



Novel stepwise approach to assess representativeness of a large multicenter observational cohort of tuberculosis patients: The example of RePORT Brazil



María B. Arriaga^{a,b,c}, Gustavo Amorim^d, Artur T.L. Queiroz^{b,e}, Moreno M.S. Rodrigues^f, Mariana Araújo-Pereira^{a,b,c}, Betania M.F. Nogueira^{b,c,g,h,i}, Alexandra Brito Souza^{j,k}, Michael S. Rocha^{b,g}, Aline Benjamin^l, Adriana S.R. Moreira^m, Jamile G. de Oliveiraⁿ, Marina C. Figueiredo^o, Megan M. Turner^o, Kleydson Alves^p, Betina Durovni^{m,q}, José R. Lapa-e-Silva^m, Afrânio L. Kritski^m, Solange Cavalcante^{l,n}, Valeria C. Rolla^l, Marcelo Cordeiro-Santos^{j,k}, Timothy R. Sterling^{o,1}, Bruno B. Andrade^{a,b,c,h,o,r,s,1,*}, for the RePORT Brazil consortium

^a Laboratório de Inflamação e Biomarcadores, Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Salvador, Brazil

^b Multinational Organization Network Sponsoring Translational and Epidemiological Research (MONSTER) Initiative, Salvador, Brazil

^c Faculdade de Medicina, Universidade Federal da Bahia, Salvador, Brazil

^d Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN, USA

^e Center of Data and Knowledge Integration for Health (CIDACS), Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Salvador, Brazil

^f Laboratório de Análise e Visualização de Dados, Fundação Oswaldo Cruz, Porto Velho, Brazil

^g Instituto Brasileiro para Investigação da Tuberculose, Fundação José Silveira, Salvador, Brazil

^h Curso de Medicina, Centro Universitário Faculdade de Tecnologia e Ciências (UniFTC), Salvador, Brazil

ⁱ Programa de Pós-graduação em Ciências da Saúde, Faculdade de Medicina da Bahia, Universidade Federal da Bahia, Salvador, Bahia, Brazil

^j Fundação Medicina Tropical Doutor Heitor Vieira Dourado, Manaus, Brazil

^k Programa de Pós-Graduação em Medicina Tropical, Universidade do Estado do Amazonas, Manaus, Brazil

^l Instituto Nacional de Infectologia Evandro Chagas, Fiocruz, Rio de Janeiro, Brazil

^m Programa Acadêmico de Tuberculose, Faculdade de Medicina, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

ⁿ Secretaria Municipal de Saúde do Rio de Janeiro (Clínica da Família Rinaldo Delamare), Rocinha, Rio de Janeiro, Brazil

^o Division of Infectious Diseases, Department of Medicine, Vanderbilt University School of Medicine, Nashville, Tennessee, USA

^p Ministério da Saúde, National Tuberculosis Control Program, Brasília, Brazil

^q Secretaria Municipal de Saúde do Rio de Janeiro, Rio de Janeiro, Brazil

^r Universidade Nilton Lins, Manaus, Brazil

^s Curso de Medicina, Universidade Salvador (UNIFACS), Laureate University, Salvador, Brazil

ARTICLE INFO

Article history:

Received 5 October 2020

Received in revised form 28 October 2020

Accepted 8 November 2020

Keywords:

Tuberculosis

Cohort study

Sample representativeness

Epidemiology

Treatment outcome

ABSTRACT

Background: A major goal of tuberculosis (TB) epidemiological studies is to obtain results that can be generalized to the larger population with TB. The ability to extrapolate findings on the determinants of TB treatment outcomes is also important.

Methods: We compared baseline clinical and demographic characteristics and determinants of anti-TB treatment outcomes between persons enrolled in the Regional Prospective Observational Research in Tuberculosis (RePORT)-Brazil cohort between June 2015 and June 2019, and the registry of TB cases reported to the Brazilian National TB Program (Information System for Notifiable Diseases [SINAN]) during the same time period. Multivariable regression models adjusted for the study site were performed using *second-generation* p-values, a novel statistical approach. Associations with unfavorable treatment outcomes were tested for both RePORT-Brazil and SINAN cohorts.

Findings: A total of 1,060 culture-confirmed TB patients were enrolled in RePORT-Brazil and 455,873 TB cases were reported to SINAN. *Second-generation* p-value analyses revealed that the cohorts were strikingly similar with regard to sex, age, use of antiretroviral therapy and positive initial smear sputum

* Corresponding author at: Laboratório de Inflamação e Biomarcadores, Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Rua Waldemar Falcão, 121, Candeal, Salvador, Bahia, 40296-710, Brazil.

E-mail address: bruno.andrade@fiocruz.br (B.B. Andrade).

¹ T.R.S. and B.B.A. equally contributed to the work.

microscopy. However, diabetes, HIV infection, and smoking were more frequently documented in RePORT-Brazil. Illicit drug use, the presence of diabetes, and history of prior TB were associated with unfavorable TB treatment outcomes; illicit drug use was associated with such outcomes in both cohorts. *Conclusions:* There were important similarities in demographic characteristics and determinants of clinical outcomes between the RePORT-Brazil cohort and the Brazilian National registry of TB cases. © 2020 The Author(s). Published by Elsevier Ltd on behalf of International Society for Infectious Diseases. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Tuberculosis (TB) is a major cause of death worldwide, particularly in people living with HIV (PLWH) (World Health Organization, 2019). The World Health Organization (WHO) estimated that in 2018, ten million people developed TB, and of those, 1.5 million died (including 251,000 PLWH) (World Health Organization, 2019). The WHO, through the End TB Strategy, (World Health Organization, 2017) has proposed two essential pillars to improve TB control worldwide: (i) robust, coordinated data systems, prioritizing existing systems where possible, and (ii) support systems that promote research and innovation (World Health Organization, 2017). Routinely collected TB data allow for monitoring and calculating epidemiological and operational indicators, which are important for TB control (World Health Organization, 2018). However, to be useful and effective, frequent data updates and quality control are essential for accurate data (Agency for Healthcare Research and Quality (US), 2014).

TB research has a vital role in generating knowledge that guides decision-making strategies (World Health Organization, 2017). However, the different components of study design (e.g., eligibility criteria, recruitment and enrollment, and retention) can include confounders and biases, which may affect the representativeness of the study population and limit generalizability of the results (He et al., 2016). This can result in erroneous estimates, and consequently lead to the adoption of inappropriate actions or strategies (Kukull and Ganguli, 2012). Thus, it is of critical importance to evaluate the representativeness of a study population and whether results from the study population reflect what occurs in the larger population (Jaehn et al., 2020; Kukull and Ganguli, 2012).

