a continental country, divided into 27 states and five macroregions (North, Northeast, South, Southeast and Midwest), comprising 557 microregions, which are defined as contiguous aggregations of the 5570 Brazilian municipalities.[17]

Data sources and variables description

This study used administrative non-identified data available from public Brazilian databases. The outcome variable, the count of GS cases, was extracted from the Notifiable Diseases Information System (SINAN in Portuguese: *Sistema de Informação de Agravos de Notificação*) using data received from the Brazilian Ministry of Health in 2019. In Brazil, it is mandatory for healthcare professionals to register GS cases through SINAN-Syphilis, where they complete a form including personal, socioeconomic, and gestational information of the patient at the time of diagnosis.[14]

GS cases registered in SINAN-Syphilis followed the definition described by the Brazilian Ministry of Health[14,18] considering active detection. This includes cases where a pregnant individual tests positive for the venereal disease research laboratory (VDRL) test with any titration during pregnancy, provided they have not received prior treatment. Additionally, cases may be considered if a confirmatory test (treponemal serology) is also positive. From the total 305,891 GS cases notified in Brazil between 2007 and 2018, we excluded: i) 131 cases for which the municipality of residence was not identified or it did not match the official municipality codes of the Brazilian Institute of Geography and Statistics (IBGE)[17]; and ii) 2 cases registered in Fernando de Noronha, a Brazilian archipelago located more than 500 km away from the Brazilian mainland, which could prevent a spatial analysis with the method considered in our study. Therefore, our analysis was based on 305,758 (99.96%) GS cases notified in SINAN-Syphilis from 2007 to 2018.

For statistical analyses of GS incidence rates and underreporting level in Brazilian microregions, we collected socio-demographic information and access to health service indicators, as well as factors associated to quality of SINAN-Syphilis form registration. The GS incidence rates were calculated per 1000 live births in the microregion. All variables considered in our study are detailed in Table 1.

Statistical modeling

In the statistical literature, correction of underreporting in count data has been done based on censored[21] and compound[22–26] Poisson approaches. In both cases, due to identifiability issues, the model depends on the availability of extra information to support the data reporting process modeling. Thus, statistical inference is performed under the Bayesian framework, in which the extra information is accommodated through appropriate prior distributions. In this work, we applied the Bayesian hierarchical Poisson model proposed by Oliveira et al. (2022).[24] Provided that the Brazilian microregions are clustered according to the quality of GS data, this model only requires specification of an informative prior distribution for the reporting level in areas belonging to the best data quality cluster. This requirement is more feasible in our case, as we do not have: (i) validation datasets[25]; (ii) information about global GS reporting level in Brazil[23]; (iii) reliable local active search survey[26]; nor (iv) informative prior for the reporting probability in all groups of areas.[22]

Clustering analysis

We applied the K-means method to group the microregions according to their quality of GS data. We considered seven indicators as proxies for GS data quality: coverage of prenatal care; adequacy of prenatal care; test acceptability; diagnosis opportunity; consistency; adequacy of case classification; and completeness score of GS care assistance indicators (see Table 1). Variables were standardized before applying the K-means method in software R,[27] using packages[28] and factoextra.[29] Definition of the possible optimal number K of clusters was based on the elbow method, which relies on evaluating the total within-cluster sums of squares.[30] A sensitivity analysis was performed considering different values for the number K of clusters around the optimal region identified with the elbow method. In each case, the clusters were hierarchically related from the best to the worst with respect to data quality. To accomplish this, we analyzed the within-cluster centroid, which contained the mean value for the seven variables used in the clustering analysis. The higher the centroid mean, the better the data quality.

Model specification

Let Y_i be the total reported (observed) GS cases in microregion i, where $i = 1, ..., 557$. In underreported

scenarios, Oliveira et al. (2022)[24] proposed assuming that

$$Y_i|\mu_i, \varepsilon_i \sim Poisson(\mu_i\varepsilon_i), \quad \text{(Eq. 1)}$$

where parameters, $\mu_i>0$ and $0<\varepsilon_i<1$ denote, respectively, the GS incidence rate and the proportion of the true (unobserved) cases reported in the i-th microregion. We modeled the GS incidence rate μ_i through the following regression structure

$$log(\mu_i) = log(P_i) + \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \phi_i + \delta_i, \quad \text{(Eq. 2)}$$

where P_i = total live births (offset), X_{1i} = Brazilian deprivation index, X_{2i} = proportion of live births with adequate prenatal care, X_{3i} = proportion of non-white live births (see descriptions in Table 1), and *log* represents the natural logarithm function. Those three covariates were selected on the basis of literature review, in which it is known that socio-demographic and adequacy to health care indicators tends to be related to the disease incidence.[16,31–33] They were available in our data sources and then were used to account for discrepancies on socio-demographic and adequacy to health care among the Brazilian microregions. Variables X_1, X_2 and X_3 were standardized before applying the model to estimate their respective effects (parameters) $\beta_0, \beta_1, \beta_2$ and β_3. Terms δ_i and ϕ_i are, respectively, random effects accounting for residual overdispersion induced by local and spatial variations. As it is usual in spatial statistics modeling, the prior distribution for the spatially structured random effect ϕ_i is represented by an intrinsic conditional auto-regressive (ICAR) model and δ_i follows a non-informative Gaussian prior.[34]

Following Oliveira et al. (2022),[24] we assume that $\varepsilon_i = 1-\gamma_1$ for all areas belonging to the best data quality group, $\varepsilon_i = 1-\gamma_1-\gamma_2$ for all areas belonging to the second best data quality group, and so on, implying that $\varepsilon_i = 1-\gamma_1-\gamma_2-\ldots-\gamma_K$ for all areas within the worse data

