Quality of tuberculosis information systems after record linkage

ABSTRACT
Objective: to analyze the quality of a tuberculosis notification information system after record linkage and spatial and temporal distribution of tuberculosis in a Brazilian state. Method: an ecological study carried between 2006 and 2016 in Sergipe, Brazil. A deterministic linkage was performed with Notifiable Diseases Information System and Mortality Information System, recording 7,873 cases and 483 deaths. The temporal trend of tuberculosis incidence was calculated. Results: there was an increase among men (2.75%), > 60 years (6.29%), higher education (4.34%) and indigenous (4.76%). A total of 190 new cases (2.9%) was found. There was an increasing trend in tuberculosis incidence with a concentration of deaths in the metropolitan region. Conclusion: the quality of the information system showed fragility in identifying cases and deaths in Sergipe. Temporal distribution showed an increasing trend in tuberculosis incidence, and spatial distribution identified higher incidences in southeastern Brazil. Descriptors: Epidemiology; Tuberculosis; Health Information Systems; Spatial Distribution; Database.

RESUMO
Objetivo: analisar a qualidade do sistema de informação da tuberculose após linkage de registros e a distribuição espacial e temporal da tuberculose em estado brasileiro. Método: estudo ecológico realizado entre 2006 e 2016 em Sergipe, Brasil. Foi realizado o linkage determinístico com as bases de dados do Sistema de Informação de Agravos de Notificação e do Sistema de Informação sobre Mortalidade, registrando 7,873 casos e 483 óbitos. A tendência temporal da incidência por tuberculose foi calculada. Resultados: houve incremento entre homens (2.75%), > 60 anos (6.29%), ensino superior (4.34%) e indígenas (4.76%). Foram encontrados 190 casos novos (2.9%). Observou-se tendência crescente da incidência de tuberculose com concentração de óbitos na região metropolitana. Conclusão: a qualidade do sistema de informação mostrou fragilidade na identificação de casos e óbitos em Sergipe. A distribuição temporal mostrou tendência crescente da incidência de tuberculose, e a distribuição espacial identificou maiores incidências na Região Sudeste. Descritores: Epidemiologia; Tuberculose; Sistemas de Informação em Saúde; Distribuição Espacial; Base de Dados.
INTRODUCTION

Tuberculosis (TB) is an important public health problem due to its magnitude, transcendence and vulnerability, affecting 10.4 million people worldwide, with Brazil being the highlight in this setting(1–2). In Brazil, TB incidence has experienced a reduction from 38.3/100 thousand inhabitants in 2009 to 34.8/100 thousand inhabitants in 2018, and a mortality rate from 2.5/100 thousand inhabitants in 2008 to 2.1/100 thousand inhabitants in 2017. Northeastern Brazil impacts by its high incidence and mortality rates(3).

The main instrument for monitoring TB cases in Brazil is the Notifiable Diseases Information System (SINAN - Sistema de Informação de Agravos de Notificação), whose function is reporting and monitoring cases, which integrates quantitative and qualitative information that constitutes a basis for calculation of clinical, epidemiological and operational indicators. The Mortality Information System (SIM - Sistema de Informação sobre Mortalidade) is used to monitor deaths by TB, in which it is possible to obtain important information for population’s health diagnosis, for implementing health policies and actions to eliminate or reduce risk situations(4–5).

However, underreporting of cases in SINAN is a reality and impairs real knowledge of the epidemiological situation of TB and the planning of actions aimed at its control(6–7). In this regard, probabilistic record linkage aims to improve the surveillance of a particular disease or disease to the extent that the combination of databases of qualitatively distinct information systems occurs, to promote case record improvement, excluding duplicates and including cases that would not have been identified if only one surveillance information system had been used(8–9). Analysis of SIM and SINAN databases through linkage allows identifying the epidemiological profile of TB cases, complementary analyses on surveillance, monitoring and care of cases(10).

Based on morbidity and mortality data, it is possible to perform a spatial analysis, which can facilitate understanding the dynamics of a public health problem in a territory, contribute to know about socio-environmental risks, identify more vulnerable areas of the disease and assist in decision-making and control measures. Thus, the spatial approach allows a better understanding of the demographic and socioeconomic conditions linked to TB(11).

OBJECTIVE

This study aims to analyze the quality of a TB notification information system after record linkage and spatial and temporal distribution of TB in a Brazilian state.

METHODS

Ethical aspects

The ethical aspects were followed according to the Research Ethics Committee with Human Beings, and the research performance was approved according to Helsinki Convention standards.