In Brazil, the National TB Program (NTBP) implemented, in 1993, the collection of detailed demographic and clinical data on all TB patients. Importantly, the notification of all TB cases and reporting of such information is mandatory, and occurs through an electronic system referred as the Notifiable Disease Information System (SINAN) (Ministério da Saúde do Brasil and Secretaria de Vigilância em Saúde, 2020b). This system is critical for monitoring the epidemiology of TB in all Brazilian regions, guiding governmental investments and policy changes to optimize patient care, minimizing the occurrence of drug-resistance, and leading to better TB outcomes and control. SINAN is also a unique source of data for scientists interested to test the generalizability of their research populations and results.

Although national TB datasets are useful, often the amount and detail of clinical and demographic data that can be collected are limited, and patient follow-up is passive rather than active. Prospective observational cohorts with active participant follow-up for TB treatment outcomes are valuable, but their cost usually limits the cohort size. Regional Prospective Observational Research in Tuberculosis (RePORT) Brazil is a multicenter cohort study with enrollment sites in 3 major regions in Brazil (Hamilton et al., 2015; Regional Prospective Observational Research for Tuberculosis, 2020). However, it is unclear whether the RePORT-Brazil cohort reflects TB epidemiology and treatment outcomes nationally. We therefore evaluated the representativeness of RePORT-Brazil by

comparing epidemiological variables and operational indicators with those in the SINAN dataset. In addition, clinical and epidemiological determinants of unfavorable TB treatment outcomes were also compared between the two cohorts. We also assessed whether data and variables collected in RePORT-Brazil could improve the utility of SINAN.

Materials and methods

Overall study design

We compared baseline clinical and sociodemographic characteristics, and the determinants of anti-TB treatment outcomes among persons with TB enrolled into RePORT-Brazil and those reported to SINAN, both between June 2015 and June 2019. The primary hypothesis was that the overall characteristics of TB patients enrolled into RePORT-Brazil and reported to SINAN were similar.

The regional prospective observational research in tuberculosis (RePORT)-Brazil

The description of RePORT sites is presented in Supplementary Methods and in (Regional Prospective Observational Research for Tuberculosis, 2020). Between June 2015 and June 2019, 1060 persons ≥ 18 years old with new or recurrent pulmonary TB (with or without extrapulmonary disease) and culture-positive sputum (Lowenstein-Jensen medium or BD BACTEC MGIT) were enrolled. Epidemiological information was collected during study visits at baseline, during and at the end of treatment, and up to 24 months after enrollment. This included sex, age, self-reported race, weight, height, education level, the consumption of alcohol, use of illicit drugs and smoking, presence of comorbidities, and HIV status. Additionally, glycated hemoglobin testing (HbA1c), chest X-ray, drug susceptibility testing for anti-TB drugs, and CD4 count (if HIV positive) were performed. In RePORT-Brazil, the treatment outcome was recorded at the last study visit (24 months after the start of treatment). For this study, 434 participants had not yet completed their 24-month visit, so the analysis of treatment outcomes did not include these participants. All data were entered into a REDCap database (Harris et al., 2009) and stored at the RePORT-Brazil data coordination center at Vanderbilt University Medical Center. The database was continuously monitored for data quality and completeness and updated as appropriate.

Notifiable diseases information system (SINAN), Brazilian Ministry of Health

SINAN is a system for the notification and investigation of transmissible diseases that has been implemented, supported, and maintained by the Brazilian Ministry of Health (Ministério da Saúde do Brasil and Secretaria de Vigilância em Saúde, 2007). The description of SINAN is presented in Supplementary Methods. Data from 455,873 TB patients were reported to SINAN between 2015 and 2019. TB cases were diagnosed by one or more of the following criteria: (a) clinical and epidemiological factors (presumptive

diagnosis), (b) bacteriology (sputum smear positive) or positive culture (solid or liquid), (c) GeneXpert MTB RIF, (d) chest radiography or (e) in the case of extrapulmonary TB by histopathology (the details regarding the diagnosis are described in the Manual of Recommendations for the Control of TB in Brazil) (Ministério da Saúde do Brasil and Secretaria de Vigilância em Saúde, 2013). After TB diagnosis, the information collected in the evaluation interview and the laboratory results were recorded on a standardized form that included the clinical form of TB, individual characteristics (sex, age, race, education, alcohol consumption, illicit drug use, smoking habits, and associated conditions and diseases), the presence of TB-HIV coinfection and test results, among others (Ministério da Saúde do Brasil and Secretaria de Vigilância em Saúde, 2013). Information regarding the anti-TB treatment outcome was also provided.

Outcome definition

For this study, a favorable treatment outcome was defined as cure or completed treatment and an unfavorable outcome was defined as treatment failure, lost to follow-up, or death during treatment. The definitions for clinical and bacteriological cure, failure, lost to follow-up, and death for both cohorts corresponded with those in the Manual of Recommendations for the Control of TB of Brazil (Ministério da Saúde do Brasil and Secretaria de Vigilância em Saúde, 2013). The definitions of treatment outcomes for RePORT-Brazil and SINAN are described in Supplementary Table 1.

Operational indicators

We also evaluated operational indicators for TB, which monitor the performance of TB Control Programs (Arakawa et al., 2015). The list of indicators is shown in Supplementary Figure 1.

Data analysis

All analyses were prespecified. Categorical variables were compared using a two-sided Pearson's chi-square test (with Yates correction) or the Fisher's two-tailed test in 2×3 or 2×2 tables, respectively. Quantitative variables were compared using the Mann Whitney *U* test. LASSO (Least Absolute Shrinkage and Selection Operator) (Trevor et al., 2009) regression analysis and mixed effects models (Gelman and Hill, 2006) were performed, in RePORT-Brazil and SINAN cohorts, to identify independent associations between clinical characteristics of TB patients and anti-TB treatment unfavorable outcomes. In LASSO regression, the optimal parameters were found by a cross-validation step, which was repeated 100 times to stabilize the results. The mixed effects model included the variable "Brazilian states" as a random effect (to avoid possible selection bias, because RePORT-Brazil participants were also reported to be part of the SINAN dataset). Results from both regressions approaches were presented in terms of point estimates and 95% confidence intervals (95% CI). P-values <0.05 were considered statistically significant. In addition to the p-value, the second-generation p-value (p δ -value) and the delta-gap (Δ) (when applicable) were calculated (Blume et al., 2018) as described in Supplementary Methods and Supplementary Figure 2.

Results

Tuberculosis epidemiology in Brazil and RePORT-Brazil sites

Brazil has 26 states and a federal district, all allocated to five macro-regions. RePORT-Brazil has sites located in the cities of Manaus, Salvador, Rio de Janeiro, and Duque de Caxias (Figure 1A).