Source and Reference	Variable Name	Definition
Outcome Variable		
SINAN-Syphilis	Gestational syphilis (GS) notified cases	GS cases are defined by an active detection through a reagent result for venereal disease research laboratory test (VDRL), with any titration, during any stage of the gestational period, since pregnant were not previously treated, and/or presented a confirmatory test for syphilis (treponemal serology). Cases were aggregated in each of the 557 Brazilian microregions for the period 2007–2018.
Factors considered in GS incidence rates modeling		
CIDACS[19]	Brazilian deprivation index (BDI)	Composite socioeconomic measure estimated for all Brazilian municipalities based on income, schooling, and household conditions information from the 2010 Brazilian Census, ranging from less deprived to more deprived areas.[19] The average BDI of the municipalities was calculated for the microregion level.
Information System on Live Births (in Portuguese: *Sistema de Informações sobre Nascidos Vivos*—SINASC)	Live births	Number of live births reported in Brazilian microregions from 2007 to 2018.
	Adequacy of prenatal care	The proportion of live births reported in Brazilian microregions between 2007 and 2018 whose mothers attended seven or more prenatal appointments.
	Non-white live births	The proportion of non-white live births registered in Brazilian microregions between 2007 and 2018.
Proxies for quality of GS cases registration		
Information System on Live Births (in Portuguese: *Sistema de Informações sobre Nascidos Vivos*—SINASC)	Coverage of prenatal care	The proportion of live births reported in Brazilian microregions between 2007 and 2018 whose mother attended at least one prenatal appointment in the same study period.
SINAN-Syphilis	Test acceptability	The proportion of GS notified cases with a qualitative result (reactive or non-reactive) for at least one non-treponemal or treponemal test in each Brazilian microregion from 2007 to 2018.
	Diagnosis opportunity	The proportion of GS notified cases in which the diagnosis was laboratory confirmed during the first or second trimester of pregnancy in each Brazilian microregion from 2007 to 2018.
	Consistency	The proportion of municipalities within the microregion for which the number of registered GS cases is greater or equal to congenital syphilis cases registered from 2007 to 2018.
	Adequacy of case classification	The proportion of GS notified cases from 2007 to 2018 with sufficient information to be classified within one of the three case definitions (primary, secondary, and tertiary or latent) established by the Brazilian Ministry of Health.
Oliveira et al. (2022)[20]	Completeness score of GS care assistance indicators	A completeness score proposed with basis on variables associated with care assistance[20] (e.g., information on partner's treatment, treponemal test realization, treatment regime specification, prenatal care attendance, and description of the evolution of case) extracted from the SINAN-Syphilis forms for GS notified cases in the period 2007–2018.

GS = gestational syphilis; SINAN: Notifiable Diseases Information System (SINAN in Portuguese: *Sistema de Informação de Agravos de Notificação*); SINASC: Information System on Live Births (SINASC in Portuguese: *Sistema de Informações sobre Nascidos Vivos*); CIDACS: Centre of Data and Knowledge Integration for Health (CIDACS in Portuguese: *Centro de Integração de Dados e Conhecimentos para Saúde*); BDI: Brazilian Deprivation Index.

Table 1: **Description of variables and data sources used in the study.**

quality cluster. Parameter γ_1 represent the proportion of underreported cases in areas classified in the highest level of data quality; γ_2 is the increment on such proportion for areas experiencing the second highest data quality level, and so on. Such parameters are hierarchically correlated through an appropriate conditional Uniform prior distribution.

The most important and attractive feature of this modeling strategy is that, to attain identifiability, it only requires an informative prior distribution for parameter γ_1, which represents the proportion of underreporting in those areas belonging to the best data quality cluster. An informative prior means a distribution that is highly concentrated in a specific subset of the parametric space. In our application to Brazilian GS data, with basis on literature review and information collected from experts' on the study of syphilis burden in Brazil, we performed a sensitivity analysis considering four different prior distributions for $\gamma_1 \sim Uniform(0, 0.02)$, $\gamma_1 \sim Uniform(0, 0.05)$, $\gamma_1 \sim Uniform(0.05, 0.10)$ and $\gamma_1 \sim Uniform(0.05, 0.10)$. Respectively, these priors imply that areas classified in the highest level of data quality report, on average, 99.0%, 97.5%, 95.0% and 92.5% of the GS cases. All these priors impose that less than 10.0% of the true GS cases are missed in such areas. Non-informative priors are elicited for parameters $\gamma_2, ..., \gamma_K$.

By varying the number of clusters K and the prior distribution for parameter γ_1, we fitted 12 different models. The WAIC[35] and the LPML[36] goodness-of-fit metrics were considered for model selection. For the WAIC, the smaller the value, the better the model fitted the data. Larger values of LPML indicate better fit. A more detailed discussion is presented in Section 1 of the Supplementary File, which includes the prior specification for all parameters of our Bayesian model as well as particularities of model implementation and validation.

It worth mentioning that, for each microregion, provided that μ_i and ε_i are estimated, the underreported GS cases, denoted here by Z_i, can be estimated from the predictive posterior distribution $Z_i|\mu_i, \varepsilon_i \sim Poisson(\mu_i(1 - \varepsilon_i))$. Then, the total GS cases can be estimated/predicted as $T_i = Y_i + Z_i$, in such a way that the ratio between Y_i and T_i provides a coverage for the SINAN-Syphilis. The Bayesian highest posterior density credible interval with level 90% was used to evaluate significance of covariates X_1, X_2 and X_3 and to quantify uncertainty in the model predictions.

Ethics approval

Not applicable. All analyses used de-identified and publicly available data.

Role of the funding source

The study's funders had no role in the study design, data collection, data analysis, interpretation, or writing of the manuscript.

Results

From January 2007 to December 2018, 305,758 GS cases were registered in Brazil corresponding to a GS incidence rate of 8.74 per 1000 live births. GS incidence rates vary across the Brazilian territory (Fig. 1), where the highest observed incidence rate at microregion and state levels were, respectively, 23.45 and 18.12, while the lowest ones were around 1.01 and 4.07, respectively. Northeast and North regions generally presented low incidence rates, as well as most microregions within Minas Gerais State (in Southeast region). Large clusters of microregions with high GS incidence rates were observed in states of Rio Grande do Sul (South region) and Mato Grosso do Sul (Midwest region), which is the only with an incidence rate above 10.2.

