Longitudinal clustering of tuberculosis incidence and predictors for the time profiles: the impact of HIV

P. Sousa,* A. Oliveira,†‡ M. Gomes,§¶ A. R. Gaio,†# R. Duarte§¶**††

*Faculty of Medicine of Porto University, Porto, †Centre of Mathematics, University of Porto, Porto, ‡School of Allied Health Sciences, Polytechnic Institute of Porto, Vila Nova de Gaia, §Epidemiology Research Unit Institute of Public Health, University of Porto, Porto, §Centro Hospitalar de Vila Nova de Gaia/Espinho EPE, Vila Nova de Gaia, ¶Mathematics Department, Faculty of Sciences, University of Porto, Porto, **Chest Disease Centre, Vila Nova de Gaia, ††Department of Clinical Epidemiology, Predictive Medicine and Public Health, University of Porto Medical School, Porto, Portugal

SUMMARY

BACKGROUND: Portugal remains the country with the highest tuberculosis (TB) incidence in Western Europe. OBJECTIVES: To identify longitudinal trends in TB incidence in Portugal from 2002 to 2012 and investigate the longitudinal effect of sociodemographic and health-related predictors among the resident population on the TB incidence rate.

METHODS: We used data from the National Tuberculosis Surveillance System and other national institutions. K-means longitudinal clustering algorithm was performed on TB incidence time profiles from districts of Portugal.

RESULTS: Three longitudinal profiles for the TB incidence rate of Portugal were identified. In all of them, TB incidence decreased over time. Among all studied sociodemographic and health-related predictors, human immunodeficiency virus (HIV) notification rate and unemployment were shown to have (positive) significant effects on TB incidence. In particular, the greatest effects were found for the HIV notification rate.

CONCLUSIONS: Our study supports the view that combined TB-HIV strategies and the improvement of social determinants can contribute to decreases in TB incidence.

KEY WORDS: human immunodeficiency virus; AIDS; public health; tuberculosis; unemployment
The Portuguese sociodemographic and health-related indicators considered in the study were as follows: 1) working age population: the proportion aged 15–64 years; 2) elderly population: the proportion aged ≥65 years; 3) unemployment: the proportion registered with the employment office as of 31 December for each year; 4) incarcerated persons: the proportion of existing inmates as of 31 December for each year; 5) prison overcrowding: the percentage of excess inmates (number of inmates – prison capacity)/prison capacity; 6) HIV notification rate; 7) use of alcohol or other drugs: the proportion of users with medical appointments for treatment for addiction provided by the public health system due to alcohol or other drug use at least once for each year; 8) physicians: the proportion of physicians or surgeons of any employment status in the population; and 9) medical appointments: the proportion of medical appointments at health centres from all clinical specialties in the population.

All proportions were multiplied by 100,000 population. For all predictors, data were collected yearly and obtained from the following national institutions: Statistics Portugal (Instituto Nacional de Estatística, INE), Employment and Vocational Training Institute (Instituto do Emprego e Formação Profissional, IEFP), Directorate General of Prison Services (Direcção Geral Dos Serviços Prisionais, DGSP), Institute of Drugs and Drug Addiction (Instituto da Droga e da Toxicodependência, IDT) and the National Health Institute Doutor Ricardo Jorge (Instituto Nacional de Saúde, INSa) (Appendix Table A.1).7

The study assessed the 18 districts in continental Portugal and the Autonomous Region of Madeira based on the territorial classification in the 2010 map of Portugal (Carta Administrativa Oficial de Portugal 2010). The Autonomous Region of Azores was excluded due to the absence of data before 2007. Continental Portugal is administratively divided into 18 districts, while the Autonomous Region of Madeira, located in the Atlantic Ocean, has two inhabited islands. Of the 18 districts, Lisbon is the most populous, with an estimated population of 2,252,719 (2011 Census).

Because we used published retrospective surveillance data without the possibility of linking patient records to patient personal data, ethical approval was not necessary for this retrospective study.