Between 2015 and 2019, Sao Paulo was the state with the highest number of notified TB cases, followed by Rio de Janeiro and Rio Grande do Sul; the states of Bahia and Amazonas were in the top 10 (Figure 1B). The top 10 cities in the number of TB cases documented during the same period included all of the cities that had RePORT sites, including Rio de Janeiro (1), Manaus (3), Salvador (4), and Duque de Caxias (10) (Figure 1B). On the evaluation of TB incidence per 100,000 persons in the Brazilian states where RePORT-Brazil has sites, different patterns were detected. Bahia had a trend in TB incidence that was similar to that observed in the whole country, whereas Amazonas had a substantially higher incidence over time, followed by Rio de Janeiro (Figure 1C, left panel). Similarly, Manaus city displayed the highest TB incidence among the cities that contained RePORT-Brazil sites, followed by Rio de Janeiro, Duque de Caxias, and Salvador. (Figure 1C, right panel). Figure 1D compares the number of cases notified to SINAN and the number of cases in RePORT Brazil in each year of the study period. The number of cases enrolled in RePORT-Brazil (1060 TB cases) represented 0.23% of the total registries in SINAN (455,873 TB cases). These analyses indicate that RePORT-Brazil sites are located in regions of Brazil that account for a significant TB burden of the country.

Characteristics of the TB cases in RePORT-Brazil and SINAN

Population characteristics of TB cases in both datasets were then compared (Table 1). In both datasets, the majority of patients were male, and the age distributions were similar (second-generation p-values not significant). The frequencies of self-reported race were somewhat similar between the two cohorts (second-generation p-value indeterminate: p δ -value = 0.5). The RePORT cohort had a higher proportion of literate persons (95.3% as compared to 83% in SINAN; ($p < 0.001$, p δ = 0; and Δ = 14.9). Compared to the national dataset, the RePORT cohort also had a higher proportion of health care workers (4.3% vs. 1.3%), proportion with no comorbidities (82.9% vs. 74.1%), prevalence of diabetes (24% vs. 8%) and HIV infection (21% vs. 13.5%), alcohol consumption (83.9% vs. 19.5%), use of illicit drugs (34.2% vs. 15.6%), smoking (52.4% vs. 24.3%), abnormal chest x-ray (97% vs. 93.3%), presence of extrapulmonary TB (11% vs. 3%) and drug-resistant *Mycobacterium tuberculosis* (17.3% vs. 9.7%), all with $p < 0.001$ and p δ = 0; Table 1).

In both cohorts, the most frequent treatment outcome was cure (69.5% in RePORT and 66.9% in SINAN). Treatment failure was more frequently documented in the RePORT cohort (4.2% vs. 1.9%), whereas patients who were transferred out to health care facilities of higher complexity were more commonly observed in the SINAN dataset (7.3% vs. 1.9%). When all categories of treatment outcomes were compared between the datasets, a statistically significant difference was observed between RePORT and SINAN ($p < 0.001$ and a p δ = 0; Table 1).

Additional comparisons are shown in Figure 2. The analyses of the second-generation p-values detected in the comparisons of the age distribution between patients stratified by sex (p δ = 0.135), diabetes comorbidity (p δ = 0, with a delta gap of 14.4), and HIV status (p δ = 0.245), indicated that there were marginal differences (except for diabetes) in age distributions within each of these subpopulations between the datasets (Figure 2A). Of note, the majority of the TB patients were new cases (treatment naive), with a similar profile between the datasets (p δ = 0.5; Table 1 and Figure 2B). The populations from the two distinct datasets exhibited similar proportions of positive acid-fast bacilli (AFB) in sputum smears (p δ = 0.5 and Figure 2C). In contrast, the proportions of TB cases with abnormal chest radiographs and positive sputum culture results were substantially higher in the RePORT cohort than SINAN (all p δ = 0 and Figure 2C); a positive sputum culture was required for enrollment in RePORT-Brazil. TB patients from