To correct for GS underreporting, we performed a sensitivity analysis by applying the Bayesian model with 12 different specifications, which are related to the combinations between the number of clusters (K) and the prior distribution elicited to parameter γ_1 (Supplementary Table S1 of the Supplementary File). The elbow method indicated that K = 9, K = 10 or K = 11 would be a good choice for the number of clusters (Fig. 2). According to the model evaluation metrics (Supplementary Table S1 of the Supplementary File), K = 9 was the best choice to aggregate the 557 Brazilian microregions according to the seven selected proxies of GS data quality. Data quality decreases from Cluster 1 (best data quality group) to Cluster 9 (worst data quality group) (Fig. 2). All the 17 microregions within Cluster 9 are located in the North (14 areas) and Northeast (3 areas) regions.

The sensitivity analysis indicated that, *a priori*, the proportion of underreported GS cases in the best data quality cluster is 1%, on average (Supplementary Table S1 of the Supplementary File). Fig. 3 displays the estimated posterior mean for the proportion of the true (unobserved) GS cases that is reported in each Brazilian microregion and state (obtained by averaging the posterior mean of microregions within each state). For microregions belonging to the best data quality group (Cluster 1), the posterior percentage of reported GS cases is 99.01%, on average, which is quite influenced to the mean of the informative prior specified to parameter γ_1 as expected. For areas in Clusters 2, on average, we estimate that more than 90% of the GS cases were correctly registered in SINAN-Syphilis. For Cluster 3 to 6, on average, the proportion of GS cases registered was estimated between 90% and 80%, whereas it was less than 80% for areas belonging to Clusters 7 to 9. In the two worst data quality clusters, our Bayesian model estimated that 34.86% and 30.21% of the GS cases were not reported, respectively. We estimate that all states in the North region registered less than 86.80% of their GS cases. Only four states (Rio Grande do Sul and Parana, in the South region, and São

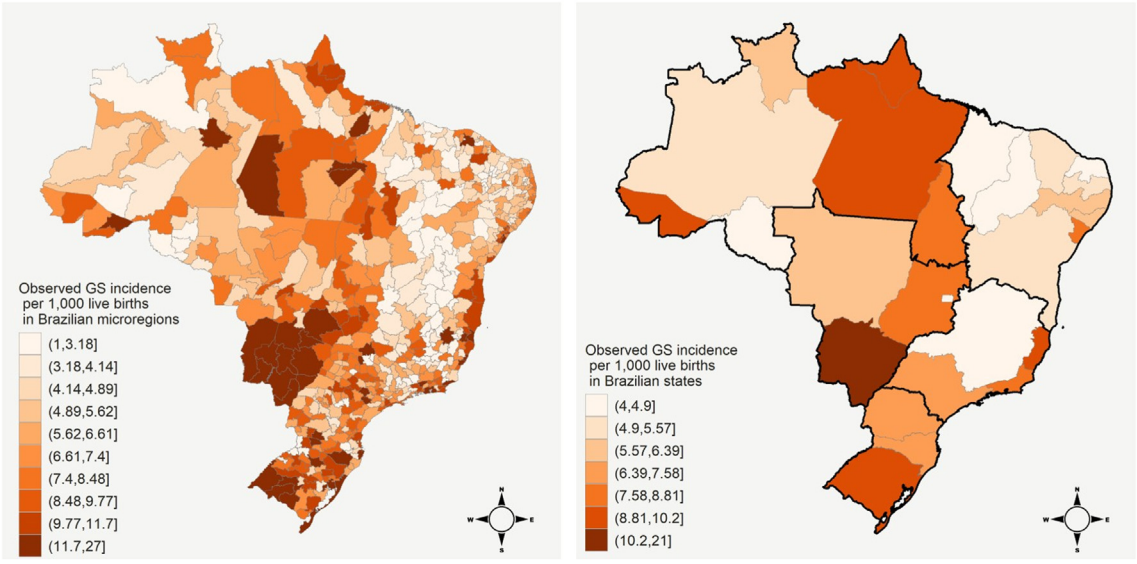

Fig. 1: Incidence rates of gestational syphilis (per 1000 live births) in the 557 Brazilian microregions (left) and the 27 Brazilian states (right), 2007–2018, according to the observed data. Brazilian regions (North, Northeast, Midwest, Southeast, and South) are separated by black lines on the right map.

Paulo and Espírito Santo, in Southeast region) reported more than 92.13% of GS cases.

After correcting for underreporting, the GS incidence rate in Brazil was estimated as 10.02 per 1000 live births in the period. We observed a substantial increase in the GS incidence rate for microregions belonging to the worst data quality clusters, which are located, in general, in the Northeast and North regions (Fig. 4). The highest estimated incidence rates at microregion and state levels were, respectively, 26.15 and 20.12, while the lowest ones were around 1.19 and 4.95, respectively. Six Brazilian states presented a posterior mean greater than 10.2 for the GS incidence rate per 1000 live births. Analysis of the effects of covariates considered for modeling the GS incidence rates is provided in Supplementary Table S2 of the Supplementary File.