Statistical analysis
The K-means longitudinal clustering algorithm was performed on the time profiles of TB incidence rate in the 19 districts.11 Two- and three-cluster structures, each with 40 randomly chosen starting points, were considered. The final number of clusters was chosen based on the epidemiological interpretation of the groups identified and on the statistical criteria of Calinsky & Harabatz and Davies & Bouldin (Figure 1).12

Apart from the time and the cluster structure identified, those predictors assumed to have a potential longitudinal effect on the TB incidence were HIV notification rate, unemployment, incarceration, prison overcrowding, alcohol or drug use, working age population, elderly population, number of physicians, number of medical appointments and interactions of these covariates with the cluster structure.

Data were grouped by district and autonomous region. The changes over time for HIV notification and the unemployment rates, adjusted for the clusters structure, were assessed by multiple regression fitted using generalised least squares (GLS), allowing for heteroscedastic and correlated errors. The longitudinal effect of the collected sociodemographic and health-related predictors on the TB incidence rate was also studied using GLS. The best variance-covariance matrix was that with a time autocorrelation structure of order 1. On graphic analysis, homoscedasticity and normality of the model residuals did not seem to be compromised. Model comparisons were based on the likelihood ratio test, whenever possible, and on the Bayesian information criterion (BIC) otherwise.

All statistical analyses were performed using the R language and software environment for statistical computation, version 3.1.2 (R Computing, Vienna, Austria). The significance level was fixed at 0.05.

RESULTS
During the period studied, a total of 33,394 new cases were notified. Males accounted for 67% of the total; the mean age was 44.09 years (standard deviation 18.57). TB cases with alcohol or other drug addictions represented 23% of the total. HIV co-infection, whether reported by patients or confirmed through serology, was present in 14% of all TB cases; however, the serological status was unknown in 26% of TB cases; 14% of patients were foreign-born; 4% were homeless or were living in community residences; and 2% had a past or current history of incarceration.

The two districts with the largest urban areas (Porto and Lisbon) consistently presented the highest TB incidence rate throughout the entire period evaluated; five other regions shared the lowest rates, namely the
Archipelago of Madeira, two coastal districts and two sparsely populated continental districts.

As expected, the most populous districts of Porto and Lisbon presented the highest absolute frequencies on every predictor considered. However, in terms of rates, none of the 19 regions was observed to be systematically at the top or bottom of a predictor prevalence. For example, Porto experienced the highest rates of unemployment and the largest proportion of working age population for several years, while the touristic Algarve region presented the highest average rates for the treatment of alcohol and drug dependencies, and Lisbon and Bragança, a northern continental district, alternated with the highest rates of incarceration. The district of Lisbon had the highest HIV notification rate throughout the study. These data are not presented in tabular form here because the data have a longitudinal structure and several predictors were studied.

The statistical classification analysis identified three (longitudinal) patterns. TB incidence rates for each region and the mean trajectory of each cluster are shown in Figure 2. The three time trajectories are all essentially decreasing functions. Cluster 1 comprises the 10 regions with the lowest incidence rates, mainly located in the central region of continental Portugal and the Autonomous Region of Madeira. The districts with the highest incidence rates, namely Porto and Lisbon, gave rise to Cluster 3. Cluster 2 encompasses the remaining seven districts, mainly from northern coastal and southern Portugal. The longitudinal mean incidence rates of Cluster 2 are roughly between the mean incidence rates from Clusters 1 and 3 (Figure 2).

The highest HIV notification rates were observed within Cluster 3 and in some regions of Cluster 2. Until 2008, the mean rates corresponding to Cluster 3 were significantly higher than those of the remaining two clusters, between which there was no statistically significant difference. In all clusters, the mean HIV notification rate decreased with time: the yearly decrease was estimated at 0.42 notified cases per 100 000 population ($P < 0.05$) for Cluster 1, 0.53 ($P < 0.05$) for Cluster 2 and 1.39 ($P < 0.05$) for Cluster 3 (Appendix Figure 2, Appendix Table A.2).
The highest unemployment rates were observed in Porto, a Cluster 3 region, and in some regions of Cluster 2. In all clusters, the mean unemployment rates fluctuated with time until 2007, after which they increased in an approximately linear way. In 2012, the rate was double the value observed in 2002. The average yearly increase was estimated at 834.51/100,000 (P < 0.001) in Cluster 1, 946.37 (P < 0.001) in Cluster 2 and 5869.30 (P < 0.001) in Cluster 3 (Appendix Figure A.2, Appendix Table A.2).