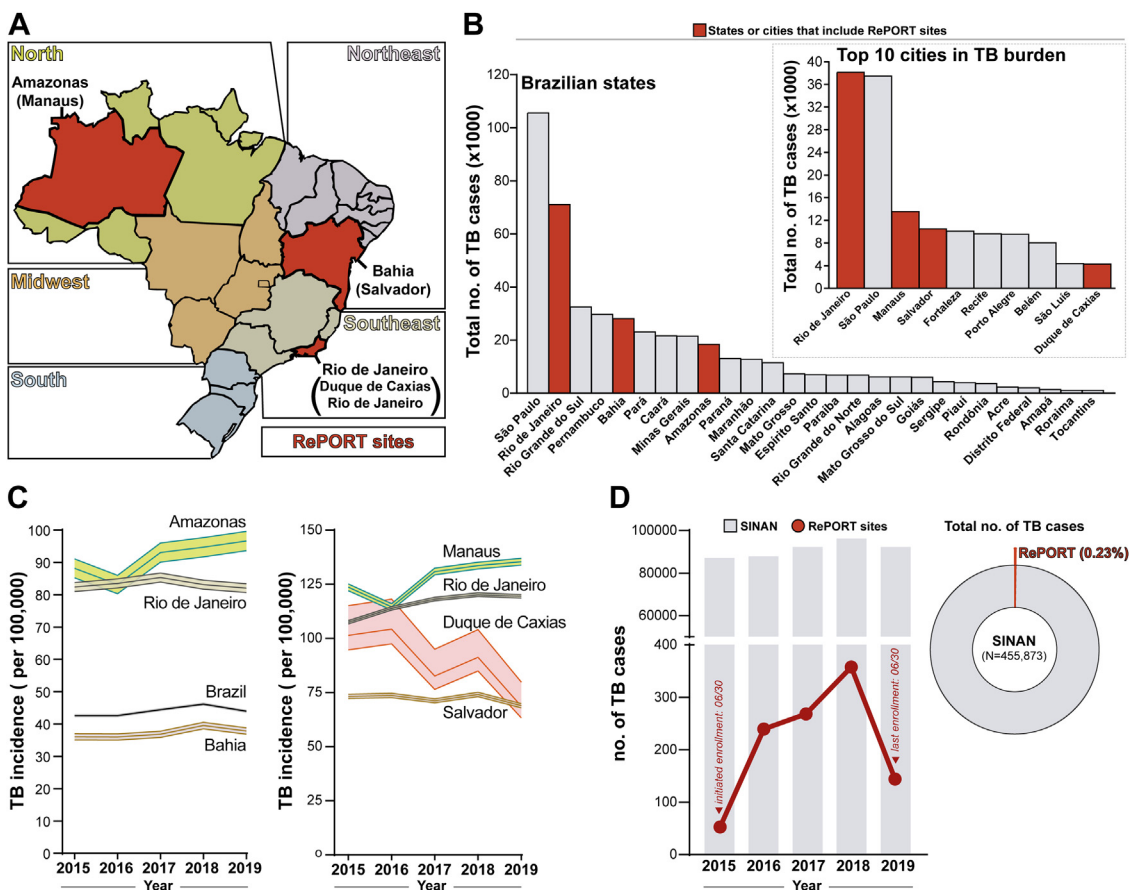


Figure 1. Overview of tuberculosis epidemiology in Brazil (2015–2019). (A) Political map of Brazil shows the different 26 Brazilian states and one federal district colored according to the 5 macroregions (geopolitical subdivision). The states that include RePORT research centers are displayed in red. (B) Total number of tuberculosis cases reported in each Brazilian state, between 2015 and 2019. The states that include RePORT research centers are indicated by red bars. The upper right panel shows the total number of tuberculosis cases between 2015 and 2019 in the top 10 Brazilian cities with the highest number of tuberculosis cases reported within the study period. The cities where RePORT has research centers are highlighted by red bars. (C) Tuberculosis incidence (per 100,000 person-years) for the years 2015–2019, in Brazil the states (left panel) or in the cities (right panel) that host RePORT sites are shown. Data represent incidence and 95% confidence interval. Data were obtained from the National Plan for the End of Tuberculosis as a Public Health Problem (*Plano Nacional pelo Fim da Tuberculose como Problema de Saúde Pública*) (Ministério da Saúde do Brasil, 2017) (D) Total number of TB cases reported by the Brazilian Notification Information System (SINAN; gray bars) and by RePORT sites (red dots and connecting lines), in the study period. Right panel shows a donut pie chart illustrating the percentage that the TB cases recruited in the RePORT Brazil protocol represents among the total number of cases reported in SINAN between 2015 and 2019. Abbreviations: TB: Tuberculosis, RePORT: Regional Prospective Observational Research for Tuberculosis, and SINAN: Sistema de Informação de Agravos de Notificação.

RePORT-Brazil more frequently reported smoking and alcohol consumption and more commonly had diabetes, than patients from SINAN ($p\delta = 0$ and Figure 2D).

Characteristics associated with unfavorable tuberculosis treatment outcomes

In RePORT-Brazil, the patients who experienced an unfavorable treatment outcome were more frequently male (75.4%), *pardo* (57.6%), and had diabetes (31.1%) and/or HIV infection (44%). In addition, alcohol consumption (92.1%), the use of illicit drugs (47.6%), and smoking (63.4%) were also associated ($p < 0.001$ in all) with unfavorable anti-TB treatment outcomes (Supplementary Table 3). Similarly, in the SINAN cohort, male sex (72.3%), *pardo* race (53%), HIV infection (24.9%), alcohol consumption (27.6%), illicit drug use (23.3%), and smoking (30.2%) were all associated with unfavorable outcomes. In addition, increasing age, previous TB (30.5%), and positive sputum culture (69.9%) also characterized individuals who had unfavorable outcomes (Supplementary Table 4). Comparisons of clinical characteristics and treatment outcomes between Report-

Brazil and SINAN patients in the cities of Duque de Caxias, Manaus, Salvador, and Rio de Janeiro are shown in Supplemental Tables 5–8.

To define the specific associations with unfavorable treatment outcomes in each cohort, LASSO regression (model 1) and mixed effects (model 2) were utilized. The LASSO regression model showed that prior TB, diabetes, illicit drug use, and HIV coinfection were all independently associated with the occurrence of unfavorable treatment outcomes in the RePORT-Brazil cohort (Figure 3, left panel), whereas increased age, male sex, alcohol consumption, illicit drug use, prior TB, and HIV coinfection were associated with increased odds of such outcomes in patients from the SINAN cohort (Figure 3, right panel). In the mixed effects model, diabetes comorbidity (OR: 1.81; 95% CI: 1.11–2.94, $p\delta = 0$; and $\Delta = 1.0$) and the use of illicit drugs (OR: 2.28; 95% CI: 1.36–3.82; $p\delta = 0$; and $\Delta = 2.2$) were associated with an unfavorable outcome in RePORT-Brazil (Figure 3, left panel). Likewise, in the SINAN cohort, the use of illicit drugs (OR: 1.85; 95% CI: 1.80–1.91; $p\delta = 0$; and $\Delta = 5.1$) was also associated with an unfavorable result, along with prior TB (OR: 1.96; 95% CI: 1.95–1.97; and $p\delta = 7.5$), HIV positive (OR: 2.81; 95% CI: 2.74–2.90; $p\delta = 0$; and $\Delta = 9.6$), and alcohol consumption (OR: 1.42; 95% CI: 1.38–2.53; $p\delta = 0$; and $\Delta = 2.4$) (Figure 3, right panel).

Table 1
Characteristics of the study participants.

| Characteristics | RePORT-Brazil (n = 1,060) | SINAN (n = 455,873) | p-value | p δ -value |
|--|------------------------------|------------------------|---------|----------------------|
| Sex-no. (%) | | | 0.027 | 0.744 |
| Male | 702 (66.3) | 316693 (69.5) | | |
| Female | 358 (33.7) | 139149 (30.5) | | |
| Age-median (IQR) | 36 (25–49) | 37 (26–52) | 0.001 | 1.0 |
| Race/Ethnicity-no. (%) | | | 0.002 | 0.5 |
| White | 214 (20.2) | 136446 (32.4) | | |
| Black | 271 (25.6) | 58061 (13.8) | | |
| Asian | 6 (0.6) | 3314 (0.8) | | |
| Pardo | 551 (52.1) | 218312 (51.9) | | |
| Indigenous | 16 (1.5) | 4817 (1.1) | | |
| Literate-no. (%) | 1008 (95.3) | 254989 (83.0) | <0.001 | 0 (Δ = 14.9) |
| Health worker-no. (%) | 45 (4.3) | 5572 (1.3) | <0.001 | 0 (Δ = 7.0) |
| Comorbidities ^a -no. (%) | | | <0.001 | 0 (Δ = 0.38) |
| Cancer | 10 (1.0) | 3933 (0.9) | | |
| Chronic Obstructive Pulmonary Disease/Emphysema | 6 (0.6) | 650 (0.1) | | |
| Kidney disease | 5 (0.5) | 328 (0.1) | | |
| Hypertension | 81 (8.2) | 46668 (10.2) | | |
| Others | 66 (6.7) | 66568 (14.6) | | |
| No comorbidity | 817 (82.9) | 337715 (74.1) | | |
| Diabetes-no. (%) | 250 (24.0) | 33961 (8.0) | <0.001 | 0 (Δ = 11.3) |
| HIV infection-no. (%) | 220 (21.0) | 49046 (13.5) | <0.001 | 0 (Δ = 11.3) |
| Antiretroviral therapy (ART) ^b -no. (%) | 193 (87.7) | 16592 (33.8) | <0.001 | 0.5 |
| Alcohol consumption-no. (%) | 889 (83.9) | 82842 (19.5) | <0.001 | 0 (Δ = 31.2) |
| Illicit drug use-no. (%) | 361 (34.2) | 64919 (15.6) | <0.001 | 0 (Δ = 12.5) |
| Smoking-no. (%) | 555 (52.4) | 102030 (24.3) | <0.001 | 0 (Δ = 14.7) |
| Prior TB-no. (%) | 170 (16.2) | 88266 (19.4) | 0.010 | 0.5 |
| Abnormal chest x-ray-no. (%) | 1027 (97.0) | 323940 (93.3) | <0.001 | 0 (Δ = 4.8) |
| Type of TB ^c -no. (%) | | | 0.001 | 0 (Δ = 13.6) |
| Pulmonary | 936 (88.4) | 384709 (84.4) | | |
| Extrapulmonary | 0 (0.0) | 57301 (12.6) | | |
| Pulmonary and Extrapulmonary | 123 (11.6) | 13629 (3.0) | | |
| Positive AFB-no. (%) | 852 (80.8) | 225900 (67.9) | <0.001 | 0.5 |
| Positive culture-no. (%) | 1053 (99.9) | 93668 (69.1) | <0.001 | 0 (Δ = 47.8) |
| Drug-susceptibility testing (DST)-no. (%) | | | <0.001 | 0 (Δ = 14.5) |
| Rifampicin resistance | 5 (0.5) | 674 (1.3) | | |
| Isoniazid resistance | 51 (4.9) | 1774 (3.4) | | |
| Rifampicin-Isoniazid resistance | 26 (2.5) | 1264 (2.4) | | |
| Any drug resistance ^d | 98 (9.4) | 1444 (2.7) | | |
| Sensitive | 859 (82.7) | 47785 (90.3) | | |
| Directly observed treatment (DOT)-no. (%) | 740 (70.3) | 150016 (48.7) | <0.001 | 0 (Δ = 10.7) |
| Treatment Outcome ^e -no. (%) | | | <0.001 | 0 (Δ = 13.6) |
| Cure | 435 (69.5) | 258355 (66.9) | | |
| Failure | 26 (4.2) | 7343 (1.9) | | |
| Relapse | 9 (1.4) | 9039 (2.3) | | |
| Death | 52 (8.3) | 32972 (8.5) | | |
| Lost to follow-up | 92 (14.7) | 50228 (13.0) | | |
| Transferred out | 12 (1.9) | 28198 (7.3) | | |