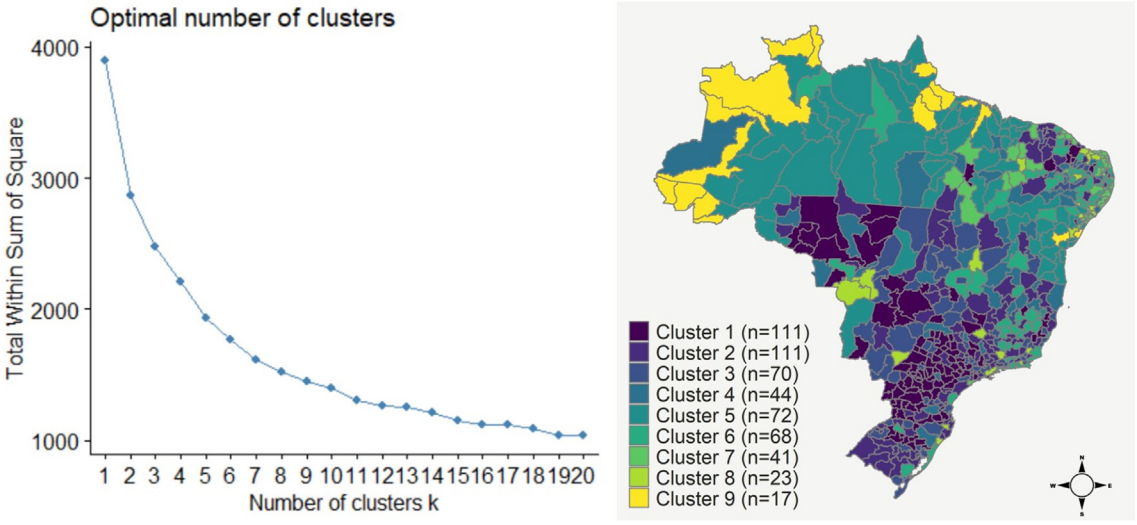

Fig. 2: Result of clustering analysis for quality of gestational syphilis reporting in Brazil, 2007–2018. Left: Plot of within cluster sum of squares versus the number of clusters K, indicating that around K = 10 is a good choice in this case according to the elbow method. Right: Map of the 557 Brazilian microregions according to the data reporting quality clusters obtained through the K-means method with K = 9, the best fit for the gestational syphilis database.

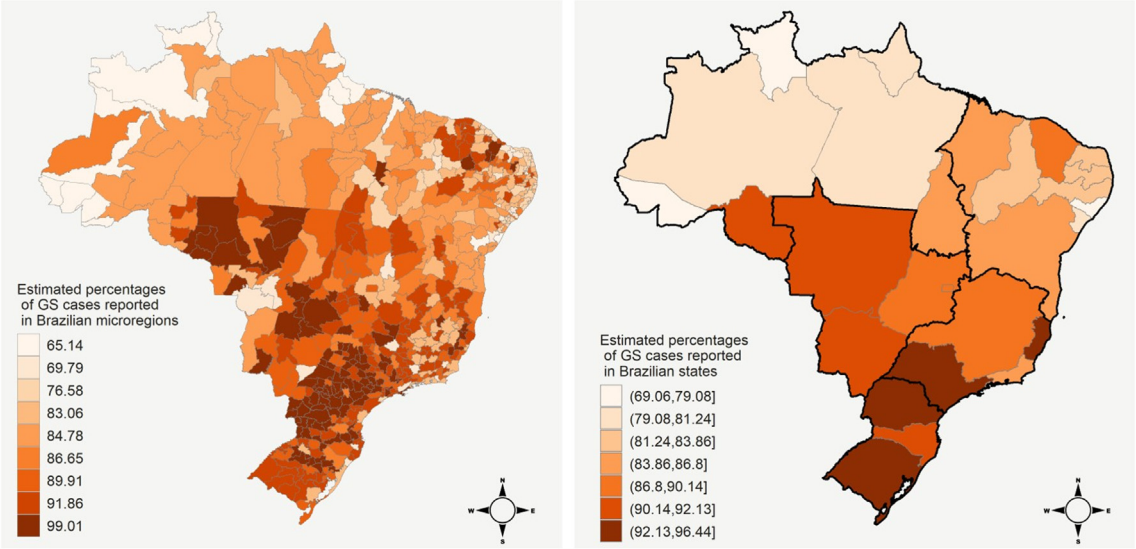

Fig. 3: Posterior mean for percentages of GS cases reported in the 557 Brazilian microregions (left) and in the 27 Brazilian states (right) provided by the fitted Bayesian model. Brazilian regions (North, Northeast, Midwest, Southeast and South) are separated by black lines in the right map.

The mean predicted total GS cases and coverage (percentage of registered cases) of SINAN-Syphilis at state level are shown on Table 2. On average, 45,196 (90%-HPD: 13,299; 79,310) GS cases were missed in Brazil from 2007 to 2018, which represents an estimated detection rate of 87.12% (90%-HPD: 79.40%; 95.83%) of all cases registered at national level. São Paulo (96.41% with 90%-HPD: 93.05%; 99.68%) and Roraima (66.60% with 90%-HPD: 56.01%; 81.83%) were, respectively, the states with the highest and smallest estimated GS coverage.

Discussion

Our study estimates that around 13% of the notifications of gestational syphilis were not registered in SINAN-Syphilis between 2007 and 2018 in Brazil. The underreporting pattern differs between regions, states

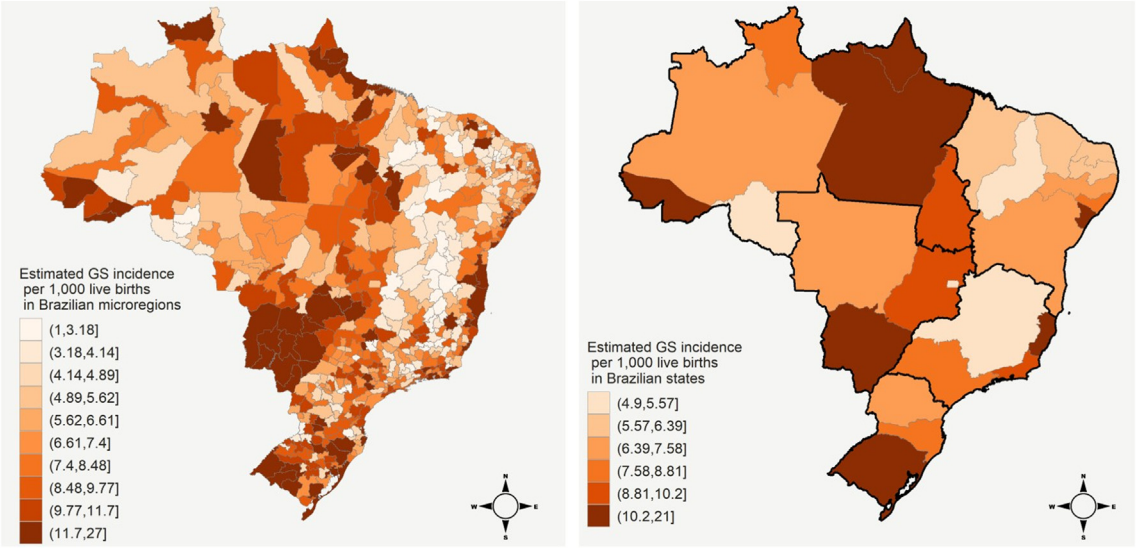

Fig. 4: Corrected incidence rates of gestational syphilis (per 1000 live births) in the 557 Brazilian microregions (left) and the 27 Brazilian states (right), 2007–2018, according to the posterior mean of the fitted Bayesian model. Brazilian regions (North, Northeast, Midwest, Southeast and South) are separated by black lines in the right map.