Study design, period and site

This is an observational, ecological study conducted from 2006 to 2016 in the state of Sergipe, Brazil. Sergipe has an estimated population of 2,288,116 inhabitants and a population density of 94.3 inhabitants/km²(12).

An instrument of observational studies in epidemiology, STROBE, was used to guide the methodology developed in this research, considering the EQUATOR network.

Population, and inclusion and exclusion criteria

The study population considered all cases and deaths recorded in SINAN as data sources for diagnosed TB cases, and in SIM, for deaths. Deaths that mentioned TB as the underlying cause were considered, comprising codes from A15 to A19 from 10th Revision of the International Classification of Diseases (ICD-10). In relation to cases and deaths of people not resident in Sergipe, those who were outside the analyzed period were excluded.

Homeless people and institutionalized individuals were included in this analysis without prejudice to results, since ecological analysis units were the municipalities.

SINAN and SIM data were made available by the Sergipe State Health Surveillance Office, and data from the resident population per year were obtained from the Brazilian Institute of Geography and Statistics (IBGE - Instituto Brasileiro de Geografia e Estatística)(13).

Study protocol

Record linkage of cases between SINAN and SIM was developed by deterministic linkage, databases were crossed by common identification variables (full name, date of birth, mother’s name); from crossing, pairs referring to a same individual were found, in order to identify TB patients with SIM registration and not notified and/or investigated in SINAN.

This study was carried out by steps systematized in a Microsoft Excel spreadsheet (Figure 1): filter application by state and municipality of residence and year of diagnosis (2006 to 2016), identification of duplicate cases and removal of duplicates in SINAN. To identify and remove duplicates, the following variables were used: patient name, notifying units, notification date, birth date, mother’s name and closure situation. The case was maintained with the date of the oldest diagnosis for situations in which a same patient was notified more than once by the same or other health institution in the coming days(6).

Subsequently, record linkage of TB cases between SINAN and SIM with the convergent variables was carried out. When there were TB cases in SINAN and SIM, other variables were added to data sheet, such as age, sex, municipality code of residence, race/color and education. A careful and visual comparison was performed to link records.

Analysis of results, and statistics

Data storage and linkage were carried out in the Microsoft Excel spreadsheet (Windows 2013 version, Microsoft Corporation; Redmond, WA, USA). For descriptive data analysis, epidemiological aspects (sociodemographic) were used through distribution of absolute, relative frequencies and 95% confidence intervals (Pearson’s chi-square test) by analysis of categorical parameters in the BioEstat 5.3 software, before and after record linkage, as well as percentage of increased cases.
From 2006 to 2016, 7,873 cases of TB were reported in SINAN. After filter application and duplicate removal, 7,643 cases were registered.

In the same period, SIM contained 483 records of TB deaths as the underlying cause. After linkage, 293 (61%) cases that were present in both databases and 190 (39%) cases were recorded only in SIM. There was an increase of approximately 2% in the total number of cases that went to death and were not reported in SINAN.

The sociodemographic characteristics of TB cases before and after record linkage show a significant increase in both men (2.75%) and women (1.88%). Age showed different increases for each age group, with an important increase in over 60 years (6.29%) and in the 40 to 59 years age group (3.53%) (Table 1).

A significant increase was identified among indigenous peoples (4.76%) and Caucasians (3.65%). As a limiting factor for filling in the variable, 389 (4.96%) cases were found in which the field was ignored and without race/color information. Concerning education, the largest increase was found in cases with complete/incomplete higher education (4.34%), and the lowest, in cases without education (1.06%). There was no information on level of education in 1,334 TB cases (17.03%) (Table 1).

Among the 7,643 cases registered in SINAN, 6,474 cases (84.7%) have been reported as new. After record linkage, 190 new TB cases were added to SINAN, which corresponded to 2.9% more new cases reported. Time series after linkage detected that the largest increase was in 2006, with an increase of 30 cases (5%), followed by 2007, with 21 cases (4.2%). There was a decrease in the percentage of increase in subsequent years, and the year with the lowest increase was 2013, with 8 cases (1.2%).

Incidence rates for TB (before and after linkage) showed highest rates in recent years, such as in 2016 (32.8%), 2014 (32.8%) and 2015 (31.2%), and lower rates in 2007 (24.6%) and 2010 (25.2%) (Table 2).

Temporal trend of TB after record linkage showed an increasing trend with APC of 1.66 (p value<0.01) (Figure 2).