Table note: Data represent no. (%) or median with interquartile range (IQR). p δ -value: second-generation p-value. Δ = delta-gap. See the Supplemental Figure 2 for the interpretation of the p δ -value. Details of the total data available in the Report-Brazil and SINAN bases in the Supplementary Table 2.

Alcohol consumption: Past or current, any consumption of alcohol. Smoking: Past or current, cigarette smoker. Illicit drug use: Past or current illicit drug use (marijuana, cocaine, heroin, or crack).

Abbreviations: TB: Tuberculosis, SINAN - Sistema de Informação de Agravos de Notificação (Brazilian Notification Information System), AFB: acid fast bacilli, and ART: Antiretroviral therapy.

^a It did not include DM and HIV.

^b ART frequency was calculated among the persons living with HIV.

^c All individuals from the RePORT cohort had a diagnosis of pulmonary tuberculosis, in some cases with presence in other anatomical sites.

^d Any drug (anti-TB) resistance except rifampicin and isoniazid: Pyrazinamide, ethambutol, streptomycin, kanamycin, and ethionamide.

^e In RePORT-Brazil study, the results of anti-TB treatment are recorded at the last study visit (24 months after the start of treatment). By the time the present analyses were performed, 434 participants had not yet completed the last visit, and analyses of treatment outcomes did not include such participants.

Epidemiological and operational indicators of tuberculosis cases in RePORT-Brazil and SINAN

Finally, the epidemiological and operational indicators were compared between the cohorts (Supplementary Figure 1A–L). Among the 11 operational indicators, significant differences were identified between RePORT-Brazil and SINAN, during each year in the study period. The frequency of new cases of pulmonary TB diagnosed through Xpert -MTB/-RIF, smear, culture, or drug susceptibility

testing among the total pulmonary TB cases was lower in each year in the SINAN cohort, likely due to RePORT inclusion criteria ($p < 0.01$, in each year) (Supplementary Figure 1B). In turn, this difference was also reflected by higher proportions of each type of TB screening test compared (Supplementary Figure 1F, and H). The coverage of HIV testing was also different between the datasets, with 70% in SINAN and 95% in RePORT ($p < 0.001$), also likely due to RePORT inclusion criteria. The ART initiation and cases receiving DOT was also significantly higher in RePORT-Brazil.