State	Number of microrregions	Observed cases	Mean predicted underreported cases	Mean predicted total cases	Mean predicted coverage (%)
North region					
Acre (AC)	5	2545	762 (279; 1245)	3307 (2824; 3790)	76.96 (67.15; 90.12)
Amazonas (AM)	13	8592	1681 (525; 2942)	10,273 (9117; 11,534)	83.64 (74.49; 94.24)
Amapá (AP)	4	2026	395 (110; 681)	2421 (2136; 2707)	83.68 (74.84; 94.85)
Pará (PA)	22	14,510	3302 (1148; 5587)	17,812 (15,658; 20,097)	81.46 (72.20; 92.67)
Rondônia (RO)	8	1819	235 (58; 427)	2054 (1877; 2246)	88.56 (80.99; 96.91)
Roraima (RR)	4	937	470 (208; 736)	1407 (1145; 1673)	66.60 (56.01; 81.83)
Tocantins (TO)	8	2577	476 (130; 850)	3053 (2707; 3427)	84.41 (75.20; 95.20)
Northeast region					
Alagoas (AL)	13	4128	1230 (454; 2037)	5358 (4582; 6165)	77.04 (66.96; 90.09)
Bahia (BA)	32	18,885	3904 (1282; 6760)	22,789 (20,167; 25,645)	82.87 (73.64; 93.64)
Ceará (CE)	33	9714	2079 (686; 3596)	11,793 (10,400; 13,310)	82.37 (72.98; 93.40)
Maranhão (MA)	21	7392	1515 (450; 2620)	8907 (7842; 10,012)	82.99 (73.83; 94.26)
Paraíba (PB)	23	4229	822 (224; 1465)	5051 (4453; 5694)	83.73 (74.27; 94.97)
Pernambuco (PE)	18	11,046	2847 (1123; 4878)	13,893 (12,169; 15,924)	79.51 (69.37; 90.77)
Piauí (PI)	15	2812	732 (247; 1255)	3544 (3059; 4067)	79.35 (69.14; 91.93)
Rio Grande do Norte (RN)	19	3217	970 (358; 1631)	4187 (3575; 4848)	76.83 (66.36; 89.99)
Sergipe (SE)	13	3682	1011 (346; 1704)	4693 (4028; 5386)	78.46 (68.36; 91.41)
Midwest region					
Distrito Federal (DF)	1	2447	281 (51; 529)	2728 (2498; 2976)	89.70 (82.22; 97.96)
Goiás (GO)	18	9957	1072 (195; 2052)	11,029 (10,152; 12,009)	90.28 (82.91; 98.08)
Mato Grosso do Sul (MS)	11	10,009	1062 (191; 2034)	11,071 (10,200; 12,043)	90.41 (83.11; 98.13)
Mato Grosso (MT)	22	3998	736 (267; 1236)	4734 (4265; 5234)	84.45 (76.39; 93.74)
Southeast region					
Espírito Santo (ES)	13	8900	1450 (428; 2503)	1035 (9328; 11,403)	85.99 (78.05; 95.41)
Minas Gerais (MG)	66	19,951	3033 (784; 5467)	22,984 (20,735; 25,418)	86.80 (78.49; 96.22)
Rio de Janeiro (RJ)	18	44,910	9113 (2999; 15,153)	54,023 (47,909; 60,063)	83.13 (74.77; 93.74)
São Paulo (SP)	63	63,387	2359 (206; 4735)	65,746 (63,593; 68,122)	96.41 (93.05; 99.68)
South region					
Paraná (PR)	39	14,513	750 (65; 1546)	15,263 (14,578; 16,059)	95.09 (90.37; 99.55)
Rio Grande do Sul (RS)	35	20,153	2077 (378; 4001)	2223 (20,531; 24,154)	90.66 (83.44; 98.16)
Santa Catarina (SC)	20	9422	832 (107; 1640)	10,254 (9529; 11,062)	91.89 (85.17; 98.88)
Brasil (BR)	557	305,758	45,196 (13,299; 79,310)	350,954 (319,057; 385,068)	87.12 (79.40; 95.83)

Table 2: **Observed cases, posterior mean of predicted number of cases and coverage of gestational syphilis for the 27 Brazilian states, 2007–2018. The 90% highest posterior density interval (90%-HPD) is presented for the predicted quantities.**

and microregions: a higher percentage of underreporting of GS cases was found in the North and Northeast regions, while, in general, in the Southeast region, the observed and corrected cases of GS were similar, considering that the lowest percentage of underreported GS cases were found in the microregions of such region. The lowest estimated coverage of GS cases was found in Roraima State (in North region), while the highest was found in São Paulo (in Southeast region). These different patterns highlight that reporting quality may be related to socio-economic and access to health care services.

To the best of our knowledge, this is the first study mapping the underreporting of GS across the country and estimating the real burden of the disease in all Brazilian microregions and states. A small-scale surveillance study in Minas Gerais State found that only 6.5% of identified GS cases were reported in SINAN-Syphilis between 2007 and 2013.[10] In Londrina, Paraná State, a study reported that congenital syphilis cases were 50% higher than GS cases from 2009 to 2015,[11] while Tiago et al. (2017),[9] using a linkage strategy, identified 46% of GS underreporting in a special indigenous sanitary district in Mato Grosso State. Moreover, Souza et al. (2019)[4] noted 13.74% of GS under-registration between 2010 and 2013 in the west of São Paulo, the most populous Brazilian state. Comparison of our results with previous studies is not possible considering their heterogeneity in scale, size, and regional focus, as well as methodological differences in the approaches.

