Decision model on the demographic profile for tuberculosis control using fuzzy logic

Laisa Ribeiro de Sá¹, Jordana de Almeida Nogueira², Ronei Marcos de Moraes³

¹ Nurse, Master in Models of Decision and Health. PhD Student at the Postgraduate Program in Models of Decision and Health, Universidade Federal da Paraíba (UFPB). João pessoa, PB, Brazil. E-mail: tsa8910@hotmail.com.
² Nurse, PhD in Nursing. Associate Professor of the Health Sciences Center, UFPB. João Pessoa, PB, Brazil. E-mail: jal_nogueira@yahoo.com.br.
³ Statistician, PhD in Applied Computing. Associate Professor of the Department of Statistics, UFPB. João Pessoa, PB, Brazil. E-mail: ronei@de.ufpb.br.

ABSTRACT
This study aimed to describe the relationship between demographic factors and the involvement of tuberculosis by applying a decision support model based on fuzzy logic to classify the regions as priority and non-priority in the city of João Pessoa, state of Paraíba (PB). As data source, we used the Notifiable Diseases Information System between 2009 and 2011. We chose the descriptive analysis, relative risk (RR), spatial distribution and fuzzy logic. The total of 1,245 cases remained in the study, accounting for 37.02% of cases in 2009. High and low risk clusters were identified, and the RR was higher among men (8.47), with 12 clusters, and among those uneducated (11.65), with 13 clusters. To demonstrate the functionality of the model was elected the year with highest number of cases, and the municipality district with highest population. The methodology identified priority areas, guiding managers to make decisions that respect the local particularities.

Descriptors: Tuberculosis; Relative Risk (Public Health); Spatial Analysis; Fuzzy Logic; Public Health Nursing.

RESUMO
Objetivou-se descrever a relação entre os fatores demográficos e o acometimento da tuberculose, aplicando-se um modelo de suporte à decisão baseado em lógica fuzzy para categorização dos locais em prioritários e não prioritários, na cidade de João Pessoa-PB. Como fonte de dados, utilizou-se o Sistema de Informação de Agravos de Notificação, entre 2009 e 2011. Optou-se por análise descritiva, risco relativo (RR), distribuição espacial e lógica fuzzy. Permaneceu no estudo 1245 casos, sendo que o ano de 2009, foi responsável por 37,02% dos casos. Identificou-se conglomerados de alto e baixo risco, sendo que o RR foi maior entre os homens (8,47), com 12 conglomerados, entre aqueles sem escolaridade (11,65), com 13 conglomerados. Para demonstrar a funcionalidade do modelo, elegeu-se o ano com maior número de casos e um bairro do município com maior contingente populacional. A metodologia permitiu identificar áreas prioritárias direcionando gestores para decisões que respeitem as particularidades locais.

Descritores: Tuberculose; Risco Relativo; Análise Espacial; Lógica Fuzzy; Enfermagem em Saúde Pública.
INTRODUCTION

Tuberculosis (TB) is an infectious disease of chronic evolution with sufficient diagnostic and therapeutic resources for a proper and successful clinical management. Despite the available technology, the disease incidence is still high, with alarming numbers. Estimates from the World Health Organization (WHO) indicate the disease as the second leading cause of death worldwide. TB was the basic cause of 1.4 million deaths in 2011, and 8.7 million people fell ill in the same year\(^1\).

Although in recent years, Brazil has owned a decline in the number of cases, it still occupies the 19th position among 22 countries, concentrating 80% of the tuberculosis burden in the world and the 104th position relative to the incidence rate. In 2012 were reported 70,047 new cases, corresponding to an incidence rate of 36.1/100,000 inhabitants\(^1\). Every year about 4,500 people die from tuberculosis, therefore, this is the fourth leading cause of death from infectious diseases and the leading cause of death among AIDS patients\(^2\).

In 2012, in the state of Paraíba, located in northeastern Brazil, were reported 1,132 new cases of the disease, with an incidence rate of 29.7/100,000 inhabitants. The state capital, João Pessoa, presented an incidence rate of 46.4/100,000 inhabitants for new cases. Over the past 10 years, this indicator has shown a falling trend, similar to the incidence rate in Brazil\(^2\).

The studies agree that TB is conditioned to social aspects, where poverty and poor sanitary conditions determinants, together with the weaknesses of management and organization of health services collaborate to the increase and/or permanence of the disease incidence rates\(^3-5\).

