medRxiv preprint doi: https://doi.org/10.1101/2022.05.29.22275727; this version posted May 30, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

A completeness indicator of gestational and congenital syphilis information in Brazil

Guilherme Lopes de Oliveira[1,2], Andrêa JF Ferreira[1], José Guilherme Santana[1,3], Raquel Martins Lana[4,5], Andrey Moreira Cardoso[4], Carlos Teles[1,6], Rosemeire L. Fiaccone[1,3,7], Rosana Aquino[1,3], Maria Auxiliadora Santos Soares[1,3], Enny S. Paixão[1,8], Idália Oliveira dos Santos[1,3], Leonardo Salvi[1], Mauricio L. Barreto[1,3], Maria Yury Ichihara[1].

[1] Centre of Data and Knowledge Integration for Health (CIDACS), FIOCRUZ, Salvador, Bahia, Brazil.
[2] Federal Centre of Technological Education of Minas Gerais (CEFET-MG), Belo Horizonte, Minas Gerais, Brazil.
[3] Institute of Collective Health, Federal University of Bahia (UFBA), Salvador, Bahia, Brazil.
[4] National School of Public Health (ENSP), FIOCRUZ, Rio de Janeiro, Rio de Janeiro (RJ), Brazil.
[5] Computational Science Programme (PROCC), FIOCRUZ, Rio de Janeiro, Rio de Janeiro, Brazil.
[6] State University of Feira de Santana, Feira de Santana, Bahia, Brazil.
[7] Institute of Mathematics and Statistics, Federal University of Bahia (UFBA), Salvador, Bahia, Brazil.
[8] London School of Hygiene and Tropical Medicine, London, United Kingdom.

Corresponding author
Guilherme Lopes de Oliveira
Av. Amazonas, 7675, Bairro Nova Gameleira, Belo Horizonte, Minas Gerais, Brazil
Postcode: 30510-000
E-mail: guilhermeoliveira@cefetmg.br

Abstract

Objective: To evaluate the quality of information on gestational syphilis (GS) and congenital syphilis (CS) on the Notifiable Diseases Information System (SINAN-Syphilis Brazil) by compiling and validating completeness indicators between 2007 and 2018. **Materials and methods:** Overall, care and socioeconomic completeness scores were compiled based on selected variables, using weights assigned by specialists. The completeness scores were analysed, considering the region and area of residence, pregnant race/colour, and year of case notification. Pearson's correlation coefficients were used to validate the scores obtained by the weighted average method, compared with the values obtained through principal component analysis (PCA). **Result:** Most of the variables selected presented a good or excellent degree of completeness for GS and CS, except for clinical classification, pregnant woman's level of education, partner's treatment, and child's race/colour, which were classified as poor or very poor. The overall (89.93% versus 89.69%) and socioeconomic (88.71% versus 88.24%) completeness scores for GS and CS, respectively, were classified as regular, while the care score (GS-90, 88%, and CS-90, 72%) was good, although there were improvements over time. Differences in the overall, care and socioeconomic completeness scores according to region, area of residence, and ethnic-racial groups were reported for syphilis notifications. A strong linear correlation (>0.90) was observed between the completeness scores estimated through the weighted average method and PCA. **Conclusion:** Improvements in the completeness of GS and CS notifications have been observed in recent years, highlighting the variables which form the care score, compared to the socioeconomic scores, despite differences between regions, area of residence, and ethnic-racial groups. The weighted average was a viable methodological alternative easily operationalised to estimate data completeness scores, allowing routine monitoring of the completeness of gestational and congenital syphilis records.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.29.22275727; this version posted May 30, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

Introduction

Problems with the quality of data registered on Brazilian Health Information Systems (HIS) remain, although undeniable progress has been made in recent years[1]. Different parameters have been used to evaluate the quality of health information, including completeness[2,3], defined as the degree of variable completion on HIS, a thermometer of the quality of information generated by these systems[3].

The incompleteness of information from HIS impairs the accuracy of records, affecting the reliability of the information, the adequate knowledge of the health-disease process, and the epidemiological patterns of events[4,5]. The capacity to plan appropriate actions to prevent and control disease and disorders within the scope of epidemiological surveillance, integrated into those of care, is reduced[6]. Then, HIS completeness is essential to promote and protect the population's health and the appropriate planning and management of resources, including investments for the continuous improvement of these systems by the Ministry of Health[2,7].

Syphilis is among the diseases of great relevance in public health, and, therefore, we need qualified information for its monitoring. As a secular disease and historically neglected, syphilis is caused by the bacterium *Treponema pallidum* and transmitted mainly sexually and vertically[8,9]. World Health Organization (WHO) estimates that approximately 12 million new syphilis cases occur annually, mainly in low-and-middle-income countries[9]. During pregnancy, labour, and postpartum phases, the mother-baby binomial has been the focus to track maternal syphilis during prenatal care, preventing vertical transmission of syphilis.

In Brazil, approximately 61,441 cases of gestational syphilis (GS) and 22,065 cases of CS, with a GS detection rate of 20.8 and CS incidence rate of 8.9 per 1,000 live births[10], were reported in 2021. CS was only included on the national list of diseases requiring mandatory notification in 1986, GS and acquired syphilis were added in 2005 and 2010, respectively[6,11], being reported through individual notification records (INR) on the Notifiable Diseases Information System (SINAN). Filling INR fields, including those not mandatory, is recommended following confirmation of a syphilis case[6]. Analysing the information generated by GS and CS notification records is fundamental to monitoring diseases indicators and ensuring the improvement of prevention, control, and care, besides indicating strategies to enhance its completion quality.