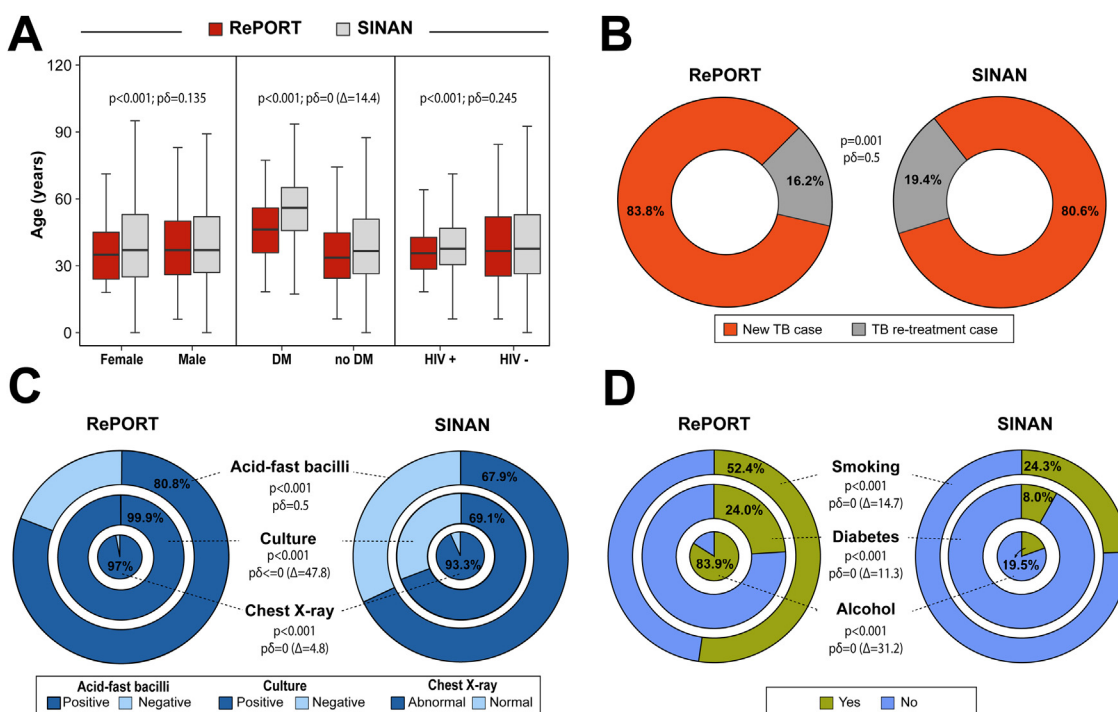


Figure 2. Characteristics of the TB cases in RePORT and SINAN (2015–2019). (A) Comparison of the age distribution (median and interquartile range values) between sex, diabetes condition, and HIV infection among RePORT participants and TB cases reported by SINAN within the study period. (B) The proportion of new and retreatment cases of tuberculosis. (C) Comparison between RePORT and SINAN TB cases (prior to the initiation of anti-TB treatment) in regard to frequency of individuals stratified by acid-fast bacilli (AFB) and *M. tuberculosis* culture results as well as of abnormal chest radiograph presentation. (D) Comparison between RePORT and SINAN TB cases with regard to smoking habit and alcohol consumption (smoking and alcohol: in the past or at the time of evaluation before anti-TB treatment). The proportion of TB-diabetes comorbidity is also shown. See the Supplemental Fig. 1 for the interpretation of the pδ-value. Abbreviations: Δ = delta-gap, pδ-value: second-generation p-value, RePORT: Regional Prospective Observational Research for Tuberculosis, SINAN: Sistema de Informação de Agravos de Notificação, and TB: Tuberculosis.

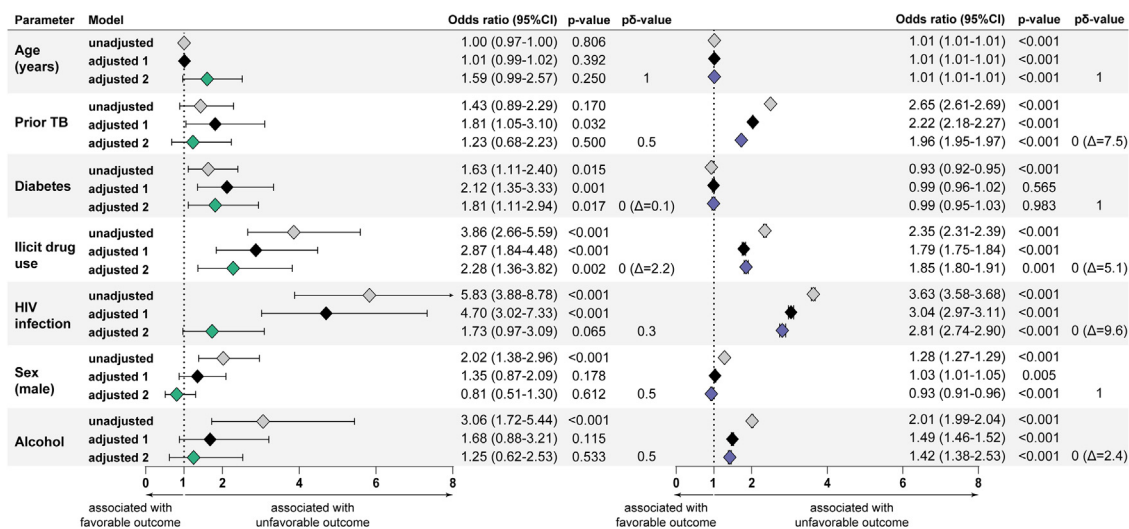


Figure 3. Characteristics associated with unfavorable treatment outcomes in tuberculosis patients. Two logistic regression models (LASSO regression and mixed effects) were performed to evaluate the independent associations of tuberculosis patients and antitubercular treatment unfavorable outcomes (failure, death, and lost to follow-up) in either RePORT (left panel) or SINAN (right panel; where the CI are present, but narrow due to the large sample size). Optimal parameters were found by a cross-validation step, which was repeated 100 times to stabilize the results (Supplementary Table 2 for RePORT and/or Supplementary Table 3 for SINAN) and were included in the adjusted models. Adjusted model 1: LASSO regression using a cross-validation step. Adjusted model 2: The mixed effects model, including the variables “Brazilian states” as a random effect. In addition to the p-value, the second-generation p-value (pδ-value) and the delta-gap (Δ) (when applicable) are shown. See the Supplemental Figure 2 for the interpretation of the pδ-value. Abbreviations: Δ = delta-gap, pδ-value: second-generation p-value, RePORT: Regional Prospective Observational Research for Tuberculosis, SINAN: Sistema de Informação de Agravos de Notificação, and TB: Tuberculosis.

Discussion

In this study, demographic characteristics such as age, sex, and race were similar in both the RePORT-Brazil and SINAN cohorts of TB patients, which support our hypothesis that the RePORT-Brazil cohort is representative of individuals with pulmonary TB in Brazil (Fernandes et al., 2018; Zhang et al., 2011). However, certain characteristics varied in the RePORT-Brazil cohort as compared to the general TB population. Literacy rate was higher among patients in RePORT, possibly because study sites are located in large cities, where access to school and overall literacy are higher than in rural areas (Ministério da Educação, 2008). Alcohol consumption, smoking, and the use of illicit drugs were significantly more commonly reported in RePORT patients. The RePORT study may have a higher detection rate of substance use as it was included in structured questionnaires and validated tools for screening. Conversely, SINAN notification templates include simple yes/no questions to evaluate these characteristics, which may be more prone to reporting or detection bias (Rocha et al., 2020). Moreover, SINAN templates are completed by health care workers who collect this information from the patient or medical charts, which often lack the detailed information needed (Ministério da Saúde do Brasil and Secretaria de Vigilância em Saúde, 2007). These factors may have led to reduced substance use identification in the national database, as noted previously (Rocha et al., 2020; Silva et al., 2010).

The proportions of patients reporting comorbidities such as cancer, chronic obstructive pulmonary disease, kidney disease, and hypertension were similar in both cohorts. The proportion reporting “other” comorbidities was higher in SINAN, and the proportions reporting diabetes or HIV were higher in RePORT-Brazil. This may be due to differences in the ascertainment of these comorbid conditions in the two cohorts (Althubaiti, 2016). Abnormal chest x-rays were reported more often in RePORT cases, possibly that reflect care provided by trained TB specialists, which is not as common in most health care facilities in Brazil (Shazzadur Rahman et al., 2019). The increased frequency of extrapulmonary manifestations of TB may also reflect treatment by experienced TB specialists and greater availability of diagnostic tests at reference centers from RePORT.

Interestingly, our study identified differences in relevant clinical characteristics such as an increased prevalence of DM, HIV, and multidrug resistance in RePORT as compared to SINAN. These disparities may not necessarily reflect differences in the prevalence of these conditions in the study populations, but rather more frequent diagnosis and notification in RePORT because of uniform HbA1c and HIV testing. These comorbidities are known to be more frequent in patients with TB (Saraceni et al., 2018; Tankeu et al., 2017).