Spatial distribution showed two thematic maps. The first (A) shows TB incidence after linkage and signals areas with a high incidence focus such as southeastern and center-southern Sergipe. The second map (B) identifies the municipalities that had TB deaths as the underlying cause and that were not reported in SINAN, identifying the metropolitan region of the state with the highest concentration of cases (Figure 3).

DISCUSSION

We identified 190 TB deaths that were not registered in SINAN, which corresponded to an increase of 2.9% more than new cases reported in SINAN. Thus, record linkage between SIM and SINAN becomes essential to identify cases of TB and underreported deaths. This comparison between databases allows a better analysis of case monitoring and assistance.13

Occurrences of unreported cases in SINAN and registered in SIM may suggest existence of barriers in access to health services and failures in the quality of these information systems, since, when outcome is death, an individual was probably not previously diagnosed by the Health Care Network, especially in Primary Care.14
Table 1 - Epidemiological characteristics of tuberculosis cases before and after record linkage of the Notifiable Diseases Information System and the Mortality Information System between 2006 and 2016, Aracaju, Sergipe, Brazil, 2020

<table>
<thead>
<tr>
<th>Epidemiological features</th>
<th>SINAN</th>
<th>SIM</th>
<th>Linkage</th>
<th>Increase</th>
<th>95% confidence interval</th>
<th>p value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.01</td>
</tr>
<tr>
<td>Female</td>
<td>2383</td>
<td>45</td>
<td>2428</td>
<td>1.88%</td>
<td>0.30 – 0.32</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>5260</td>
<td>145</td>
<td>5405</td>
<td>2.75%</td>
<td>0.69 – 0.72</td>
<td></td>
</tr>
<tr>
<td>Age group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.03</td>
</tr>
<tr>
<td>0 to 19 years</td>
<td>597</td>
<td>03</td>
<td>600</td>
<td>0.5%</td>
<td>0.07 – 0.08</td>
<td></td>
</tr>
<tr>
<td>20 to 39 years old</td>
<td>3656</td>
<td>39</td>
<td>3695</td>
<td>1.06%</td>
<td>0.46 – 0.48</td>
<td></td>
</tr>
<tr>
<td>40 to 59 years old</td>
<td>2374</td>
<td>84</td>
<td>2458</td>
<td>3.53%</td>
<td>0.30 – 0.32</td>
<td></td>
</tr>
<tr>
<td>60 years and older</td>
<td>1016</td>
<td>64</td>
<td>1080</td>
<td>6.29%</td>
<td>0.13 – 0.15</td>
<td></td>
</tr>
<tr>
<td>Race/color</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.01</td>
</tr>
<tr>
<td>White</td>
<td>1255</td>
<td>30</td>
<td>1285</td>
<td>2.39%</td>
<td>0.15 – 0.17</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>987</td>
<td>22</td>
<td>1009</td>
<td>2.22%</td>
<td>0.12 – 0.14</td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>82</td>
<td>03</td>
<td>85</td>
<td>3.65%</td>
<td>0.02 – 0.12</td>
<td></td>
</tr>
<tr>
<td>Mixed-ethnicity</td>
<td>4933</td>
<td>110</td>
<td>5043</td>
<td>2.22%</td>
<td>0.63 – 0.65</td>
<td></td>
</tr>
<tr>
<td>Indigenous</td>
<td>21</td>
<td>01</td>
<td>22</td>
<td>4.76%</td>
<td>0.02 – 0.07</td>
<td></td>
</tr>
<tr>
<td>No information</td>
<td>365</td>
<td>24</td>
<td>389</td>
<td>6.57%</td>
<td>0.04 – 0.05</td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.02</td>
</tr>
<tr>
<td>None</td>
<td>656</td>
<td>07</td>
<td>663</td>
<td>1.06%</td>
<td>0.07 – 0.09</td>
<td></td>
</tr>
<tr>
<td>Elementary school</td>
<td>4170</td>
<td>91</td>
<td>4261</td>
<td>2.18%</td>
<td>0.53 – 0.55</td>
<td></td>
</tr>
<tr>
<td>High school</td>
<td>1224</td>
<td>39</td>
<td>1263</td>
<td>3.18%</td>
<td>0.15 – 0.16</td>
<td></td>
</tr>
<tr>
<td>Higher education</td>
<td>299</td>
<td>13</td>
<td>312</td>
<td>4.34%</td>
<td>0.03 – 0.04</td>
<td></td>
</tr>
<tr>
<td>No information</td>
<td>1294</td>
<td>40</td>
<td>1334</td>
<td>3.09%</td>
<td>0.16 – 0.17</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>7643</td>
<td>190</td>
<td>7833</td>
<td>2.42</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: *parameter estimate with Pearson’s chi-square test/SINAN - Information System for Notifiable Diseases/SIM - Mortality Information System.