Few national studies have evaluated the completeness of SINAN-Syphilis information[7,12,13] and the factors that influence its variation, such as region, year of notification, area of residence, and race/colour of the individual, as has been evidenced in other studies with HIS in Brazil[2,7,14]. There is also a lack of valuable indicators which are easily operationalised, which enable health managers and professionals to monitor the completeness of the information registered on HIS. This study aims to evaluate the quality of the information by compiling and validating SINAN-Syphilis completeness indicators in Brazil between 2007 and 2018.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.29.22275727; this version posted May 30, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

Materials and Methods:

This descriptive and analytical study evaluated the degree of completeness of the variables in the GS and CS records available on the Ministry of Health SINAN-Syphilis between 2007 and 2018[6]. We build binary indicators and completeness scores for the overall variables selected and for those classified as socioeconomic or related to care. Completeness degree was classified according to Romero and Cunha[15] as follows: excellent (> 95%), good (90%-95%), regular (80%-90%), poor (50%-80%) and very poor (≤ 50%).

The following variables were selected: **GS** – pregnant woman's age range; area of residence; level of education, race/colour; gestational age; qualitative results of treponemal and non-treponemal tests (VDRL); treatment regime; clinical classification and partner's treatment; and **CS** - age range; child's sex and race/colour; area of residence; mother's level of education and race/colour; qualitative results of VDRL for the child; case evolution; attending prenatal care; treatment regime; time of diagnosis and partner's treatment.

1. Building the completeness indicators and scores

(a) Binary completeness indicator: to evaluate the individual completion of each variable, the proportion of individual notification records with complete information was obtained, considering the indicator of the presence (1) or absence of information, information ignored, or inconsistent (0) in the selected variables.

(b) Completeness scores: to evaluate the completion of a set of variables, the weighted average method was used, based on the values of the binary indicators defined in (a) and weights defined on a scale from 1 to 4, varying from the lowest to the highest importance, respectively. The weights were assigned independently by specialists (epidemiologists and health professionals), considering their conception of the relevance of each variable to analyse epidemiological patterns from the perspective of the prevention and control of GS and CS. Subsequently, these weights were evaluated and agreed upon by the study researchers for their use in the estimates of completeness scores (Table 1). The overall score considers all variables selected, while the socioeconomic and care scores consider the variables classified as socioeconomic and care, respectively. The three scores vary between 0 and 1, and the higher the score value, the better the completeness of the set of variables considered. The completeness scores were analysed according to the year of case notification, region, area of residence, and pregnant woman's race/colour.

2. *Validating the completeness scores*

The validation was based on comparing the scores obtained by the weighted average method with those obtained through principal component analysis (PCA), a multivariate statistical technique used to reduce dimensionality, index generation (scores), and clustering of individuals or variables. The tetrachoric correlation matrix of the binary completeness indicators was used to extract the principal components[16]. The binary indicators of the variables with excellent completeness (> 95%) were not considered in the analysis due to

medRxiv preprint doi: https://doi.org/10.1101/2022.05.29.22275727; this version posted May 30, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

the low variability between the categories, and the "partner's treatment" variable was disregarded in the CS analysis due to the high percentage of data either missing or ignored. We calculated the PCA scores for each record using the regression method[17]. The average of scores associated with the first two principal components extracted in the analysis was considered for the overall score. For socioeconomic and care scores, only the first principal component was considered. Lastly, we analyse the relationship between scores calculated by the weighted average and PCA methods through scatter plots and estimate Pearson's linear correlation coefficient. R software, version 3.6.0, and the psych package were used in the PCA[18].

Ethical Aspects

This study used anonymised administrative health data provided by the Ministry of Health, exempting submission to the Ethics and Research Committee.

Results:

305,891 GS and 197,974 CS cases were notified between 2007 and 2018, registering an increase of 9.1 and 5.3 times, respectively, in notification cases in 2018. Notification of GS was higher in the Southeast (44.8%) and Northeast (21.3%) regions, followed by the South (14.4%), North (10.8%), and Central-West (8.6%). A similar pattern was found for CS. Considering the area of residence of GS notifications, 88.3% occur in the urban area, 7.7% in rural, 0.7% from the peri-urban area and 3.2% of the notifications did not fill out this information. Most CS notifications (89.3%) were from urban areas, 6% from rural and 0.8% from peri-urban areas, and 3.9% of the cases did not include this information. The prevalence of GS cases was concentrated among women who self-declared mixed/brown (48.3%), white (29.3%), black (12.3%), asian descendant (0.9%) and indigenous (0.8%), with information on race/colour being absent in 8.5% of the notifications. For CS, 48.9% of case notifications was mixed/brown, 25.4% white, 5.0% black, 0.4% indigenous, 0.24% asian descendant, and 20.1% of notifications were without any information on race/colour.

Analysis of the completeness of GS records, measured by the binary indicator (Table 1), was classified as excellent (> 95%) for the pregnant woman's age range, area of residence, and gestational age; good (90%-95%) for the pregnant woman's race/colour, qualitative results of treponemal test, VDRL test and treatment regime; poor (50%-80%) for the pregnant woman's level of education and clinical classification; and very poor (<50%) for the partner's treatment information. The CS notification records presented excellent completeness (>95%) for the mother and child's age range variable, area of residence, qualitative results of VDRL test, and clinical evolution; good (90%-95%) for child's sex, prenatal care, and the time of diagnosis; regular (80%-90%) for the mother's race/colour, treponemal test and treatment regime; and poor (50%-80%) for the mother's level of education, child's race/colour, and partner's treatment.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.29.22275727; this version posted May 30, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

Moreover, overall and socioeconomic completeness scores for GS (89.93% and 88.7%) and CS (89.69% and 88.71%) were classified as regular (Table 1), while the care score was good (GS-90.88% and CS-90.72%) over the years.

