Spatiotemporal Analysis of Lung Cancer Incidence and Case Fatality in Villa Clara Province, Cuba

Norma E. Batista MD and Oscar A. Antón MD

INTRODUCTION
Cancer has historically been a main cause of death in Cuba, with lung cancer the number one cause of cancer death in both sexes. Cancer morbidity and mortality rates are the basic measures of cancer impact in the community. Cancer mortality has been one of the major applications of geographic analysis and has made important progress in recent decades thanks to access to mortality statistics and to development and availability of geographic information systems. Cuba does not have a strong tradition of etiologic research using spatial analysis. High levels of lung cancer morbidity and mortality in Villa Clara and growing interest in spatial analysis as an epidemiologic tool motivated this study.

OBJECTIVE
To identify spatial and/or spatiotemporal clusters of lung cancer morbidity and case fatality in the province of Villa Clara, and to demonstrate the value of cluster analysis as an epidemiologic tool.

METHODS
Descriptive observational study based on administrative data, using the technique of space-time scan statistics. The study focused on new cases diagnosed in 2004 and case-fatality for those cases through 2009. Variables used were: cases diagnosed, deaths, date of diagnosis, date of death, municipality and Cartesian geocoding for each municipality.

RESULTS
The study identified significant spatial and spatiotemporal clusters of greater than expected lung cancer incidence (municipalities of Encrucijada, Camagüey, Sagua la Grande, Caibarién and Santa Clara) and case fatality (Encrucijada, Camagüey, Sagua la Grande, Caibarién, Santa Clara, Placetas and Manicaragua).

CONCLUSIONS
Although the results are not explanatory, the spatial and spatiotemporal patterns of excess lung cancer risk and case-fatality can support hypothesis generation for research and eventual interventions for targeted prevention and management.

KEYWORDS
Pulmonary neoplasms, epidemiological methods, spatiotemporal analysis, cluster analysis, space-time clustering, incidence, case fatality, Cuba
Environmental tobacco smoke (ETS), also known as second-hand smoke, has been classified by the International Agency for Research on Cancer as a Group 1 carcinogen, that is, there is convincing evidence that it produces cancer in humans.[17] A non-smoker’s risk of contracting lung cancer has been shown to increase 20%–30% when exposed to ETS. ETS also produces an increase in all-cause death rates, as well as changes in the immune system.[17–20] Among the 41 countries that reported data on population exposure to ETS in 2002, Cuba ranked in the top five, and in the top 24 in a 2006 report on 82 countries.[21,22]

In 2001 in Villa Clara, 45%–49.9% of families were exposed to ETS. In Cuba, research indicates that persons with higher educational levels may tend to underestimate ETS risk, constituting failed responsibility for protection of their own and others’ health. The extent of this problem in Cuba denotes a knowledge gap that could compromise future population health outcomes in the country.[23]

The importance of identifying high-risk areas for intervention explains the current keen interest among epidemiologists in spatiotemporal distribution of diseases, and in particular, cluster analysis. A cluster is defined as the occurrence of a significantly greater number of cases than expected in a specific geographic area or specific time period, or both. Epidemiological demonstration of a cluster does not necessarily imply causal association between the possible risk factor and the disease. An etiologic study is required to confirm association when evidence indicates ill persons have been exposed to a common risk factor.[26]

For population health problems, characterizing their geographic and spatial differentiation reveals patterns that can provide important clues to causation and identification of systemic problems that may require intervention. Integrating such information into the public health surveillance system can offer important evidence for health service management and decisionmaking. Thus, a population group’s lifestyle and behaviors can determine exposure patterns, concentrating or excluding risk factors. Spatial aggregation or clustering can provide evidence of shared risk factors, and identify regions where strong patterns predominate of common ethnicity, eating habits, soil type, agricultural activities, climate, social organization, inter alia.[24]

The study reviewed records of patients residing in the province of Villa Clara aged >15 years who were diagnosed with lung cancer (387) during the period from January 1 to December 31, 2004; and among these, deaths occurring through 2009 (311).

**Data source and analysis** Data on lung cancer incidence in 2004 and deaths in these cases from 2004 through 2009 obtained from databases of the Provincial Health Statistics Bureau, Villa Clara. Incidence rates were calculated per 10,000 population using the number of lung cancer cases diagnosed in 2004 and the census population reported for that year. Case fatality was calculated as percentage of patients diagnosed in 2004 who died before the close of 2009.