Among the sociodemographic and health variables analysed, only HIV notifications and unemployment rates were found to have a statistically significant effect on the multiple model for the evolution over time of the TB incidence rate, after adjusting for time and cluster structure (Appendix Table A.3).

The effect of each of the predictors in the final model was interpreted for fixed values of the remaining variables in the model (Appendix Table A.3). Following that structure, 1) an increase of 10 HIV notifications/100,000 was associated with two new TB cases/100,000 (P < 0.05), and 2) an increase of 1000 unemployed subjects/100,000 was associated with one new TB case/100,000 (P < 0.05).

Finally, for any fixed values of the explanatory variables in the model, including time, Cluster 1 (vs. Cluster 3) was expected to have the lowest (vs. highest) average TB incidence rate (P < 0.05). Cluster 2 was predicted to lie between the two (P < 0.05) (Appendix Figure A.3, Appendix Table A.3).

DISCUSSION

Our study identified longitudinal trends in the TB incidence rate in Portugal from 2002 to 2012 and assessed the effect of socio-demographic and health-related predictors on the time profiles identified. We found three longitudinal patterns of TB incidence in the 19 Portuguese regions evaluated; all clusters essentially featured decreasing trends in incidence. HIV co-infection and unemployment were identified as those factors with the highest impact on TB time profiles.

HIV infection is a well-known risk factor for active TB.13 There is a strong spatio-temporal association between TB and HIV,14 even in low-incidence regions.14,15 Although the HIV incidence rate in Portugal has declined steadily since 2001, it remains one of the highest in Western Europe, at 13.6/100,000. Lisbon, the largest urban centre, had the highest HIV incidence rate in the country, at 31.8/100,000.16 According to our models, HIV notification was the factor with the greatest impact on TB time profiles.

Several strategies to reduce HIV infection have been implemented in the country. National strategies include free distribution of condoms since 2006 and the syringe exchange programme for drug users established in 1993. Mobile and fixed centres for rapid HIV diagnosis have been used to reduce the time to diagnosis.16 To improve diagnosis and treatment adherence, HIV diagnosis and highly active antiretroviral therapy are freely available to all patients,
including those with non-legal status. Other strategies have been developed to manage hard-to-reach populations, namely drug users, with improvements in TB detection and treatment completion.

At the end of 2012, the national TB and HIV programmes were fully integrated, and well-defined joint strategies were established. In 2013, 77% of all TB patients were screened for HIV, and 15% were positive. Information on TB screening among HIV-positive patients is lacking; however, data show that in 2013, 26% of patients with the acquired immune-deficiency syndrome (AIDS) had TB as an AIDS-defining disease.

In our study, unemployment was also positively associated with increased TB incidence. This finding may be related to the socio-economic and demographic profiles often linked to the unemployed, including low income, poor access to health care and poor nutrition. Unemployment has been associated with increased risk of TB.

Our study had some limitations. Due to lack of data from other institutions for some years, it was not possible to include all variables of interest, such as the homeless, those on minimum social welfare, the foreign-born and those dependent on alcohol or drugs. Second, although the model assumed non-zero correlations between observations from different years, in practice this time dependence referred to the infectious period. In Portugal, the median time between the onset of symptoms and diagnosis ranges from 36 to 92 days. The analysis could be enhanced by the use of shorter intervals; however, most of the data would not be available. The Portuguese HIV surveillance system also suffers from some classical problems: reporting delays in the notification process, mainly in the most recent cases, and underreporting of cases. The HIV notification and TB incidence rates may therefore be biased. Furthermore, the absence of molecular epidemiology limits the present analysis to associations between notifications, without insights into the contribution of HIV to TB transmission and into the enhanced susceptibility of HIV patients to progress rapidly to TB disease. Finally, it was not possible to determine the prison establishment for 1–2% of the inmates for each year.

The strengths of the study include the TB case detection rate in Portugal (84–93%) during the period studied: this longitudinal study covered 11 years and about 35 000 reported cases of active TB, and almost the entire national territory, except for the Azores. Finally, it should be highlighted that we cross-checked data from several national institutions.