Between 2007 and 2015 for Brazil (Figure 1), the overall completeness score was evaluated as regular for GS (87.7% in 2007 to 89.9% in 2009) and CS (85.5% in 2007 to 89.9% in 2014) and good from 2016. For both diseases, the socioeconomic score was also classified as regular over the years, with a growing tendency to improve completeness between 2011 and 2017 for GS (87.5% and 89.8%), and between 2007 and 2017 for CS (85.5% and 89.4%). The care score was classified as regular for GS between 2007 and 2012 and good in the following years, while for CS, it remained regular until 2015, with a gradual increase in the completeness scores in the subsequent years. We highlighted that, from 2008, the care score was higher than the socioeconomic score for both diseases. The North and South Brazilian regions presented, over time, improvement in completeness scores for GS and CS, especially for socioeconomic scores (Figure 1). This improvement occurred in the Central-West and Southeast regions for GS, while for CS, it was more evident in the Southeast and Northeast regions.

It is essential to highlight the regional differences observed in the gestational and congenital completeness scores (Figure 2): the overall and socioeconomic scores were classified as good in the North (GS-92.2% and CS-92.6% and GS-93.1% and CS-93.7%, respectively) and South regions (GS-91% and GS-90.5% and CS-92% and CS-91.2%, respectively), and regular in the other regions. The care score for GS was classified as good in all regions (North-91.4%; Central-West-90.5%; Southeast-91.3%; and South-91.5%), except for the Northeast region (89.4%), classified as regular. For CS, the care score was classified as good for the North (91.9%), Southeast (90.8%), and South regions (92.5%), and regular for the Northeast (89.7%) and Central-West (89.1%) regions.

Gestational and congenital syphilis cases presented information for the area of residence in 96.8% and 96.1% of notification records, respectively (Figure 3). It was observed that the overall and care scores classifications were good in the urban (GS-90.4% and 91.1%; CS-90.1% and 91%), rural (GS-91.3% and 90.8%; CS-91.5% and 90.8%) and peri-urban area of residence (GS-90.6% and 90.5%; CS-90.5% and 91.1%). However, the socioeconomic score presented a good quality for GS in rural (92.0%) and peri-urban areas of residence (90.8%) but regular in the urban area (89.5%). For CS, the socioeconomic score was good for the rural area (99.2%) and regular for the other areas of residence (urban-88.9% and peri-urban- 89.7%). However, when information about the area of residence was ignored (GS-3.2% and CS-3.9%), the classification of completeness scores worsened (overall and care scores were poor, while the socioeconomic score was regular).

Considering 91.5% and 79.85% of the GS and CS cases (Figure 4), respectively, for which information on race/colour was filled, we observed that the overall, socioeconomic and care completeness scores for GS were classified as good for all the racial categories (black, mixed/brown, indigenous, asian descendant, and white). For CS, the overall and

medRxiv preprint doi: https://doi.org/10.1101/2022.05.29.22275727; this version posted May 30, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

socioeconomic scores presented good completeness for all the racial categories. In cases where race/colour information was ignored (GS-8.5% and CS-20.15%), all the completeness scores were classified as poor, except the socioeconomic score of GS records, which was regular.

Validation of the overall, socioeconomic and care scores were achieved by comparing them using the weighted average and PCA methods, and the result presented a strong linear correlation (>0.90) in the scatter plots (Figure 5). In the case of PCA, the explanation percentage for the principal components retained in the analysis varied between 47.15% and 63.30% for the six different scores calculated. The result obtained by both methods was similar.

Discussion

The results indicate a good or excellent degree of completeness for most of the SINAN-Syphilis variables included in gestational and congenital syphilis notification records, except for clinical classification, pregnant/mother's level of education, partner's treatment, and child's race/colour, which were classified as poor or very poor. The overall and socioeconomic completeness scores were similar for GS and CS records, both classified as regular, while the care score was classified as good and improved during the period studied. However, differences in scores completeness were verified according to region, area of residence, and ethnic-racial groups. These differences were also observed in studies that evaluated the completeness of information on GS and CS notification cases in Latin American and Caribbean countries[19]. This situation emphasised the need to improve the capacity of countries to collect high-quality data for diseases of interest to public health[7,19].

GS and CS notifications have increased more than 9 and 5 times, respectively, in recent years, possibly due to the expansion of primary health care coverage, in particular the Family Health Strategy, and the Bolsa Familia Program, which conditions the receipt of the benefit to the attendance of pregnant women to prenatal care and children to medical appointments to monitor growth and development[14,20]. Improvement in case detection has not been followed by an increase in the completeness of information, with significant regional differences across Brazilian regions, which corroborate previous results described in the literature for other infectious diseases[12,14,21]. Problems with information completeness of SG and SC records may be related to differences in the level of commitment of health professionals with regards to the adequate registration of case notification records, the reduced number of health professionals in health services to complete a significant number of records, as well as the non-prioritisation of syphilis by health managers[7,12]. Considering that, often, notification is seen as a bureaucratic activity of secondary importance, it is essential to train health professionals on the need to generate quality information for epidemiological surveillance, ensuring the integrated flow of information between the responsible sectors, which can reduce differences at the area of residence and regional level in the quality of the information available in syphilis notification records[5,7,22].

medRxiv preprint doi: https://doi.org/10.1101/2022.05.29.22275727; this version posted May 30, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

The care score completeness for CS was classified as regular for the indigenous group and good for the other racial categories, indicating possible difficulties in the monitoring of indigenous children and, consequently, in the adequate case notifications, in addition to the absence or insufficiency in the training of indigenous health teams[23]. Many factors may contribute to the incompleteness of information, such as the availability of time to complete the INR fields, considering the demands of the service, the existence of linguistic and cultural barriers, and difficulty in entering data onto SINAN by the base-centres and/or Special Indigenous Sanitary Districts, since besides not being notifying units, they do not interact with those in the municipal health system[13,23].