Table 2 - Absolute cases, percentage, increase and incidence rate of tuberculosis in the Notifiable Diseases Information System before and after record linkage with the Mortality Information System between 2006 and 2016, Aracaju, Sergipe, Brazil, 2020

<table>
<thead>
<tr>
<th>Year</th>
<th>SINAN cases (n (%))</th>
<th>SIM cases (n (%))</th>
<th>Linkage (n)</th>
<th>Increase</th>
<th>Incidence rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td>599 (9.2)</td>
<td>44 (9.1)</td>
<td>629</td>
<td>5.0%</td>
<td>30.1</td>
</tr>
<tr>
<td>2007</td>
<td>495 (7.6)</td>
<td>36 (7.4)</td>
<td>516</td>
<td>4.2%</td>
<td>24.6</td>
</tr>
<tr>
<td>2008</td>
<td>594 (9.1)</td>
<td>36 (7.4)</td>
<td>612</td>
<td>3.0%</td>
<td>29.2</td>
</tr>
<tr>
<td>2009</td>
<td>581 (8.9)</td>
<td>46 (9.5)</td>
<td>597</td>
<td>2.7%</td>
<td>28.5</td>
</tr>
<tr>
<td>2010</td>
<td>511 (7.8)</td>
<td>39 (8.0)</td>
<td>528</td>
<td>3.3%</td>
<td>25.2</td>
</tr>
<tr>
<td>2011</td>
<td>578 (8.9)</td>
<td>48 (9.9)</td>
<td>598</td>
<td>3.4%</td>
<td>28.6</td>
</tr>
<tr>
<td>2012</td>
<td>515 (7.9)</td>
<td>43 (8.9)</td>
<td>529</td>
<td>2.7%</td>
<td>25.3</td>
</tr>
<tr>
<td>2013</td>
<td>622 (9.6)</td>
<td>45 (9.3)</td>
<td>630</td>
<td>1.2%</td>
<td>30.1</td>
</tr>
<tr>
<td>2014</td>
<td>670 (10.3)</td>
<td>57 (11.8)</td>
<td>686</td>
<td>2.3%</td>
<td>32.8</td>
</tr>
<tr>
<td>2015</td>
<td>638 (9.8)</td>
<td>44 (9.1)</td>
<td>652</td>
<td>2.1%</td>
<td>31.2</td>
</tr>
<tr>
<td>2016</td>
<td>671 (10.3)</td>
<td>45 (9.3)</td>
<td>687</td>
<td>2.3%</td>
<td>32.8</td>
</tr>
<tr>
<td>Total</td>
<td>6.474</td>
<td>483</td>
<td>6.664</td>
<td>2.9</td>
<td>28.9</td>
</tr>
</tbody>
</table>

Note: SINAN - Notifiable Diseases Information System/SIM - Mortality Information System.

Figure 2 – Trend in the tuberculosis incidence rate per 100,000 inhabitants after linkage with the Notifiable Diseases Information System and the Mortality Information System Aracaju, Sergipe, Brazil, 2020

Note: APC - Annual Percentage Change.

Figure 3 - Tuberculosis infection rate after linkage, considering the Notifiable Diseases Information System and the Mortality Information System and cases of deaths by tuberculosis not registered in the Notifiable Diseases Information System, Aracaju, Sergipe, Brazil, 2020

Note: SINAN - Information System for Notifiable Diseases/SIM - Mortality Information System.
The sociodemographic profile of reported cases showed an increase in both males and females, but with greater sensitivity to males (2.75%), which reinforces the profile pointed out in the scientific literature\(^{15-16}\). Exposure to risk situations, such as alcohol abuse, smoking and other drugs, lower timely demand for health services in relation to women, the idea of male invulnerability to the disease, waiting time for care and difficulties in accessibility (prior scheduling, hours of care) of health services are factors that influence a higher proportion of TB cases in males\(^{17}\).

A for age group, the highest number of cases, both before and after record linkage, is concentrated in an age group from 20 to 39 years, but there was an important increase in older than 60 years (6.29%) and in people aged 40 to 59 years (3.53%). Individuals of reproductive age (20–59 years) have higher notification numbers, a standard that occurs nationally\(^{18-19}\) and interferes with the economic and social conditions of patients and their family.