Our findings have important public health implications, and support the view that combined TB-HIV strategies and improvements in social determinants can contribute to the decrease in TB incidence.

Acknowledgements

ARG was partially funded by the European Regional Development Fund through the COMPETE Programme and by the Portuguese Government through the Fundação para a Ciência e a Tecnologia under the PEst project (C/MA/TU0144/2013).

Conflicts of interest: none declared.

References


APPENDIX

Table A.1  Data source of variables analysed

<table>
<thead>
<tr>
<th>Variable</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Working age population</td>
<td>INE, annual estimates of resident population at the municipal and age group level, 2002–2012</td>
</tr>
<tr>
<td>Elderly population</td>
<td>INE, annual estimates of resident population at the municipal and age group level, 2002–2012</td>
</tr>
<tr>
<td>Resident population</td>
<td>INE, annual estimates of resident population at the municipal and age group level, 2002–2012</td>
</tr>
<tr>
<td>Incarceration</td>
<td>DGSP, capacity and existing inmates on 31 December 2002–2012</td>
</tr>
<tr>
<td>Prison overcrowding</td>
<td>DGSP, capacity and existing inmates on 31 December 2002–2012</td>
</tr>
<tr>
<td>Alcohol or other drugs</td>
<td>IDT, annual reports ‘The Drugs and Drug Addiction Situation in the Country’, information at the district level, 2002–2011. In 2012, the available data referred only to patients with problems related to drug abuse. The value for 2012 was therefore estimated by calculating the mean of 2010 and 2011</td>
</tr>
<tr>
<td>HIV infection</td>
<td>INSA, notified cases of HIV infection, 2002–2012</td>
</tr>
<tr>
<td>Physicians</td>
<td>INE, number of physicians at the municipal level, 2002–2012</td>
</tr>
<tr>
<td>Medical appointments</td>
<td>INE, number of medical appointments in health centres at the municipal level, 2002–2012</td>
</tr>
</tbody>
</table>

INE = Instituto Nacional de Estatística (Statistics Portugal); IEFP = Instituto do Emprego e Formação Profissional (Employment and Vocational Training Institute); DGSP = Direcção Geral Dos Serviços Prisionais (Directorate General of Prison Services); IDT = Instituto da Droga e da Toxicodependência (Institute of Drugs and Drug Addiction); HIV = human immunodeficiency virus; INSA = Instituto Nacional de Saúde Doutor Ricardo Jorge (Doutor Ricardo Jorge National Health Institute).

Table A.2  Estimates identified by the longitudinal models for HIV notification and unemployment rates, 2002–2012. Cluster 3 was used as the reference category

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard error</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evolution in time of the HIV notification rate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>35.43</td>
<td>4.59</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Cluster 1</td>
<td>–25.81</td>
<td>5.03</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>–20.16</td>
<td>5.21</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Time (year–1)*</td>
<td>–1.39</td>
<td>0.52</td>
<td>0.008</td>
</tr>
<tr>
<td>Time*Cluster 1</td>
<td>0.97</td>
<td>0.57</td>
<td>0.091</td>
</tr>
<tr>
<td>Time*Cluster 2</td>
<td>0.86</td>
<td>0.59</td>
<td>0.145</td>
</tr>
<tr>
<td>Evolution in time of unemployment rates</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>73.792.49</td>
<td>8 022.79</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Cluster 1</td>
<td>–66 458.02</td>
<td>8 757.08</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>–53 225.60</td>
<td>9 025.07</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Time (year–1)*</td>
<td>5 869.30</td>
<td>511.72</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Time*Cluster 1</td>
<td>–2.60</td>
<td>0.46</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Time*Cluster 2</td>
<td>1.34</td>
<td>0.49</td>
<td>0.007</td>
</tr>
</tbody>
</table>

* Years 2002, 2003,… 2012 were coded 1, 2,…11.

HIV = human immunodeficiency virus.