It should be highlighted that the "partner's treatment" variable had a very poor and poor degree of completeness for GS and CS, respectively. We supposed that this occurs because, since September of 2017, the Brazilian Ministry of Health modified the definition of CS cases by including the criterion "adequate treatment of the pregnant woman" and disregarding the information on the simultaneous "partner's treatment". Therefore, this variable has become a complementary and non-mandatory filling field in INR[24]. Insufficient completeness of the field "concomitant treatment of the partner", influenced by the non-mandatory registration of the information, harms the monitoring of syphilis cases, mainly CS, which reduces the possibility to assess the real transmission scenario of syphilis and the quality of prenatal care[12,25,26]. Studies indicate that the incompleteness of SINAN information, particularly for CS, and the high underreporting of cases, limit the use of this system in studies of incidence and factors related to the growth of syphilis in Brazil[7,27,28].

Moreover, many variables present in INR are considered important markers for the monitoring and epidemiological evolution of GS and CS in the country (i.e., partner's treatment for CS) and support the adoption of therapeutic conduct of the disease (i.e., clinical classification and time of diagnosis of the disease). Other fields, such as the mother's level of education and race/colour of the mother or child, are used as markers of social and racial inequalities related to infection and/or associated negative outcomes[4,7,28], which highlighted the importance to improve quality data in SINAN-Syphilis. Also, we should highlight that the incompleteness of information related to education level had been observed in several HIS throughout the national territory, limiting its use in research[29].

Data incompleteness directly reflects on the quality of information by not demonstrating the reality of the epidemiological profile of syphilis cases notified in the country[30]. This scenario complicates the planning of strategies and actions to prevent and control syphilis, contributes to the inefficiency of public policies to tackle the disease, and interferes in achieving the goals of eliminating and eradicating syphilis cases proposed by the WHO[9,14,30]. This situation becomes more complex when we consider regional, residential, and ethnic-racial variations in the information completeness scores in a country of continental dimensions, such as Brazil. Therefore, it is suggested that future studies analyse the factors associated with the incompleteness of data and its variations among different groups and regions, which will allow the adoption of strategies to reduce differentials in the quality of SINAN-Syphilis information.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.29.22275727; this version posted May 30, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

Our results indicate that there has been an improvement in the degree of completeness of SINAN-Syphilis information during the study period, despite variations observed. We present a proposal to evaluate the quality of SINAN-Syphilis information, which is still little explored in the national scenario, building scores based on the weighted average of completeness indicators that are easily operationalised. There are some limitations related to the definition of weights for the weighted average method, based on the experience of specialists and not through the correlation structure present in the data obtained by the PCA. However, weighting is commonly used to compile health indicators and, in our case, produces results consistent with those obtained via PCA. This study only accessed the completeness of information, one of the parameters used to evaluate the quality of the information, not considering cases of underreporting or agreement of information, which are also important parameters to be considered when we evaluate the quality of HIS[3].

Conclusion

There have been improvements in the completeness of GS and CS notifications in recent years, mainly with the care score variables, compared to the socioeconomic variables, although with differences between Brazilian regions, ethnic-racial groups, and area of residence for both diseases. However, essential information to describe the epidemiological profile and monitor GS and CS still have poor completeness, compromising the planning of epidemiological surveillance and health care actions. We recommended promoting health teams' training and adopting strategies by health managers, aiming at improvements in the complete filling of INR, given the prioritisation of other demands in health services. Lastly, health professionals and managers can incorporate indicators of easy compilation and interpretability for monitoring the completeness of SINAN-syphilis data, as proposed in this study, into the routine of evaluating the quality of information aiming at its improvement.

Contributions

GLO, JSG, AJFF, and MYI contributed to the concept and design of the study. GLO and JSG prepared the data and conducted the manuscript analyses. GLO, JGS, AJFF, and MYI prepared the preliminary versions of the manuscript and interpreted the study data. RML, AMC, CT, MASS, RLF, ESP, RA, IOS, LS, and MLB contributed to data interpretation and critically reviewed the manuscript. All the authors approved the final version and are responsible for all aspects of the study, including guaranteeing its accuracy and integrity.

Funding

Cidacs/Fiocruz-Bahia is supported by grants from CNPq/MS/The Gates Foundation (401739/2015-5) and the Wellcome Trust (202912/Z/16/Z). EPS is funded by the Wellcome Trust (213589/Z/18/Z). The Syphilis project receives funding from CNPq/Brazil (Grant number 442873/2019-0).

medRxiv preprint doi: https://doi.org/10.1101/2022.05.29.22275727; this version posted May 30, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

References

1 Laguardia J, Domingues CMA, Carvalho C, Lauerman CR, Macário E, Glatt R. Sistema de informação de agravos de notificação em saúde (Sinan): desafios no desenvolvimento de um sistema de informação em saúde. *Epidemiol E Serviços Saúde* 2004; **13**: 135–46.

2 Marques CA, Siqueira MM de, Portugal FB. Avaliação da não completude das notificações compulsórias de dengue registradas por município de pequeno porte no Brasil. *Ciênc Saúde Coletiva* 2020; **25**: 891–900.

3 Lima CR de A, Schramm JM de A, Coeli CM, Silva MEM da. Revisão das dimensões de qualidade dos dados e métodos aplicados na avaliação dos sistemas de informação em saúde. *Cad Saúde Pública* 2009; **25**: 2095–109.