However, the older the age, the higher the proportion of underreporting in SINAN, with higher proportions for individuals over 50 years old. This result can be justified by the difficulty in diagnosing TB in elderly people\(^{20-21}\) and by existence of other comorbidities that mask TB occurrence, increasing underreporting cases and increasing difficulties of access to health services in this age group. Correlation between age and mortality from TB demonstrates the importance of early diagnosis in individuals older than 60 years\(^{22}\).

Concerning race/color, notifications before and after record linkage recorded the highest occurrences of cases in mixed-ethnicity and white people, similar to other national surveys\(^{23-24}\). However, a significant increase among indigenous peoples (4.76%) points out a need for a greater focus in this population, which is vulnerable to TB.

Concerning level of education, there was an increase in this study in cases with complete/incomplete higher education (4.34%). It is reinforced that patients with higher education may have more access to health services and diagnosis, even in the advanced stage of TB or at death, than those with fewer years studying and low social status, but who even at death may not be diagnosed with TB\(^{8,10}\).

The findings identified that 39.3% of the patients who died did not have access to the diagnosis in a timely manner. This setting portrays late diagnosis and treatment, fragile promotion and prevention actions in a territory, as well as a lack of search for respiratory symptoms, identification of suspected cases and investigation of contacts\(^{23}\).

In a spatial distribution of TB cases not registered in SINAN, there is a greater concentration of deaths in the metropolitan region of Sergipe, especially in municipalities that also have high rates of disease incidence. Some municipalities that had low incidences reported deaths by TB and had no notification in SINAN. These are places close to municipalities that have high incidence rates; hypothetically, early TB diagnosis is weakens because as municipalities are close to others with high rates of incidence, the potential for disease transmission expands, contaminating areas previously not affected.

TB death is considered a sentinel event, because it is preventable and indicative of weaknesses in patient care by a health service, representing a concrete occasion to identify determinants and conditions in the disease process, not only for individuals, but for their interface with family and community\(^{25-27}\).

Thus, SINAN becomes an essential tool to assist epidemiological monitoring and support decision-making regarding TB management, but unreported cases interfere with the representation of the real epidemiological profile of TB\(^{26}\).

**Study limitations**

This study had limitations regarding secondary data that, although they are from the Brazilian National Information System, had missing characteristics. We know that the time frame and the study design do not explain the phenomenon with characteristics of cause and effect; however, it is possible to generate hypotheses and prompt new research. The identified limitations did not interfere with data analysis and presentation of results. This study presents extremely important statistical and thematic relevance for the Brazilian and worldwide setting.

**Contributions to nursing, health, and public health policies**

Information retrieval after record linkage promoted an improvement in the quality of information, since, from linkage, it was possible to identify the concentration of underreported cases, the mapping and the design of a more reliable epidemiological profile. Deaths that were not recorded by SIM indicate low uptake of cases or failures in notification by the programs and end up interfering with a safe monitoring of the epidemiological situation of a given location.

Record linkage, as an alternative to this problem, it can be a tool for improving the information on injuries in terms of completeness, consistency and reliability of the compulsory notification system\(^{28}\). The adoption of linkage record strategies is an important tool used by many countries to establish estimates of incidence and number of cases and to increase the reliability of epidemiological data\(^{29}\). Therefore, record linkage brought important contributions to TB surveillance.

The study developed can help developing public policies, especially regarding implementing information systems that can communicate with each other and that reflect the reality in health. Improving the quality of these systems can help the work process and, consequently, improve the quality of care for individuals with TB.

**CONCLUSION**

In conclusion, there was a greater increase in men, over 60 years old, with complete/incomplete higher education and indigenous people. After linkage, a significant increase in new cases and a growing trend in the TB incidence rate was identified in Sergipe, with the highest concentration of deaths in the metropolitan region. Due to the low quality of the information systems, it was identified that municipalities with lower incidence rates registered deaths that were not reported in SIM. The linkage technique was essential for improving the quality of data in the information systems, as it showed more reliable information to monitor TB and decision-making regarding actions relevant for TB control in Sergipe.

**ACKNOWLEDGMENTS**

We would like to thank the State Health Office under the coordination of Health Surveillance, for the availability of databases and support, as well as the Graduate Program in Health Sciences at Universidade Federal de Sergipe, for academic support.
REFERENCES