We highlight the role of inequalities in access to health services and GS cases detection. Despite the expansion of prenatal coverage in Brazil in the last decade and improvements in actions to reduce GS, the quality of prenatal services is still low in many Brazilian

cities, particularly in the North and Northeast regions.[13] Previous studies also pointed out that the quality of syphilis surveillance in Brazil is intrinsically related to the adequacy of prenatal care,[31] since it is recommended that the serological test for *T. pallidum* among pregnant women should be requested, at least, at the first prenatal care visit, in the beginning of the third trimester of gestational period, and at the moment of childbirth.[14] Therefore, inequality in healthcare access contributes directly to underreporting of syphilis cases in the country, particularly among women living in vulnerable contexts, in rural and remote areas, or those facing cultural barriers in access to healthcare services.[13,15,16,31,37] Also, the inability of some health professionals to identify syphilis symptoms in the early stage during prenatal care have been associated with underreporting of the disease.

Besides, the access to syphilis tests during the gestational period and the reporting of cases have been directly associated with highest municipal Human Development Index (HDI), most of them located in the Southeast and South region of Brazil.[13,31,32,37] The availability of syphilis rapid test in healthcare service, the timely access to results of diagnostic tests, in addition to postponing the search for test results due to the absence of symptoms in some stages of the infection contribute to GS underreporting. Therefore, the highest quality of prenatal care services is closely related to access to syphilis tests, and could contribute to reduce failures in reporting GS.[13,31,32,37] Aligned with our results, we found that the highest incidence rates of underreported cases of GS occurred in North and Northeast, regions that concentrated the microregions with the highest deprivation levels, and with less coverage and access to diagnostic test and prenatal care, factors previously associated with underreporting and the GS incidence rate in Brazil.[31,32,37]

The increasing incidence of GS represents a growing concern for health professionals and public health in Brazil,[4,14] and failure to correctly diagnose syphilis during pregnancy contributes to non-treatment, since the disease could be asymptomatic, or lead to inadequate adherence to treatment, particularly when the diagnosis occurs at the end of the gestational period. Additionally, stigma and fear of discrimination may discourage pregnant individuals from seeking care for treatment, further contributing to the inadequate management of syphilis during pregnancy, which increases the occurrence of adverse health outcomes for pregnant individuals. Also, failures in the diagnosis of GS could increase the vertical transmission of syphilis, as well as it enhances the chance of adverse effects on the newborn.[6,38] Moreover, it is important to highlight that syphilis should be considered during pregnancy as a fetal emergency due to the high proportion of early or late fetal deaths related to lack of treatment.[14,39] Thus, underreporting of GS can occult the real epidemiological profile of the disease in the country and reduce the possibility of achieving the goals of the World Health Organization regarding GS and congenital syphilis.[40] Besides, it is important highlighting that the variation of GS reporting coverage identified across Brazilian microregions and states expresses differences in capabilities and human resources of health systems, particularly related to prenatal and childbirth care across the country. This scenario creates the urgency of actions and programs for each region to improve GS diagnosis and registration of cases in the SINAN-Syphilis.

Strengths and limitations

The study is a first attempt to estimate the underreporting of GS in Brazil and correct incidence rates for the whole country, which could help to address the country's gap of knowledge regarding GS notification. We provide robust subsidies to Brazilian health authorities to better increase surveillance for GS across the country, highlighting municipalities where improvements in surveillance and case reporting should be prioritized. Our results could allow policymakers and stakeholders to improve policy and action to reduce GS, and, therefore, contribute to reducing mother-to-child transmission and the adverse effects on birth outcomes related to congenital syphilis. Also, it could be used by the Brazilian Ministry of Health in planning the *Previne Brasil*, a public policy that will transfer financial resources to municipalities based on their health indicators, such as the proportion of pregnant with at least six prenatal care visits and tested for syphilis.[41]

Despite the relevance of our results, the study presented some limitations. First, the percentages of underreported cases were calculated for the entire study period, not allowing for analysis of possible improvements or setbacks in the quality of information over time. Second, the model imposes the same reporting probability for areas within the same data quality cluster. However, we found this is the best methodology to be applied, given the restricted type of prior information we have available on underreporting of GS in Brazil.

Conclusion

Brazilian microregions and states in the North and Northeast regions presented the highest values of underreporting of GS, which directly affects the knowledge of the epidemiological profile and the estimation of the social, economic, and health burdens associated with the disease. Conversely, microregions and states in the Southern region presented the lowest values of underreported cases. This scenario highlights the regional disparities in the incidence rate of GS and the quality of prenatal care in the Brazilian context, as well as emphasizes the need to improve the quality of information registered in SINAN-Syphilis, considering the importance of disease surveillance across the

country. Furthermore, strategies to prevent, control, and maintain the surveillance of GS based only on the data made available in the national health information system without correcting for underreported cases can lead to misguided actions and health policies, particularly related to health finance and the distribution of human resources across the regions.

This study provides robust evidence that can be used to enhance national surveillance system and particularly specific policies for GS detection, early treatment and disease control, based in prior information and using a statistical methodology that could be applied by health manager of many countries.

Contributors

G.L.O., M.Y.I. and A.J.F.F. conceptualized this study and contributed to its design.

G.L.O. conducted the statistical analysis.

G.L.O., M.Y.I., A.J.F.F., C.A.S.S.T., R.A., R.F., R.L., M.P., E.S.P., and A.M.C. contributed to the result analysis and interpretation.

G.L.O., M.Y.I., A.J.F.F., C.A.S.S.T., I.O.S and M.A.S performed literature review.

G.L.O. and A.J.F.F. formatted tables, figures and wrote the first draft.

M.Y.I. and M.L.B acquired the data, funding and coordinated the study.

G.L.O., A.J.F.F., C.A.S.S.T. and M.Y.I. have accessed and verified the underlying data.

G.L.O., A.J.F.F., E.S.P., M.Y.I. and M.L.B were responsible for the decision to submit the manuscript.

All authors have access to all the data reported in the study, wrote the manuscript and approved the final version.

Data sharing statement

The GS data are publicly available from the Brazilian National Notifiable Diseases Information System (SINAN) at http://portalsinan.saude.gov.br/sinan-net. Additionally, all data collected for the study and posterior results are available from the corresponding author upon reasonable request.

Editorial disclaimer

The Lancet Group takes a neutral position with respect to territorial claims in published maps and institutional affiliations.