4 Saraceni V, Domingues RMSM, Vellozo V, *et al.* Vigilância da sífilis na gravidez. *Epidemiol E Serviços Saúde* 2007; **16**: 103–11.

5 Correia LO dos S, Padilha BM, Vasconcelos SML. Métodos para avaliar a completitude dos dados dos sistemas de informação em saúde do Brasil: uma revisão sistemática. *Ciênc Saúde Coletiva* 2014; **19**: 4467–78.

6 Ministério da Saúde. Sistema de Informação de Agravos de Notificação–SINAN: normas e rotinas. 2006.

7 Saraceni V, Pereira GFM, da Silveira MF, Araujo MAL, Miranda AE. Vigilância epidemiológica da transmissão vertical da sífilis: dados de seis unidades federativas no Brasil. *Rev Panam Salud Pública* 2017; **41**: 1.

8 Hook III EW. Syphilis. *The Lancet* 2017; **389**: 1550–7.

9 World Health Organization. Report on global sexually transmitted infection surveillance. 2018.

10 Brasil M da S. Boletim Epidemiológico de Sífilis 2021. Departamento de Doenças de Condições Crônicas e Infecções Sexualmente Transmissíveis. 2022. http://www.aids.gov.br/pt-br/pub/2021/boletim-epidemiologico-de-sifilis-2021 (accessed Jan 31, 2022).

11 Azevedo Dantas L, Jerônimo SHN de M, Teixeira GA, *et al.* Perfil epidemiológico de sífilis adquirida diagnosticada e notificada em hospital universitário materno infantil. *Enferm Glob* 2017; **16**: 217–45.

12 Soares MAS, Aquino R. Completude e caracterização dos registros de sífilis gestacional e congênita na Bahia, 2007-2017. *Epidemiol E Serviços Saúde* 2021; **30**. DOI:10.1590/S1679-49742021000400018.

13 Braz RM, Oliveira P de TR de, Reis AT dos, Machado NM da S. Avaliação da completude da variável raça/cor nos sistemas nacionais de informação em saúde para aferição da equidade étnico-racial em indicadores usados pelo Índice de Desempenho do Sistema Único de Saúde. *Saúde Em Debate* 2013; **37**: 554–62.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.29.22275727; this version posted May 30, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

14 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**: 490–6.

15 Romero DE, Cunha CB da. Avaliação da qualidade das variáveis sócio-econômicas e demográficas dos óbitos de crianças menores de um ano registrados no Sistema de Informações sobre Mortalidade do Brasil (1996/2001). *Cad Saúde Pública* 2006; **22**: 673–81.

16 Kirk DB. On the numerical approximation of the bivariate normal (tetrachoric) correlation coefficient. *Psychometrika* 1973; **38**: 259–68.

17 MINGOTI SA. Análise de Dados Através de Métodos de Estatística Multivariada. Uma Abordagem Aplicada, 1ª. UFMG, 2007.

18 Foundation for Statistical Computing. R: The R Project for Statistical Computing. 2021. https://www.r-project.org/ (accessed July 11, 2021).

19 Serruya SJ, Duran P, Martinez G, *et al.* Maternal and congenital syphilis in selected Latin America and Caribbean countries: a multi-country analysis using data from the Perinatal Information System. *Sex Health* 2015; **12**: 164–9.

20 Baroni L, Alves RFS, Boccolini CS, *et al.* Database on the coverage of the “Bolsa-Família” conditioning cash-transfer program: Brazil, 2005 to 2021. *BMC Res Notes* 2021; **14**: 435.

21 Conceição HN da, Câmara JT, Pereira BM. Análise epidemiológica e espacial dos casos de sífilis gestacional e congênita. *Saúde Em Debate* 2020; **43**: 1145–58.

22 Muguande OF, Ferraz ML, França E, Gontijo ED. Avaliação da qualidade do Sistema de Vigilância Epidemiológica de Doença de Chagas Aguda em Minas Gerais, 2005-2008. *Epidemiol E Serviços Saúde* 2011; **20**: 317–25.

23 Tiago Z da S, Picoli RP, Graeff SV-B, *et al.* Subnotificação de sífilis em gestantes, congênita e adquirida entre povos indígenas em Mato Grosso do Sul, 2011-2014*. *Epidemiol E Serviços Saúde* 2017; **26**: 503–12.

24 Ministério da Saúde. Departamento de Doenças de Condições Crônicas e Infecções Sexualmente Transmissíveis. Nota Informativa nº 02-SEI/2017 - DCCI/SVS/MS |. 2017. http://www.aids.gov.br/pt-br/legislacao/nota-informativa-no-02-sei2017-diahvsvsms (accessed Oct 18, 2021).

25 World Health Organization. The global elimination of congenital syphilis : rationale and strategy for action. *Lélimination Mond Syph Congénit Raison Dêtre Strat* 2007; : 38.

26 Ministério da Saúde. Manual de Normas e Rotinas do SINAN. 2007.

27 Heringer R. Desigualdades raciais no Brasil: síntese de indicadores e desafios no campo das políticas públicas. *Cad Saúde Pública* 2002; **18**: S57–65.

28 Benzaken AS, Pereira GFM, Cunha ARC da, Souza FMA de, Saraceni V. Adequacy of prenatal care, diagnosis and treatment of syphilis in pregnancy: a study with open data from Brazilian state capitals. *Cad Saude Publica* 2019; **36**: e00057219–e00057219.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.29.22275727; this version posted May 30, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

29 Lima DJM, Chagas ACMA, Mendes IC, Oriá MOB, Aquino P de S, Pinheiro AKB. Completude e consistência dos dados de gestantes HIV positivas notificadas [Completeness and consistency of data on notified HIV-positive pregnant women]. *Rev Enferm UERJ* 2014; **22**: 321–6.