To generate records required by the program on geographic coordinates, populations and cases, various databases were used:

- Information on lung cancer cases diagnosed in 2004, and the outcome of these through study termination on December 31, 2009, from Villa Clara’s Provincial Health Statistics Bureau;
- Information on total population from the National Statistics Bureau[7];
- Records on geocoding (Cartesian) of the province’s municipalities from MapInfo v. 8.5.

To determine the existence of clusters by spatial analysis unit, Kulldorff’s space-time scan statistic (SaTScan Program v. 7.01) was used.[37] The main advantages of this statistical tool include: it prevents selection bias; for each cluster a unique value of p is obtained, which is adjusted by the multiple tests derived from the large quantity of potential clusters evaluated; if the null hypotheses is rejected, the methodology indicates the approximate cluster site responsible.

Each type of analysis has a different purpose: spatial analysis disregards time and detects anomalous concentrations in space, while spatiotemporal analysis detects clusters coinciding in space and time. The space-time scan statistical method explained in detail by Kulldorff creates a type of circular window (space) or cylindrical window (space-time) with a circular base of varying radius, corresponding to a geographical area; the cylinder’s height corresponds to time. These are centered
around one of many possible centroids in the study, with each circle or cylinder representing a possible cluster. Each cylinder has a specific number of observed cases and the program calculates the number of expected cases under the null hypothesis for the total risk population.\[38\] We set $\alpha$ at 0.05 for rejecting the null hypothesis of ‘no cluster.’

Relative risk (RR) estimates reported denote risk of incidence of cancer or mortality from cancer in the cluster, compared with risk in the rest of the province. Tables were created to display clusters with a statistically significant (p < 0.05) excess of observed over expected cases, under the hypothesis of a Poisson distribution and using the likelihood ratio for parameter estimation. Cartographic representations were made with MapInfo v. 8.5, importing statistical results into the geographic information system. Municipalities forming part of significant spatiotemporal clusters were colored and those in spatial clusters identified by a star.

**RESULTS**

**Incidence** Overall lung cancer incidence for the province was 4.8 per 10,000 population. Seven municipalities had higher rates: Caibarién (7.3), Camajuaní (6.3), Placetas (6.0), Sagua la Grande (5.9), Quemado de Gúines (5.3), Encrucijada (5.1) and Cifuentes (5.0) (Table 1).

**Spatial analysis** Municipalities at highest risk are displayed in Figure 1. A maximum-likelihood spatial cluster was identified that included the municipalities of Encrucijada, Camajuaní, Cifuentes, Sagua and Caibarién, with 133 cases observed, compared to 106.5 expected (p = 0.044).

**Spatiotemporal analysis** A significant spatiotemporal cluster was detected that included the municipalities of Cifuentes, Encrucijada, Sagua la Grande and Santa Clara. During December 2004, there were 32 cases observed versus 14.5 expected (p = 0.030), that is, more than twice the risk observed in the rest of the province (Figure 1).

**Table 1:** Lung cancer incidence (2004) and case fatality (2004–2009), Villa Clara, Cuba

<table>
<thead>
<tr>
<th>Municipality</th>
<th>Cases 2004 (No.)</th>
<th>Incidence (per 10,000 population)</th>
<th>Deaths 2004–2009 (No.)</th>
<th>Case fatality (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corralillo</td>
<td>1</td>
<td>0.4</td>
<td>1</td>
<td>100.0</td>
</tr>
<tr>
<td>Quemado de Gúines</td>
<td>12</td>
<td>5.3</td>
<td>9</td>
<td>75.0</td>
</tr>
<tr>
<td>Sagua la Grande</td>
<td>32</td>
<td>5.9</td>
<td>27</td>
<td>84.4</td>
</tr>
<tr>
<td>Encrucijada</td>
<td>17</td>
<td>5.1</td>
<td>14</td>
<td>82.4</td>
</tr>
<tr>
<td>Camajuaní</td>
<td>40</td>
<td>6.3</td>
<td>28</td>
<td>70.0</td>
</tr>
<tr>
<td>Caibarién</td>
<td>28</td>
<td>7.3</td>
<td>26</td>
<td>92.9</td>
</tr>
<tr>
<td>Remedios</td>
<td>17</td>
<td>3.7</td>
<td>13</td>
<td>76.5</td>
</tr>
<tr>
<td>Placetas</td>
<td>42</td>
<td>6.0</td>
<td>35</td>
<td>83.3</td>
</tr>
<tr>
<td>Santa Clara</td>
<td>107</td>
<td>4.5</td>
<td>78</td>
<td>72.9</td>
</tr>
<tr>
<td>Cifuentes</td>
<td>16</td>
<td>5.0</td>
<td>14</td>
<td>87.5</td>
</tr>
<tr>
<td>Santo Domingo</td>
<td>23</td>
<td>4.3</td>
<td>20</td>
<td>87.0</td>
</tr>
<tr>
<td>Ranchuelo</td>
<td>24</td>
<td>4.2</td>
<td>20</td>
<td>83.3</td>
</tr>
<tr>
<td>Manicaragua</td>
<td>28</td>
<td>3.9</td>
<td>26</td>
<td>92.9</td>
</tr>
<tr>
<td>Total</td>
<td>387</td>
<td>4.8</td>
<td>311</td>
<td>80.4</td>
</tr>
</tbody>
</table>