The abstract translated to Portuguese was submitted by the authors and we reproduce it as supplied. It has not been peer reviewed. Our editorial processes have only been applied to the original abstract in English, which should serve as a reference for this manuscript.

Declaration of interests

We declare no competing interests.

Acknowledgements

This work was supported by a grant from the National Council for Scientific and Technological Development of Brazil, the Bill and Melinda Gates Foundation, and the Wellcome Trust.

Funding: The Syphilis Project receives funding from CNPq/Brazil (Grant number 442873/2019-0). CIDACS/Fiocruz-Bahia is supported by grants from CNPq/MS/The Gates Foundation (401739/2015-5) and the Wellcome Trust (202912/Z/16/Z). E.P.S. is funded by the Wellcome Trust (213589/Z/18/Z). The funders had no role in the study design, data collection and analysis, preparation of this manuscript, interpretation of results, or the decision to submit the manuscript for publication.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.lana.2023.100564.

References

1 PAHO. *Global progress report on HIV, viral hepatitis and sexually transmitted infections, 2021*; 2021. https://www.who.int/publications-detail-redirect/9789240027077. Accessed September 6, 2022.
2 Du M, Yan W, Jing W, et al. Increasing incidence rates of sexually transmitted infections from 2010 to 2019: an analysis of temporal trends by geographical regions and age groups from the 2019 Global Burden of Disease Study. *BMC Infect Dis*. 2022;22:574. https://doi.org/10.1186/s12879-022-07544-7.
3 dos Santos MM, Lopes AKB, Roncalli AG, de Lima KC. Trends of syphilis in Brazil: a growth portrait of the treponemic epidemic. *PLoS One*. 2020;15:e0231029. https://doi.org/10.1371/journal.pone.0231029.
4 de Souza TR, Moreira AD, Matozinhos FP, Lana FCF, Arcêncio RA, de Freitas GL. Tendência temporal de internações por sífilis congênita entre 2008 e 2018, em Minas Gerais. *Rev Eletr Enferm*. 2021;23:1–8. https://doi.org/10.5216/ree.v23.64978.
5 PAHO. *With rising trends of syphilis and congenital syphilis in some countries in the Americas, PAHO calls for reinforcement of public health measures - PAHO/WHO*. Pan American Health Organization; 2022. https://www.paho.org/en/news/5-7-2022-rising-trends-syphilis-and-congenital-syphilis-some-countries-americas-paho-calls. Accessed August 25, 2022.
6 Korenromp EL, Rowley J, Alonso M, et al. Global burden of maternal and congenital syphilis and associated adverse birth outcomes—estimates for 2016 and progress since 2012. *PLoS One*. 2019;14:e0211720. https://doi.org/10.1371/journal.pone.0211720.
7 Silva ÂAO, Leony LM, Souza WVD, et al. Spatiotemporal distribution analysis of syphilis in Brazil: cases of congenital and syphilis in pregnant women from 2001–2017. *PLoS One*. 2022;17:e0275731. https://doi.org/10.1371/journal.pone.0275731.
8 Soares KKS, do Prado TN, Zandonade E, Moreira-Silva SF, Miranda AE. Análise espacial da sífilis em gestantes e sífilis congênita no estado do Espírito Santo, 2011-2018. *Epidemiol Serv Saúde*. 2020;29:e2018193. https://doi.org/10.5123/S1679-49742020000100018.
9 Tiago Z da S, Picoli RP, Graeff SV-B, da Cunha RV, Arantes R. Subnotificação de sífilis em gestantes, congênita e adquirida entre povos indígenas em Mato Grosso do Sul, 2011-2014. *Epidemiol Serv Saúde*. 2017;26:503–512. https://doi.org/10.5123/s1679-49742017000300008.
10 Lafetá KRG, Martelli Júnior H, Silveira MF, Paranaíba LMR. Sífilis materna e congênita, subnotificação e difícil controle. *Rev Bras Epidemiol*. 2016;19:63–74. https://doi.org/10.1590/1980-5497201600010006.
11 Lazarini FM, Barbosa DA. Intervenção educacional na Atenção Básica para prevenção da sífilis congênita. *Rev Latino-Am Enfermagem*. 2017;25:e2845. https://doi.org/10.1590/1518-8345.1612.2845.
12 Belo MM de A, de Oliveira CM, de Barros SC, Maia LT de S, do Bonfim CV. Estimativa da subnotificação dos óbitos por sífilis congênita no Recife, Pernambuco, 2010-2016: relacionamento entre os sistemas de informações sobre mortalidade e de agravos de notificação. *Epidemiol Serv Saúde*. 2021;30:e2020501. https://doi.org/10.1590/s1679-49742021000300009.
13 Saraceni V, Miranda AE. Relação entre a cobertura da Estratégia Saúde da Família e o diagnóstico de sífilis na gestação e sífilis congênita. *Cad Saúde Pública*. 2012;28:490–496. https://doi.org/10.1590/S0102-311X2012000300009.
14 Domingues CSB, Duarte G, Passos MRL, et al. Protocolo Brasileiro para Infecções Sexualmente Transmissíveis 2020: sífilis congênita e criança exposta à sífilis. *Epidemiologia e Serviços de Saúde*. 2021;30:e2020597. https://doi.org/10.1590/s1679-4974202100005.esp1.
15 Benzaken AS, Pereira GFM, da Cunha ARC, de Souza FMA, Saraceni V. Adequacy of prenatal care, diagnosis and treatment of syphilis in pregnancy: a study with open data from Brazilian state capitals. *Cad Saúde Pública*. 2019;36:e00057219. https://doi.org/10.1590/0102-311X00057219.
16 Trivedi S, Taylor M, Kamb ML, Chou D. Evaluating coverage of maternal syphilis screening and treatment within antenatal care to guide service improvements for prevention of congenital syphilis in Countdown 2030 Countries. *J Glob Health*. 2020;10:010504. https://doi.org/10.7189/jogh.10.010504.
17 Instituto Brasileiro de Geografia e Estatistica. *Divisão regional do Brasil em regiões geográficas imediatas e regiões geográficas intermediárias: 2017*; 2017. https://biblioteca.ibge.gov.br/index.php/bibl