30 Komka, MR, Lago EG. Sífilis congênita: notificação e realidade. 2007. https://pesquisa.bvsalud.org/portal/resource/pt/lil-492467 (accessed Oct 18, 2021).

medRxiv preprint doi: https://doi.org/10.1101/2022.05.29.22275727; this version posted May 30, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

Tables and Figures:

Table 1. Analysis of the completeness of the overall, socioeconomic and care variables and scores. SINAN-Syphilis, 2007-2018.

Congenital Syphilis			Gestational Syphilis		
Variables	**Completeness n (%)**	**Weights**	**Variables**	**Completeness n (%)**	**Weights**
Socioeconomic					
Pregnant woman's level of education	141,563 (71.51)	4	Pregnant woman's level of education	218,457 (71.44)	4
Pregnant woman's race/colour	177,000 (89.41)	4	Pregnant woman's race/colour	279,863 (91.51)	4
Pregnant woman's age range	191,947 (96.96)	3	Area of residence	295,881 (96.75)	3
Area of residence	190,281 (96.11)	3	Pregnant woman's age range	305,794 (99.99)	3
Child's race/colour,	158,055 (79.84)	2	-	-	-
Child's age range	197,922 (99.97)	2	-	-	-
Sex	186,368 (94.14)	1	-	-	-
Socioeconomic Score	88.24	-	*Socioeconomic Score*	88.71	-
Care					
Partner's treatment	156,672 (79.14)	4	**Partner's treatment***	**89,004 (29.12)**	-
Treponemal test	172,673 (87.22)	4	Treponemal test	276,013 (90.25)	4
Treatment regime	176,192 (89.00)	4	Treatment regime	290,434 (94.97)	4
Attended prenatal care	185,884 (93.89)	4	Non-treponemal test (VDRL)	290,009 (94.83)	4
Time of diagnosis	187,688 (94.80)	4	Clinical classification	276,013 (72.10)	3
Non-treponemal test (VDRL)	191,435 (96.70)	4	Gestational age	305,213 (99.80)	3
Evolution	188,956 (95.44)	3	-	-	-
Care Score	90.72	-	*Care Score*	90.88	-
Overall Score	89.69	-	*Overall Score*	89.93	-

* Partner's treatment was not used to calculate the GS scores.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.29.22275727; this version posted May 30, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

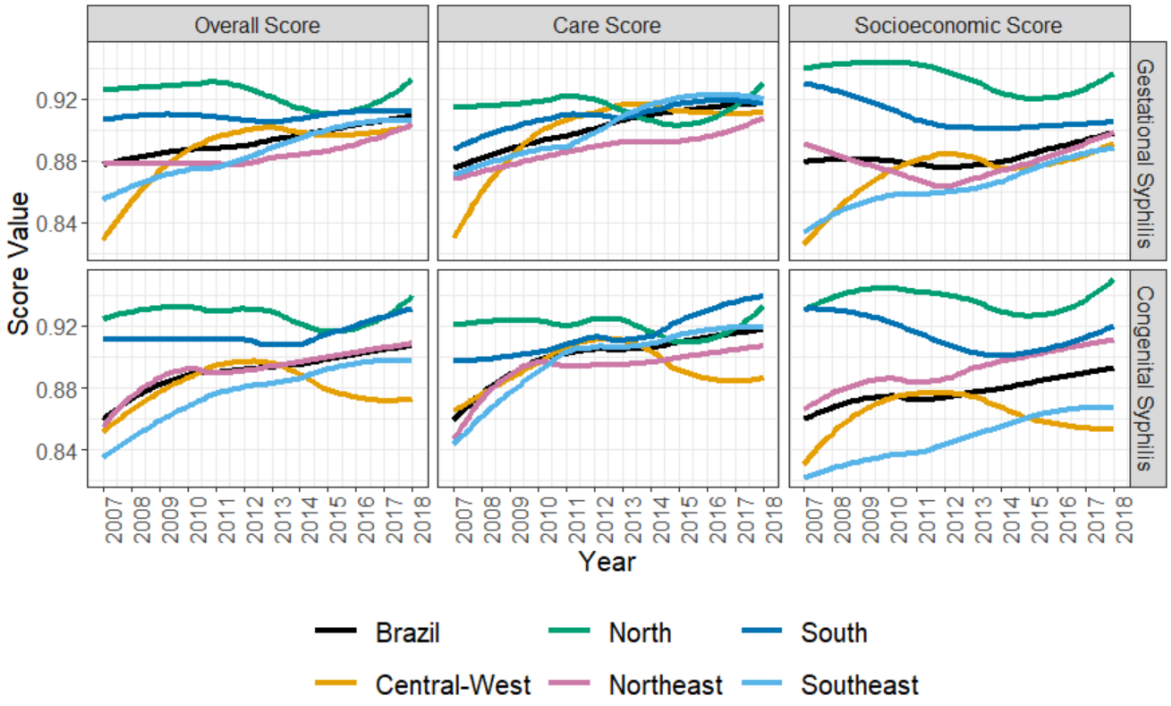

Figure 1. Analysis of completeness scores according to the year of case notifications and Brazilian regions SINAN-Syphilis, 2007-2018.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.29.22275727; this version posted May 30, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