**Table 2:** Lung cancer case fatality (2004–2009), Villa Clara, Cuba

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Municipality</th>
<th>Time</th>
<th>Deaths observed</th>
<th>Deaths expected</th>
<th>RR</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial</td>
<td>Caibarién, Camajuaní, Cifuentes, Encrucijada, Sagua la Grande</td>
<td>—</td>
<td>133</td>
<td>106.5</td>
<td>1.37</td>
<td>0.044</td>
</tr>
<tr>
<td>Spatiotemporal</td>
<td>Cifuentes, Encrucijada, Sagua la Grande, Santa Clara</td>
<td>Dec 2004</td>
<td>32</td>
<td>14.5</td>
<td>2.31</td>
<td>0.030</td>
</tr>
</tbody>
</table>

**Figure 1:** Spatial and spatiotemporal lung cancer incidence clusters in Villa Clara, Cuba (2004)

**Figure 2:** Spatial and spatiotemporal lung cancer case fatality clusters in Villa Clara, Cuba (2004–2009)

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Municipality</th>
<th>Time</th>
<th>Deaths observed</th>
<th>Deaths expected</th>
<th>RR</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial</td>
<td>Caibarién, Camajuaní, Cifuentes, Encrucijada, Sagua la Grande</td>
<td>—</td>
<td>109</td>
<td>85.6</td>
<td>1.42</td>
<td>0.055</td>
</tr>
<tr>
<td>Spatiotemporal</td>
<td>Caibarién, Camajuaní, Cifuentes, Encrucijada, Sagua la Grande</td>
<td>Mar 1, 2004–Jun 30, 2005</td>
<td>87</td>
<td>19.0</td>
<td>5.96</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>Manicaragua, Placetas, Santa Clara</td>
<td>Apr 1, 2004–Aug 31, 2005</td>
<td>104</td>
<td>34.5</td>
<td>4.03</td>
<td>0.001</td>
</tr>
</tbody>
</table>
Case fatality The provincial case fatality rate for 2004–2009 was 80.4% (Table 1). The municipalities with highest case fatality (not counting Corralillo, where only one case was diagnosed in 2004) were Caibarién and Manicaragua (92.9%), Cifuentes (87.5%) and Santo Domingo (87.0%).

Spatial analysis A single spatial cluster was found that included the municipalities of Encrucijada, Camajuaní, Cifuentes, Sagua la Grande and Caibarién, with a relative risk of 1.42 (Figure 2).

Spatiotemporal analysis The same municipalities as in the spatial cluster (Encrucijada, Camajuaní, Cifuentes, Sagua la Grande and Caibarién) formed a spatiotemporal cluster for the period of March 1, 2004–June 30, 2005, with 87 deaths versus 19 for an RR of 5.96. Another spatiotemporal cluster was identified in Manicaragua, Santa Clara and Placetas municipalities; the RR for the period April 1, 2004–August 31, 2005 was 4.03, with 104 deaths versus 34.5 expected (Figure 2).