Table A.3  Estimates from the multiple model for the evolution in time of tuberculosis incidence rates, 2002–2012. Cluster 3 was used as the reference category

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard error</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>41.48</td>
<td>4.41</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>HIV notification rate</td>
<td>0.20</td>
<td>0.08</td>
<td>0.010</td>
</tr>
<tr>
<td>Unemployment rate</td>
<td>0.001</td>
<td>0.0005</td>
<td>0.002</td>
</tr>
<tr>
<td>Cluster 1</td>
<td>–26.09</td>
<td>3.67</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>–16.84</td>
<td>3.45</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Time (year–1)*</td>
<td>–2.60</td>
<td>0.46</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Time*Cluster 1</td>
<td>1.34</td>
<td>0.49</td>
<td>0.007</td>
</tr>
<tr>
<td>Time*Cluster 2</td>
<td>0.96</td>
<td>0.5</td>
<td>0.056</td>
</tr>
</tbody>
</table>

* Years 2002, 2003,… 2012 were coded 1, 2,…11.

HIV = human immunodeficiency virus.
Figure A.1  Time evolution of HIV notification rates (per 100,000 population) by cluster and region, 2002–2012. Source: INSA = National Health Institute Doutor Ricardo Jorge, Lisbon, Portugal. HIV = human immunodeficiency virus.

Figure A.2  Time evolution of unemployment rates (per 100,000 population) by cluster and region, 2002–2012. Source: Instituto do Emprego e Formação Profissional, Lisbon, Portugal.
Figure A.3  Expected values for TB incidence rates (per 100,000 population) for the sample averages of the HIV notification and unemployment rates (per 100,000 population) by cluster, 2002–2012. TB = tuberculosis; HIV = human immunodeficiency virus.
CADRE : Le Portugal reste le pays d’Europe de l’Ouest qui a l’incidence de tuberculose (TB) la plus élevée.

OBJECTIF : Identifier les tendances longitudinales de l’incidence de la TB au Portugal de 2002 à 2012 et étudier l’effet longitudinal des indicateurs sociodémographiques et sanitaires de la population résidente sur le taux d’incidence de la TB.

MÉTHODE : Nous avons utilisé les données du système national de surveillance de la TB et d’autres institutions nationales. L’algorithme de regroupement longitudinal des moyennes a été réalisé sur le profil d’incidence dans le temps de la TB dans les districts du Portugal.

RÉSULTATS : Trois profils longitudinaux relatifs au taux d’incidence de la TB au Portugal ont été identifiés. Dans les trois cas, l’incidence de la TB a décru avec le temps. Parmi tous les indicateurs sociodémographiques et sanitaires qui ont été étudiés, le taux de notification du virus de l’immunodéficience humaine (VIH) et le chômage se sont avérés avoir un effet positif significatif sur l’incidence de la TB. L’effet a, en particulier, été le plus élevé pour le taux de notification du VIH.

CONCLUSION : Notre étude soutient le fait que des stratégies combinées vis-à-vis de la TB et du VIH et l’amélioration des déterminants sociaux peuvent contribuer à la diminution de l’incidence de la TB.

MARCO DE REFERENCIA: El Portugal sigue siendo el país con la más alta incidencia de tuberculosis (TB) en Europa occidental.

OBJETIVOS: Determinar la evolución longitudinal de la incidencia de TB en el Portugal del 2002 al 2012 e investigar el efecto longitudinal de los determinantes sociodemográficos y los relacionados con la salud de la población residente, en la tasa de incidencia de TB.

MÉTODOS: Se utilizaron los datos del Sistema Nacional de Vigilancia de la Tuberculosis y de otras instituciones nacionales. Se aplicó el algoritmo de partición K-medias para el análisis de los datos longitudinales de las curvas de incidencia frente al tiempo en los distritos de Portugal.

RESULTADOS: Se detectaron tres patrones longitudinales en la tasa de incidencia de TB. En los tres casos la incidencia disminuyó con el transcurso del tiempo. De todos los factores sociodemográficos y relacionados con la salud que se estudiaron, la tasa de notificación de la infección por el virus de la inmunodeficiencia humana (VIH) y el desempleo ejercieron un efecto significativo positivo en la incidencia de TB. Los efectos de mayor magnitud correspondieron a la tasa de notificación de la infección por el VIH.

CONCLUSION: Los resultados del estudio respaldan la pertinencia de la utilización de las estrategias integradas de atención de la TB y la infección por el VIH y el mejoramiento de los determinantes sociales.