Thus, new answers must be found to promote fairness, ensuring access to patients, seeking not only their care and well-being, but in a broader sense, the consolidation of the Unified Health System (SUS) in its guidelines. The alternatives lie in the readjustment of the health care system in the assistance to these patients, redefinition of procedures and organizational charts, redefinition of institutional missions of civil society entities, and the search for alternatives to solve the problem. Among those alternatives, communication should occupy a prominent place\(^6\).

In this perspective, and with the significant number of TB cases, arose the need to analyze the demographic profile of the disease in the city of João Pessoa (state of Paraíba – PB) to facilitate the decision-making process related to this disease. As the variables associated with the involvement of tuberculosis have some degree of uncertainty, a system based on fuzzy rules becomes an interesting alternative for a decision model that allows the categorization of priority regions using each variable separately.

Studies using the proposed model in various morbidities were found in the literature, but only assessing cases in their entirety\(^7-8\). However, no studies associating cases of tuberculosis according to demographic variables through a spatial analysis were found, and this is proposed in the present study.

Given the above, this study aimed to describe the relationship between demographic factors and the involvement of tuberculosis. It also applied a decision support model based on fuzzy logic to classify the regions as priority and non-priority, according to each variable.

METHODS

This is an epidemiological, retrospective study, based on records of tuberculosis cases in the city of João Pessoa.

It included the cases reported in the database of the Notifiable Diseases Information System (Sinan) of the Health Secretariat of the state of Paraíba between 1 January 2009 and 31 December 2011. The study excluded all cases in which data could lead to systematic errors, such as cases from other municipalities without information about the home district.

Subsequently, was held a database refinement to identify the variables of interest: the year of notification, gender, educational level, district (residence unit of the notified case and not necessarily where the infection
occurred). These data were arranged in a spreadsheet and analyzed according to the frequency and proportional distribution to generate the relative risk (RR), spatial scan statistic (CSS) and fuzzy logic. From these variables, tables and graphs were generated for a better presentation of results.

The Brazilian Census conducted in Brazil in the years 2000 and 2010 was used to calculate the incidence rates. The denominators were estimated from projections of the resident population in the city, using the population base of 1,000 inhabitants in each district of João Pessoa.

The relative risk (RR) was represented by the ratio of the incidence rate in a district divided by the incidence rate of the entire region (João Pessoa). The use of maps to spatially represent the RR allowed primary analyzes, such as the identification of regions that presented the highest and lowest risks of disease occurrence. The RR was divided into class intervals, previously establishing a color for each interval. Risk maps were prepared according to the year of occurrence and variable studied (total, gender and educational level). For the interpretation of the RR map, the following criteria were defined (Table 1):

<table>
<thead>
<tr>
<th>Relative Risk</th>
<th>Interpretation of Relative Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greater than or equal to 0 and less than 0.5</td>
<td>Districts where the relative risk ranged from 0 to 0.5, i.e., relative risk lower than half the overall risk of the study area.</td>
</tr>
<tr>
<td>Greater than or equal to 0.5 and less than 1.0</td>
<td>Districts where the relative risk ranged from 0.5 to 1.0, i.e., higher than half the overall risk of the study area.</td>
</tr>
<tr>
<td>Greater than or equal to 1.0 and less than 1.5</td>
<td>Districts where the relative risk ranged from 1.0 to 1.5, so the relative risk is greater than the overall risk of the study area in less than 50%.</td>
</tr>
<tr>
<td>Greater than or equal to 1.5 and less than 2.0</td>
<td>Districts where the relative risk ranged from 1.5 to 2.0, so the relative risk is higher than the overall risk of the study area in more than 50%.</td>
</tr>
<tr>
<td>Greater than or equal to 2.0</td>
<td>Districts where the relative risk was higher than 2.0.</td>
</tr>
</tbody>
</table>

*Adapted table (7-8,10).

Subsequently, the CSS method was used to detect spatial clusters, allowing to outline the critical regions using graphical computational algorithms. For spatial analysis, we adopted the probabilistic model of Poisson that associates the area information with a single point within the polygon, the centroid, which is the mass center of each region of the studied area.

In the CSS method, the restriction is in relation to the percentage of the population at risk, determining the increase of the search radius up to β% of the population. This study used β% of the population of 1%, 3%, 5% and 10%. The maps from the respective β% should be compared with the risk map of every year to find the map that best fits. Thus, from a visual inspection of the maps representing the phenomenon under study, we compare the CSS map with the risk map, opting for the CSS map that most closely resembles the risk map. The spatial clusters identified were presented on maps in which the red dots represent the centroids.