Similarly, all participants in RePORT-Brazil were required to have positive *M. tuberculosis* sputum cultures and drug susceptibility testing, in contrast to SINAN, which resulted in a selection bias. When compared to the national data, we found that in 2019 only 76.1% of patients with TB were tested for HIV, despite the fact that the national guidelines recommend HIV testing in all patients with a TB diagnosis and that low-cost HIV testing is available nationally. Testing for DM is not routinely recommended by national guidelines; patients may have blood glucose testing if they exhibit risk factors or classical symptoms of DM (Ministério da Saúde do Brasil and Secretaria de Vigilância em Saúde, 2013). In SINAN, the prevalence of DM is likely only self-reported DM and may therefore underrepresent the actual prevalence.

Regarding the detection of drug-resistant TB (DR-TB), the national statistics also indicate possible underdiagnosis. In 2019, nationally, 37.7% of patients with newly diagnosed TB had Xpert-MTB/-RIF performed, and only 24% had sputum culture performed, which suggests that, except for rifampicin, drug sensitivity was unknown for many cases. Among cases of retreatment, 30.4% had

sputum culture performed and only 50% of those completed the recommended diagnostic workup, including drug sensitivity tests. This reflects the ongoing limited access to Xpert-MTB-RIF technology nationally (Ministério da Saúde do Brasil and Secretaria de Vigilância em Saúde, 2020a; Silva et al., 2018) and the lack of routine solicitation from physicians or capacity to perform sputum cultures and drug sensitivity tests in cases of presumed TB. The comparison of these two datasets indicates that the broad testing strategy adopted in the RePORT-Brazil protocol more accurately detected the presence of important comorbidities, exposures, and drug resistance in TB cases.

RePORT-Brazil and SINAN had a similar proportion of cases with cure, death, and loss to follow-up. Reporting on these outcomes is less difficult than other TB outcomes possibly due to the simplicity of their definitions. Nevertheless, RePORT had a higher proportion of treatment failures and lower frequency of individuals who were transferred out, than SINAN. These discrepancies may be because patients enrolled in RePORT-Brazil were actively followed up for 24 months from treatment initiation. In such scenarios, the treatment outcome is likely obtained in a more reliable fashion, particularly in cases of treatment failure, loss to follow-up, and transferred out. In SINAN, information on outcome is usually recorded at the end of treatment. In public health care centers, it is not recommended that patients return after finishing anti-TB treatment. Indeed, other studies (Baldez do Canto and Borges Nedel, 2020; Rocha et al., 2020; Silva et al., 2010) have described limitations related to the underreporting of treatment outcome results and lack of follow-up for patients who were transferred out or discontinued treatment.

This study had some limitations. RePORT-Brazil had more active follow-up than SINAN, and the ability to follow-up prospectively in the latter cohort was mainly because of the large number of patients. In addition, routine follow-up for clinical care usually does not extend beyond the end of TB treatment. Another limitation was that follow-up for study endpoints in RePORT-Brazil is ongoing. However, for this analysis, we truncated follow-up so that it was similar to SINAN. Nonetheless, the present study adds to current knowledge in the field by demonstrating the close similarities of the RePORT-Brazil and SINAN cohorts regarding patient characteristics and treatment outcomes. This is beneficial for both cohorts. For the smaller RePORT-Brazil cohort, this supports the generalizability of findings from current and future translational studies (Jaehn et al., 2020). For the larger SINAN cohort, it means that smaller, less-expensive population-based studies in RePORT-Brazil may provide important data that can inform TB treatment strategies throughout all of Brazil.

In this study, we found similarities between both databases, though significant differences were identified regarding TB diagnosis, due primarily to RePORT-Brazil inclusion criteria. This study highlights critical gaps in comprehensive TB care and identifies data that could be collected in SINAN that would facilitate improvements in the care and outcomes of TB patients in Brazil (Ministério da Saúde do Brasil and Secretaria de Vigilância em Saúde, 2013).

Summary

We compared a multicenter tuberculosis cohort enrolled from three regions of Brazil with the Brazilian National Tuberculosis Program registry, using innovative statistical approaches. The cohort was representative of TB patients overall, facilitating generalizability of study outcomes.

Ethical approval

All clinical investigations were conducted according to the principles of the Declaration of Helsinki. The RePORT-Brazil

protocol, informed consent, and study documents were approved by the institutional review boards at each study site and at Vanderbilt University Medical Center. Participation in RePORT-Brazil was voluntary, and written informed consent was obtained from all such participants. For the data extracted from SINAN, the anonymity of study subjects was preserved, and all data were de-identified.

Funding

The study was supported by the Intramural Research Program of the Fundação Oswaldo Cruz (B.B.A.), Intramural Research Program of the Fundação José Silveira (B.B.A., M.S.R., B.M.F.N.), Departamento de Ciência e Tecnologia (DECIT)-Secretaria de Ciência e Tecnologia (SCTIE)-Ministério da Saúde (MS), Brazil [25029.000507/2013-07 to V.C.R.], and the National Institutes of Allergy and Infectious Diseases [U01-AI069923 to T.R.S. and U01-AI115940 to B.B.A.]. M.B.A. received a fellowship from the Fundação de Amparo à Pesquisa da Bahia (FAPESB). M.A.-P. received a fellowship from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Finance code: 001). B.B.A., J.R.L.S., and A.K. are senior investigators of the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil.

Disclaimer

The funder of the study had no role in the study design, data collection, or data analysis; however, a representative of the Brazilian Ministry of Health (K.A.) was involved in data interpretation and writing the report.

Conflicts of interest

None declared.

CRediT authorship contribution statement

María B. Arriaga: Data curation, Investigation, Formal analysis, Methodology, Resources, Software, Writing - original draft, Writing - review & editing. **Gustavo Amorim:** Data curation, Investigation, Formal analysis, Methodology, Resources, Software, Writing - review & editing. **Artur T.L. Queiroz:** Data curation, Investigation, Formal analysis, Methodology, Software, Writing - review & editing. **Moreno M.S. Rodrigues:** Data curation, Formal analysis, Software, Writing - review & editing. **Mariana Araújo-Pereira:** Formal analysis, Software, Writing - original draft, Writing - review & editing. **Betania M.F. Nogueira:** Investigation, Writing - original draft, Writing - review & editing. **Alexandra Brito Souza:** Investigation, Writing - review & editing. **Michael S. Rocha:** Investigation, Writing - review & editing. **Aline Benjamin:** Investigation, Writing - review & editing. **Adriana S.R. Moreira:** Investigation, Writing - review & editing. **Jamile G. de Oliveira:** Investigation, Writing - review & editing. **Marina C. Figueiredo:** Conceptualization, Investigation, Formal analysis, Funding acquisition, Project administration, Software, Writing - review & editing. **Megan M. Turner:** Data curation, Writing - review & editing. **Kleydson Alves:** Investigation, Writing - review & editing. **Betina Durovni:** Investigation, Funding acquisition, Writing - review & editing. **José R. Lapa-e-Silva:** Investigation, Funding acquisition, Writing - review & editing. **Afrânio L. Kritski:** Investigation, Funding acquisition, Writing - review & editing. **Solange Cavalcante:** Investigation, Funding acquisition, Writing - review & editing. **Valeria C. Rolla:** Conceptualization, Investigation, Funding acquisition, Writing - review & editing. **Marcelo Cordeiro-Santos:** Conceptualization, Investigation, Funding acquisition, Writing - review & editing. **Timothy R. Sterling:** Conceptualization,