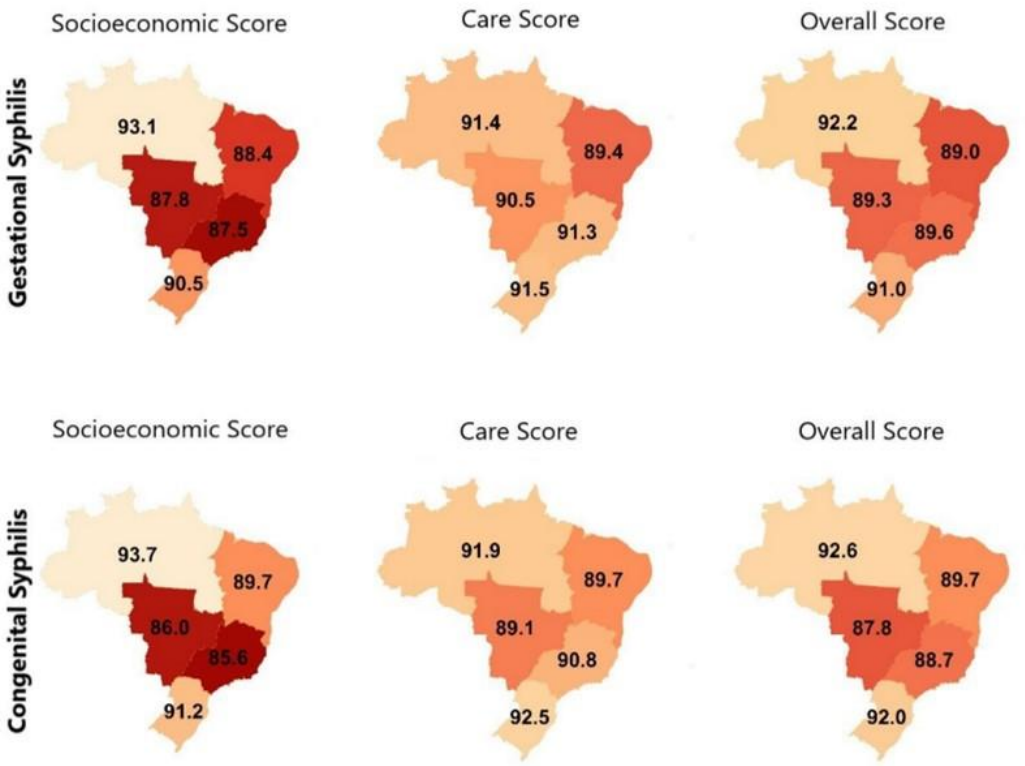

Figure 2. Analysis of completeness scores according to the region of residence for gestational and congenital syphilis, SINAN-Syphilis, 2007-2018.

Legend:

The numerical values correspond to the score value observed in each region.

The darker scores correspond to the lowest completeness scores and the lightest scores to the highest ones.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.29.22275727; this version posted May 30, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

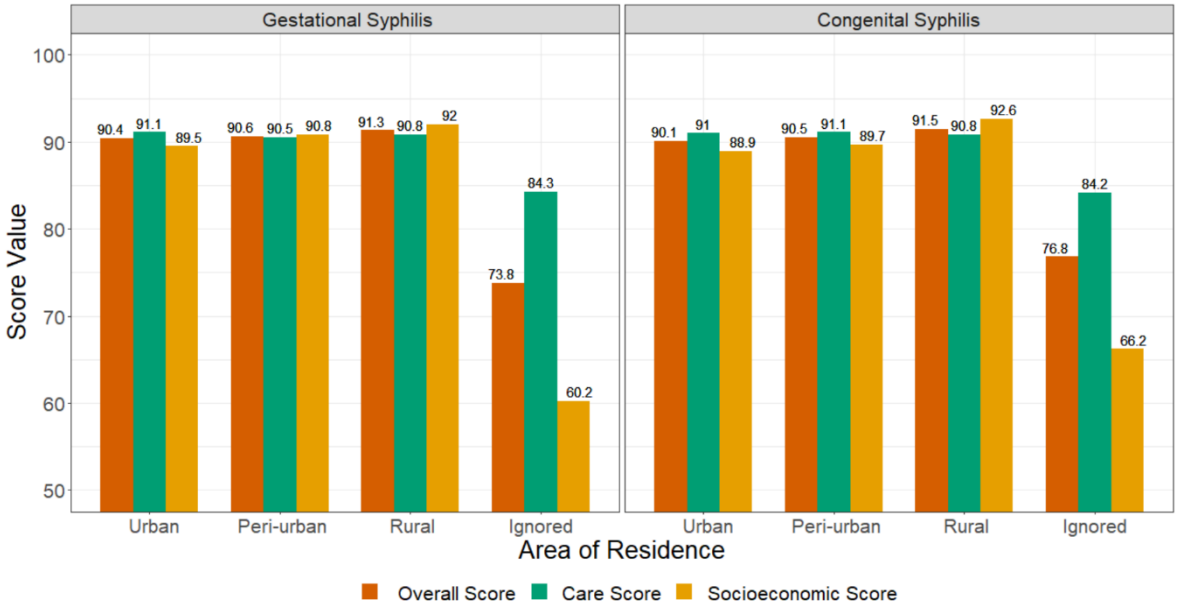

Figure 3. Analysis of completeness scores according to the area of residence of pregnant women for gestational and congenital syphilis, SINAN-Syphilis, 2007-2018.