DISCUSSION

One constraint in this investigation was the lack of similar studies in the national literature for comparison purposes. However, several international studies have applied this method to lung cancer. One was conducted by Fukuda, who identified the principal incidence clusters in Japan’s metropolitan regions, related primarily to habits and socioeconomic conditions.[39] Another study conducted in Costa Rica identified high-incidence clusters of bronchopulmonary cancer in the country’s northeast and Atlantic regions, implying a relationship with major risk factors such as smoking and environmental pollution.[26]

Incidence The lung cancer incidence clusters identified in this study indicate the need for further research on risk factors—mainly tobacco use, which constitutes the most significant etiological agent. A study conducted by Varona reported that some 23% of all deaths from cancer in Cuba in 2010 were due to lung cancer, of which 85% were attributed to smoking (92% in men and 75% in women).[40] An analytical cross-sectional epidemiological study in Matanzas Province, Cuba found that patients who smoked were four times more likely to die from lung cancer.[41]

Population aging is one of the most important contributors to rising numbers of cancer cases, and Villa Clara Province has the oldest population in the country (18.2% aged ≥60 years).[42] All of the province’s municipalities are among the 40 oldest in the country (of a total of 169 municipalities).[7] Placetas, Cifuentes, Camajuaní and Encrucijada have the oldest populations in Villa Clara and are also where higher incidence and case fatality clusters were identified in this study.

Along with exposure to risk factors, genetic mutations predisposing to lung cancer may increase the probability of developing the disease. There is a hypotheses on varying susceptibility to carcinogens, based on an individual’s ability to form genotoxic intermediates, detoxify these intermediates and repair the damage to DNA, all inheritable genetic features that would affect potential for suffering lung cancer and other cancers as well.[43]

Family aggregations can also be explained by the fact that individuals in the same family tend to have similar habits, such as smoking, drinking, diet and occupation, hence the value of studying clusters.[44]

The spatiotemporal incidence cluster in December 2004 could be due to year-end reporting procedures or other factors that warrant further examination in future studies. With a disease of such long latency, a biological explanation is implausible.

Case fatality Spatial analysis revealed close geographic overlap of incidence and case fatality (see Figures 1 and 2). However, case fatality patterns do not depend on incidence but are influenced more by stage at diagnosis and by treatment applied, both of which can affect patient survival.

Both spatiotemporal case fatality clusters were within two years of diagnosis, consistent with lung cancer’s established poor prognosis, one of the worst for any cancer. It is asymptomatic over most of its evolution and, in most patients, by time of diagnosis, the disease is already well advanced and nonresponsive to most treatments.[40,45] Clearly screening and close examination of at-risk groups in primary health care are essential. This can overcome obstacles at various levels: individual (such as awareness), primary care (identification of at-risk individuals, streamlining of diagnosis and prompt referral to specialized care), and secondary care (prompt application of clinical protocol for diagnosed cases).[46] A study in Villa Clara Province found a direct association between patient survival and specific aspects of treatment: early tumor stage at diagnosis, morphologic and histologic diagnosis, and diagnostic methods used.[47] The finding that Placetas Municipality did not form part of any incidence clusters but did appear in a spatiotemporal case fatality cluster merits further study, as it could suggest health system factors at play, such as delays in diagnosis or errors in treatment protocols.

The importance of this type of study resides in its ability to raise questions and provide clues that may lead to causal explanations further down the road. Use of this tool in the surveillance system can help increase health equity and system efficiency by focusing medical services where they are most needed, and of course can help set priorities for research.

CONCLUSIONS

Application of the SaTScan analytical method revealed significant spatial and spatiotemporal lung cancer clusters, in both incidence and case fatality. While our findings are not explanatory, they do identify areas at higher risk and thus can support generation of hypotheses for research and eventual intervention.

Efforts are needed to determine factors contributing to the observed excess risk in areas of Villa Clara, to prevent the disease, initiate earlier detection and treat it more effectively, as well as to identify systemic factors that could account for differences in case fatality. Bringing about a major change in lung cancer burden will require improvements across the entire spectrum of prevention, early diagnosis and treatment.

It would also be valuable to extend application of this tool to epidemiological surveillance systems in other health jurisdictions in Cuba, to deepen epidemiological exploration of population health status, identify possible contextual determinants to facilitate better risk factor control, and seek solutions in health systems and services to problems identified through cluster detection. Such an investment would strengthen health and disease surveillance and provide a valuable addition to Cuba’s universal health care system.
THE AUTHORS
Norma E. Batista Hernández (Corresponding author: normabh@ucm.vcl.sld.cu), physician specializing in biostatistics and family medicine, biomedical research unit, Medical University of Villa Clara, Santa Clara, Cuba.

Oscar A. Antón Fleites, family physician, Marta Abreu University Teaching Polyclinic, Santa Clara, Cuba.

Submitted: January 5, 2012
Approved for publication: July 12, 2013
Disclosures: None