Investigation, Funding acquisition, Project administration, Resources, Software, Supervision, Writing - review & editing. **Bruno B. Andrade:** Conceptualization, Data curation, Investigation, Formal analysis, Funding acquisition, Methodology, Project administration, Resources, Software, Supervision, Writing - original draft, Writing - review & editing.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgments

The authors thank the study participants. We also thank the teams of clinical and laboratory platforms of RePORT Brazil. A special thanks to Elze Leite (FIOCRUZ, Salvador, Brazil), Eduardo Gama (FIOCRUZ, Rio de Janeiro, Brazil), Elcimar Junior (FMT-HVD, Manaus, Brazil), and Hilary Vansell (VUMC, Nashville, USA) for administrative and logistical support.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.ijid.2020.11.140>.

References

- Agency for Healthcare Research and Quality (US). Registries for Evaluating Patient Outcomes. 2014.
- Althubaiti A. Information bias in health research: definition, pitfalls, and adjustment methods. *J Multidiscip Healthc* 2016;9:211–7.
- Baldez do Canto V, Borges Nedel F. Completeness dos registros de tuberculose no Sistema de Informação de Agravos de Notificação (Sinan) em Santa Catarina, Brasil, 2007–2016. *Epidemiol Serv Saúde* 2020;29(3).
- Blume JD, D'Agostino McGowan L, Dupont WD, Greevy Jr. RA. Second-generation p-values: improved rigor, reproducibility, & transparency in statistical analyses. *PLoS One* 2018;13(3):e0188299.
- Fernandes P, Ma Y, Gaeddert M, Tsacogianis T, Marques-Rodrigues P, Fregona G, et al. Sex and age differences in Mycobacterium tuberculosis infection in Brazil. *Epidemiol Infect* 2018;146(12):1503–10.
- Gelman A, Hill J. Data Analysis Using Regression and Multilevel/Hierarchical Models. 2006.
- Hamilton CD, Swaminathan S, Christopher DJ, Ellner J, Gupta A, Sterling TR, et al. RePORT international: advancing tuberculosis biomarker research through global collaboration. *Clin Infect Dis* 2015;61Suppl 3:S155–9.
- Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42(2):377–81.
- He Z, Ryan P, Hoxha J, Wang S, Carini S, Sim I, et al. Multivariate analysis of the population representativeness of related clinical studies. *J Biomed Inform* 2016;60:66–76.
- Jaehn P, Rehling J, Klawunn R, Merz S, Holmberg C, group AGs. Practice of reporting social characteristics when describing representativeness of epidemiological cohort studies—a rationale for an intersectional perspective. *SSM Popul Health* 2020;11:100617.
- Kukull WA, Ganguli M. Generalizability: the trees, the forest, and the low-hanging fruit. *Neurology* 2012;78(23):1886–91.
- Ministério da Educação. National Report from Brazil Brazil. 2008.
- Ministério da Saúde do Brasil Sd Ve S. Plano Nacional pelo Fim da Tuberculose como Problema de Saúde Pública. . p. 54.
- Ministério da Saúde do Brasil, Secretaria de Vigilância em Saúde. Sistema de Informação de Agravos de Notificação Brasil. 2007.
- Ministério da Saúde do Brasil, Secretaria de Vigilância em Saúde. Manual de Recomendações para o Controle da Tuberculose no Brasil. . p. 288.
- Ministério da Saúde do Brasil, Secretaria de Vigilância em Saúde. Boletim Epidemiológico Tuberculose. 2020.
- Ministério da Saúde do Brasil, Secretaria de Vigilância em Saúde. Sistema de Informação de Agravos de Notificação. 2020 Available from: <http://portalsinan.saude.gov.br/>. [Accessed 14 August 2020].
- Regional Prospective Observational Research for Tuberculosis. RePORT-Brazil. 2020 Available from: www.reportbrazil.org. [Accessed 08 August 2020].
- Rocha MS, Bartholomay P, Cavalcante MV, Medeiros FC, Codenotti SB, Pelissari DM, et al. Notifiable Diseases Information System (SINAN): main features of tuberculosis notification and data analysis. *Epidemiol Serv Saude* 2020;29(1):e2019017.
- Saraceni V, Benzaken AS, Pereira GFM, Andrade KB, Oliveira PB, Arakaki-Sanchez D, et al. Tuberculosis burden on AIDS in Brazil: a study using linked databases. *PLoS One* 2018;13(11):e0207859.

- Shazzadur Rahman AAM, Langley I, Galliez R, Kritski A, Tomeny E, Squire SB. Modelling the impact of chest X-ray and alternative triage approaches prior to seeking a tuberculosis diagnosis. *BMC Infect Dis* 2019;19(1):93.
- Silva GDMd, Bartholomay P, Cruz OG, Garcia LP. Assessment of the tuberculosis surveillance system in Rio de Janeiro/RJ, Brazil, 2001 to 2006. *Cad saúde colet* 2010;18(3).
- Silva DR, Sotgiu G, D'Ambrosio L, Pereira GR, Barbosa MS, Dias NJD, et al. Diagnostic performances of the Xpert MTB/RIF in Brazil. *Respir Med* 2018;134:12–5.
- Tankeu AT, Bigna JJ, Nansseu JR, Endomba FTA, Wafeu GS, Kaze AD, et al. Global prevalence of diabetes mellitus in patients with tuberculosis: a systematic review and meta-analysis protocol. *BMJ Open* 2017;7(6):e015170.
- Trevor Hastie, Robert Tibshirani, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second Edition Springer Series in Statistics 2, editor. 2009.
- World Health Organization. Framework for Implementing the “End TB Strategy” in the African Region 2016–2020. World Health Organization; 2017.
- World Health Organization. Guidance for Tuberculosis Programme Managers, Analysis and use of Health Facility Data. . p. 1–44.
- World Health Organization. Global Tuberculosis Report 2019. . p. 283.
- Zhang X, Andersen AB, Lillebaek T, Kamper-Jorgensen Z, Thomsen VO, Ladefoged K, et al. Effect of sex, age, and race on the clinical presentation of tuberculosis: a 15-year population-based study. *Am J Trop Med Hyg* 2011;85(2):285–90.