After the spatial analysis, was applied the decision model based on fuzzy logic. The fuzzy logic originates from the concept of fuzzy sets and represents a scheme capable of translating inaccurate information into numerical values, making it possible to solve problems that classical logic could not solve. Then, a fuzzy set is defined as a class of objects with continuous degrees of association. This set is characterized by membership functions assigned to each object of the set, varying between zero and one.

One of the most common ways to express this knowledge is through rules of the condition-action type. Thus, all knowledge is represented by a set of rules in...
which the conditions are established from a set of linguistic terms, which are associated with the input/output variables. Therefore, the fuzzy rule is a unit capable of capturing some specific knowledge \(^{(14)}\).

At this stage, were created the input variables (total relative risk, totalCSS, female relative risk, male relative risk, male CSS, uneducated relative risk, educated CSS, educated relative risk - at least literate, educated CSS - at least literate) and the output variables (priority and non-priority) and the membership functions for each of them. Each of these variables has its linguistic terms representing a fuzzy set \(^{(14)}\). In the final result, it was possible to identify the priority and non-priority regions of the city of João Pessoa, with their respective degrees of membership.

The ethical aspects established by Resolution 196/96, which regulates research on humans, were respected. Since this is a study with search in a secondary database, we used the Institutional Consent term issued by the Municipal Secretariat of Health in João Pessoa-PB under number 20.900/2013.

RESULTS

In the period between 2009 and 2011, were reported 1,329 cases of tuberculosis in the city of João Pessoa-PB. Based on the inclusion criteria, 1,245 cases remained in the study, of which 461 (37.02%) were reported in 2009, 417 (33.49%) in 2010 and 367 (29.47%) in 2011. The male gender was predominant (853/68.5%), as well as subjects with some level of education (983/78.95%).

To exemplify, in the analysis of relative risk and CSS, were used the maps of tuberculosis in João Pessoa in 2009, for the total, female, male, uneducated and educated (Figure 1). Note that maps were built for the entire study period.

The map of total RR showed the behavior of tuberculosis in the districts of João Pessoa-PB, where the RR varied between 0 and 5.90 per 1,000 in habitants. The spatial analysis of cases through statistical CSS identified clusters a thigh and low risk. Nineteen districts showed spatial clusters for the year 2009, with 5% of the population. Comparing the map of total CSS with the map of total risk, there is a coincidence of regions with the greatest RR ratios and those with significant spatial clusters.

Analyzing the maps of risk and CSS by gender, the RR was higher among men (8.47), with the presence of 12 clusters. The RR was lower among women, reaching a maximum value of 4.66. However, the CSS method identified 13 significant clusters. In the analysis of the male population, was also observed that some districts showed RR two or more times higher when compared with the total RR and the female gender.

In relation to the educational level, the RR reached 11.65 among uneducated cases and 6.76 among educated individuals. In both groups were identified 13 clusters, and in the first, there was a higher concentration in the districts of the central area of the city and/or districts with economically vulnerable population.
In the analysis of fuzzy variables involving relative risk were defined five linguistic terms, namely: very high risk, high risk, medium risk, very low risk and low risk (Chart 1). The following membership functions of the fuzzy variables were of trapezoidal type: total relative risk, female relative risk, male relative risk, uneducated

Figure 1: Maps of relative risk and CSS of tuberculosis cases according to residence district. João Pessoa, PB, Brazil, 2009.
relative risk, educated relative risk. The binary membership function for the results of the CSS method were named withScan and withoutScan. The following membership functions of fuzzy variables were of triangular type: total CSS, female CSS, male CSS, uneducated CSS and educated CSS. The output variable obtained priority and non-priority as a linguistic term, and membership functions of trapezoidal type (Chart 1).

![Membership functions](image)

**Chart 1**: Membership functions of the input variables of relative risk and CSS, and membership functions of the output variable. João Pessoa, PB, Brazil, 2009.

For the construction of the rules database, were used ten fuzzy variables obtained through the annual maps of the relative risk and CSS maps for total, female, male and uneducated population, as well as educated population. Five banks were created with ten rules defined for each, totaling 50 rules. Some of these rules are presented below for exemplification:

- IF (total relative risk is very high) AND (total scan is withScan), THEN (decision is priority)
- IF (total relative risk is very high) AND (total scan is withoutScan), THEN (decision is non-priority);

After construction of the rules, was applied the decision model based on fuzzy logic to classify the districts as priority and non-priority for tuberculosis in relation to gender, educational level and total. This model was applied to each set of rules for all 64 districts, in all years, demonstrating its functionality. It can be used as a low cost method by managers and professionals, for an effective combat of the morbidity according to the target population and the regions considered as priority.
To exemplify the application of the model based on fuzzy logic, we chose a district with high population and the year 2009, a period with the highest occurrence of cases (Charts 2-4).