ioteca-catalogo?view=detalhes&id=2100600. Accessed September 6, 2022.
18 Secretaria de Estado da Saude de Sao Paulo. *Guia de Bolso para o Manejo da Sifilis em Gestante e Sifilis Congenita. Centro de Referencia e Treinamento DST/Aids-SP. Programa Estadual DST/Aids-SP*; 2016. https://saude.campinas.sp.gov.br/doencas/sifilis/guiadebolsodasifilis_2edicao2016.pdf. Accessed September 6, 2022.
19 Allik M, Ramos D, Agranonik M, et al. Developing a small-area deprivation measure for Brazil: technical report. Brazil https://cidacs.bahia.fiocruz.br/ibp/wp-content/uploads/2020/12/technical-report_compressed.pdf; 2020. Accessed September 6, 2022.
20 Oliveira GL, Ferreira AJF, Santana JG, et al. A completeness indicator of gestational and congenital syphilis information in Brazil. *Rev Saude Publica*. 2023;57. https://doi.org/10.11606/s1518-8787.2023057004789.
21 de Oliveira GL, Loschi RH, Assunção RM. A random-censoring Poisson model for underreported data. *Stat Med*. 2017;36:4873–4892. https://doi.org/10.1002/sim.7456.
22 Schmertmann CP, Gonzaga MR. Bayesian estimation of age-specific mortality and life expectancy for small areas with defective vital records. *Demography*. 2018;55:1363–1388. https://doi.org/10.1007/s13524-018-0695-2.
23 Stoner O, Economou T, Drummond Marques da Silva G. A hierarchical framework for correcting under-reporting in count data. *J Am Stat Assoc*. 2019;114:1481–1492. https://doi.org/10.1080/01621459.2019.1573732.
24 de Oliveira GL, Argiento R, Loschi RH, Assunção RM, Ruggeri F, Branco MD. Bias correction in clustered underreported data. *Bayesian Analysis*. 2022;17:95–126. https://doi.org/10.1214/20-BA1244.
25 Dvorzak M, Wagner H. Sparse Bayesian modelling of under-reported count data. *Stat Model Int J*. 2016;16:24–46. https://doi.org/10.1177/1471082X15588398.
26 de Oliveira GL, Oliveira JF, Pescarini JM, et al. Estimating under-reporting of leprosy in Brazil using a Bayesian approach. *PLoS Neglected Trop Dis*. 2021;15:e0009700. https://doi.org/10.1371/journal.pntd.0009700.
27 R Core Team. *R: a language and environment for statistical computing*; 2021. https://www.R-project.org/. Accessed September 6, 2022.
28 Charrad M, Ghazzali N, Boiteau V, Niknafs A. NbClust: an R package for determining the relevant number of clusters in a data set. *J Stat Software*. 2014;61:1–36. https://doi.org/10.18637/jss.v061.i06.
29 Kassambara A, Mundt F. *Extract and visualize the results of multivariate data analyses*; 2020. https://cran.r-project.org/web/packages/factoextra/readme/README.html. Accessed September 6, 2022.
30 Tibshirani R, Walther G, Hastie T. Estimating the number of clusters in a data set via the gap statistic. *J Roy Stat Soc B*. 2001;63:411–423. https://doi.org/10.1111/1467-9868.00293.
31 Soares MAS, Aquino R. Association between the incidence rates for gestational syphilis and congenital syphilis and prenatal care coverage in the State of Bahia, Brazil. *Cad Saúde Pública*. 2021;37:e00209520. https://doi.org/10.1590/0102-311x00209520.
32 Freitas CHSM, Forte FDS, Galvão MHR, Coelho AA, Roncalli AG, Dias SMF. Inequalities in access to HIV and syphilis tests in prenatal care in Brazil. *Cad Saúde Pública*. 2019;35:e00170918. https://doi.org/10.1590/0102-311X00170918.
33 de Figueiredo DCMM, de Figueiredo AM, de Souza TKB, Tavares G, Vianna RP de T. Relação entre oferta de diagnóstico e tratamento da sífilis na atenção básica sobre a incidência de sífilis gestacional e congênita. *Cad Saúde Pública*. 2020;36:e00074519. https://doi.org/10.1590/0102-311X00074519.
34 Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. *Ann Inst Stat Math*. 1991;43:1–20. https://doi.org/10.1007/BF00116466.
35 Watanabe S. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. http://jmlr.org/papers/v11/watanabe10a.html. Accessed July 14, 2023.
36 Ibrahim JG, Chen M-H, Sinha D. *Bayesian survival analysis*. New York, NY: Springer; 2001. https://doi.org/10.1007/978-1-4757-3447-8.
37 Roncalli AG, Rosendo TMSS, Santos MM, Lopes AKB, Lima KC. Effect of the coverage of rapid tests for syphilis in primary care on the syphilis in pregnancy in Brazil. *Rev Saude Publica*. 2021;55:94. https://doi.org/10.11606/s1518-8787.2021055003264.
38 de Souza JM, Giuffrida R, Ramos APM, Morceli G, Coelho CH, Rodrigues MVP. Mother-to-child transmission and gestational syphilis: spatial-temporal epidemiology and demographics in a Brazilian region. *PLoS Neglected Trop Dis*. 2019;13:e0007122. https://doi.org/10.1371/journal.pntd.0007122.
39 Paixao ES, Ferreira AJ, Dos Santos IO, et al. Mortality in children under 5 years of age with congenital syphilis in Brazil: a nationwide cohort study. *PLoS Med*. 2023;20:e1004209. https://doi.org/10.1371/journal.pmed.1004209.
40 World Health Organization. *Global guidance on criteria and processes for validation: elimination of mother-to-child transmission of HIV and syphilis*. 2nd ed. Geneva: World Health Organization; 2017. https://apps.who.int/iris/handle/10665/259517. Accessed September 6, 2022.
41 Ministerio da Saude. *Portaria no 2.979, de 12 de novembro de 2019. Previne Brasil - modelo de financiamento para a APS*; 2019. http://aps.saude.gov.br/. Accessed September 6, 2022.