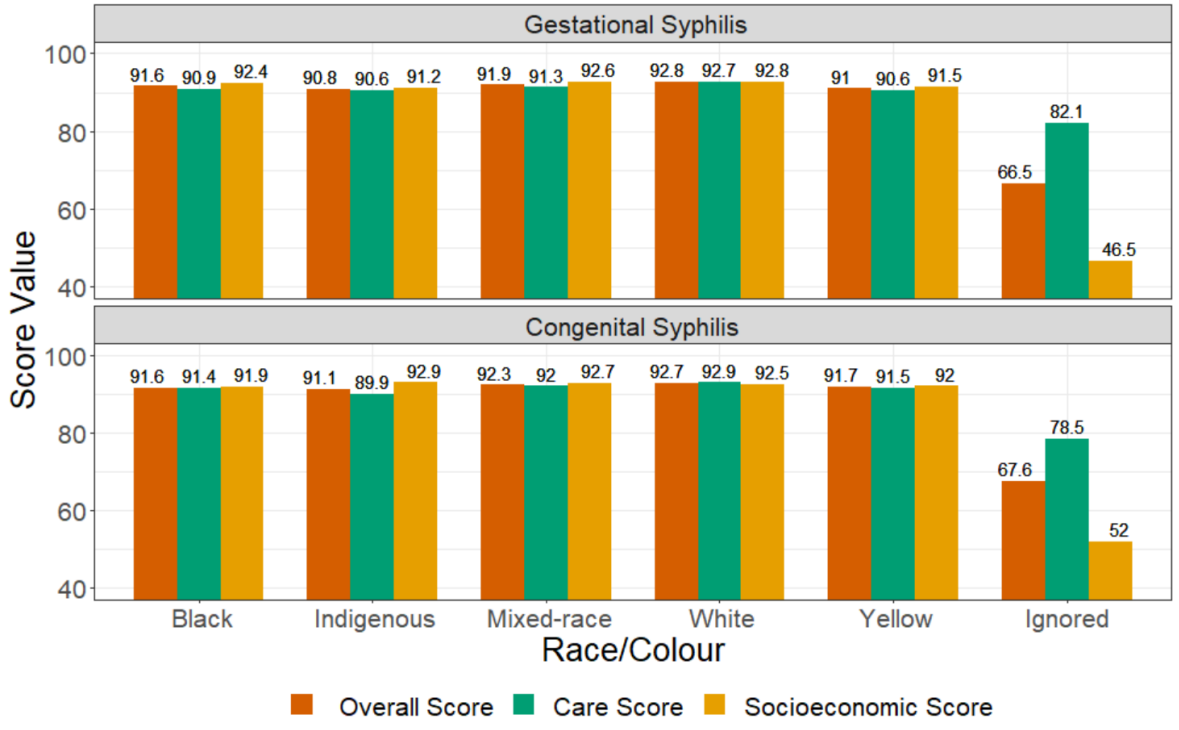

Figure 4. Analysis of completeness scores according to the pregnant woman's race/colour for gestational and congenital syphilis, SINAN-Syphilis, 2007-2018.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.29.22275727; this version posted May 30, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

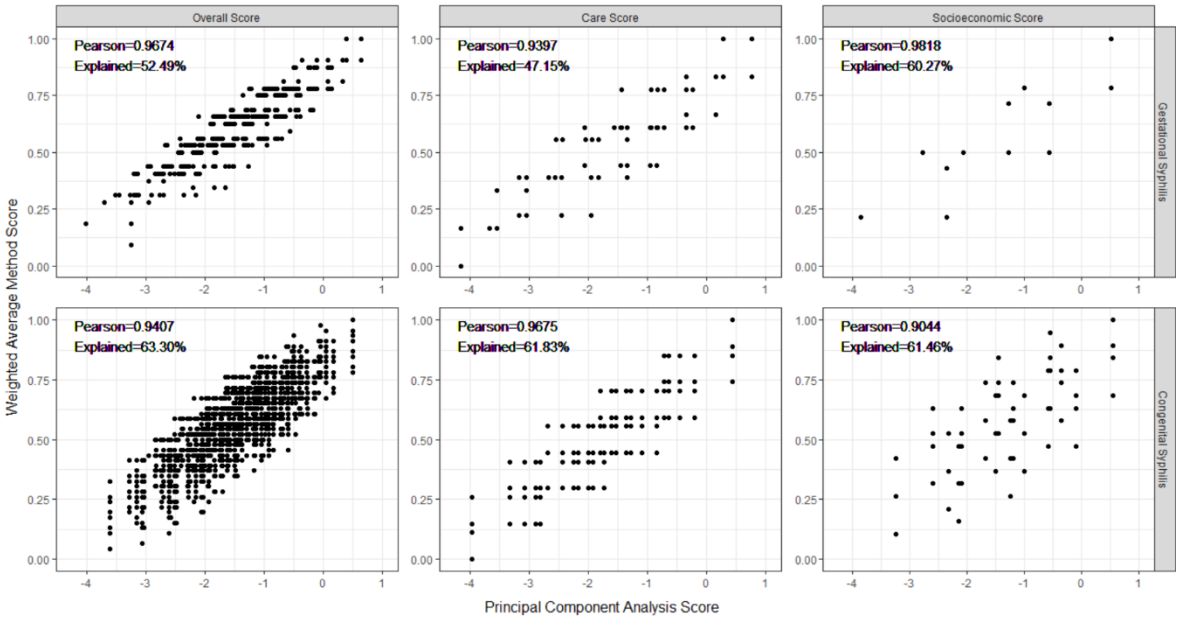

Figure 5. Comparison between the completeness scores obtained through the weighted average method and those obtained through principal component analysis, SINAN-Syphilis, 2007-2018.

Legend:

We have the GS comparisons on the upper line and the CS results on the lower line. The comparative graphs for the socioeconomic scores are presented in the left-hand column; the comparison of the care scores in the central column and the overall scores are compared in the right-hand column. The graphs present Pearson's linear correlation coefficient (Pearson's R). The explanation percentage obtained in the PCA is presented on the horizontal axis of the graphs.

Supplementary material

Table S1. Analysis of the completeness percentage and scores according to the year, region, area of residence, and race/colour of the pregnant woman, considering gestational and congenital syphilis, SINAN-Syphilis, 2007-2018.