Each column represents a membership function for each of the variables in the rule. Each row represents a rule that is activated according to the input data, which may or may not activate the last column, according to the value of the input variables. Thus, the output function shows the suggestion of decision to be made by the manager for the district; as priority (partially or completely filled by blue color on the right side) or as non-priority (partially or completely filled by blue color on the left side).

Following the same conception, the model confirmed the expected result, in which the elected district was classified as priority in relation to the total population (Chart 2), male gender (Chart 3) and educated (Chart 4). It was classified as non-priority in relation to the female gender (Chart 3) and uneducated (Chart 4).
Chart 2: Result of the total linguistic fuzzy model in the district of Mangabeira- João Pessoa, PB, 2009.
### Male Linguistic Fuzzy Model, Magabeira District (2009)

<table>
<thead>
<tr>
<th>Total Relative Risk</th>
<th>Total CSS</th>
<th>Female Relative Risk</th>
<th>Female CSS</th>
<th>Male Relative Risk</th>
<th>Male CSS</th>
<th>Uneducated Relative Risk</th>
<th>Uneducated CSS</th>
<th>Educated Relative Risk</th>
<th>Educated CSS</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.74</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DISCUSSION

The disease concentration in the most populous districts or those of lower economic class may be due to its characterization as a social problem, resulting from the interaction of several factors, among which the low family income, poor education, poor/non-existent housing, large families, high density communities, food malnutrition, alcoholism, associated diseases, migration flows, deficiency of the health system\textsuperscript{(15)}.

The majority of studies found corroborate the local scene because they characterize the disease as a social problem resulting from the interaction of several factors, of which the most cited were the male gender, low family income, poor education, poor/non-existent housing, large families, alcoholism, smoking, mental illness, AIDS and occupational risk\textsuperscript{(4,16–20)}.

The gender as a risk factor can be justified by the fact that historically men seek health services for preventive control with less frequency, especially when it comes to basic health care, and gender issues are very strong when it comes to the health-disease process\textsuperscript{(21)}.

The actions related to care have always been associated with women, and the access of men to health services is hindered by labor issues, when there is a mismatch between the health services hours and the working hours of male subjects. As a result, men tend to search for health services after the worsening of clinical picture. In addition to be seen as the home providers, and feeling more difficulty in expressing themselves, men’s needs are not met instantly.

In relation to educational level, the study results show that most of the cases reported occur in educated people. However, the application of the decision model found that most districts were classified as priority for uneducated people, confirming this is also a risk factor for TB. Corroborating the above in terms of education, the longer the duration of studies, the lesser the occurrence of TB, showing that more years of study result in less occurrence of TB\textsuperscript{(22)}.

The decision model based on fuzzy logic is considered a fundamental methodological support to respond to the health process uncertainties and the illness of subjects. Moreover, it improves the understanding of individual risk and is essential in epidemiological studies.

The theory of fuzzy sets appears as a useful methodological tool in social areas by shaping the uncertainties inherent in its phenomena, especially in epidemiology\textsuperscript{(23)}. Several epidemiological studies show the importance of the theory of fuzzy sets when studying the health and disease process, contributing in the construction of diagnostics and expert systems. Among the studies using the theory of fuzzy sets satisfactorily, we can highlight a few: application of fuzzy clusters model for analysis and identification of different types of tuberculosis\textsuperscript{(24)}; creation of a fuzzy model to predict the result of TB cases with ongoing treatment\textsuperscript{(25)}.

CONCLUSION

The results showed that the decision model based on fuzzy logic is a low cost tool and allowed the classification of all the city districts as priority or non-priority, according to each variable, enabling the visualization of peculiarities, and guiding managers for decisions that respect the particularities of each location. Note that this is an innovative approach because it visualizes the totality, as well as each variable separately, allowing to know the vulnerability of each region for the effective combat of the morbidity.

Thus, a further understanding of the vulnerable populations to the disease and the region become possible, which facilitates the definition of public policies for the priority and specific areas and the population of that region. Also, policies for the combat and prevention of morbidity can be defined together with health professionals, making it possible to decrease the onset of tuberculosis.
REFERENCES


25. Kalhori SRN, Zeng XJ. Fuzzy Logic Approach to Predict the Outcome of Tuberculosis Treatment Course Destination. In:
Disponível em:

Received: 12/11/2013.
Accepted: 09/15/2014.
Published: 06/30/2015.