

# **Análisis estadístico espacial para la identificación de conglomerados de cáncer de mama en la ciudad de La Paz, B. C. S.**

***Martín Agúndez Amador***

Instituto Tecnológico de La Paz, Baja California Sur, México  
*magundez@uabcs.mx*

***Carlos Sánchez Saldaña***

Instituto Tecnológico de La Paz, Baja California Sur, México

***Guillermo Martínez Flores***

Centro Interdisciplinario de Ciencias Marinas, Instituto Politécnico Nacional  
La Paz, Baja California Sur, México  
*gmflores@ipn.mx*

***Reyna de Jesús Romero Geraldo***

Instituto Tecnológico de La Paz, Baja California Sur, México

***Jorge Enrique Luna Taylor***

Instituto Tecnológico de La Paz, Baja California Sur, México

## **Resumen**

Estadísticas de la Organización Mundial de la Salud (OMS), muestran que el cáncer se encuentra entre las primeras causas de muerte a nivel mundial; en 2012 se le atribuyeron 8.2 millones de muertes. Entre los tumores malignos que afectan a la población femenina, se destaca el cáncer de mama. Es motivo de análisis, el hecho de que en los estados del norte del país, entre los que se encuentra Baja California Sur, este concentrada la mayor tasa de incidencias del cáncer de mama. La realización de

estudios sobre la distribución espacial de esta enfermedad, apoyados en Sistemas de Información Geográfica y herramientas de análisis espacial, puede resultar muy importante para respaldar decisiones de políticas en campos como la salud pública y educación.

En este trabajo se realizó un estudio basado en análisis estadístico espacial, para identificar posibles conglomerados de cáncer de mama en la ciudad de La Paz, B. C. S. Los resultados de este estudio muestran la presencia de conglomerados significativos de alto y bajo riesgo relativo de la enfermedad, y permiten generar hipótesis iniciales sobre los posibles factores de riesgo que incrementan la incidencia de cáncer en la localidad.

**Palabras Clave:** Análisis espacial, cáncer de mama, conglomerados, sistemas de información geográfica.

### **Abstract**

*Statistics from the World Health Organization (WHO) show that cancer is among the leading causes of death worldwide; in 2012 it was attributed 8.2 million deaths. Among the cancers that affect the female population, breast cancer is highlighted. It is a matter of analysis, the fact that in the northern states of the country, It is a matter of analysis, the fact that in the northern states of the country, among which is Baja California Sur, this concentrated the highest rate of incidence of breast cancer this concentrated the highest rate of incidence of breast cancer. The studies on the spatial distribution of the disease, supported by Geographic Information Systems (GIS) and spatial analysis tools, can be very important to support policy decisions in fields such as public health and education.*

*This paper presents a study based on spatial statistical analysis was performed to identify possible clusters of breast cancer in the city of La Paz, BCS. The results of this study show the presence of significant clusters of high and low relative risk of disease,*

*and can generate initial hypotheses about possible risk factors that increase the incidence of cancer in the town.*

**Keywords:** *Spatial analysis, breast cancer, clusters, geographic information systems.*

## **1. Introducción**

El cáncer de mama es uno de los tipos de cáncer de mayor incidencia a nivel mundial. Según la OMS, cada año se detectan 1.38 millones de nuevos casos y fallecen 458 mil personas por esta causa. Se estima que para el año 2030, esta cifra se elevará un 46%. Por otra parte, de acuerdo al Instituto Nacional de Estadística y Geografía (INEGI, 2014), la tasa de mortalidad observada por cáncer de mama, durante los últimos 6 años en México, es de 15 casos por cada 100 mil mujeres. Observándose que es en los estados del norte del país es donde se presentan las mayores tasas de incidencia de este padecimiento. Esta distribución geográfica hace suponer que deben existir factores ambientales, o de estilo de vida, en estas poblaciones que favorezcan la incidencia de esta enfermedad.

Se estima que el 67% de los casos de cáncer están ligados a algún factor ambiental o de estilo de vida (NCI, 2013). Un substancial conjunto de evidencias científicas indican que la exposición a químicos y a la radiación contribuye a incrementar la incidencia de cáncer de mama (Gray et al., 2009). Se puede pensar en relaciones que varían a lo largo del espacio tales como el uso de plaguicidas o cualquier otra sustancia relacionada con el ambiente; así como en situaciones de distancia social y económica que pueden estar relacionadas al riesgo de cáncer de mama (Santamaría-Ulloa, 2009).

### **1.1 Medio ambiente y distribución espacial**

La identificación de patrones espaciales en eventos de la salud es fundamental para comprender la exposición y prevenir eventos futuros, más allá de si el proceso que subyace es contagioso, influenciado por el ambiente o relacionado a la variabilidad

genotípica (Alegret, 2006). El estudio de la distribución espacial de las enfermedades y de la mortalidad, es una aproximación inicial a sus posibles causas (López-Abente and Ibáñez, 2002). A través de estos estudios se puede obtener información de cómo es que influye la falta de acceso a los servicios médicos de ciertas poblaciones en la mortalidad por cáncer (Alegret, 2006).

Este tipo de estudios, apoyados en Sistemas de Información Geográfica (SIG) y herramientas de análisis espacial, son aplicables a muchos campos, y resultan muy importantes para respaldar decisiones de políticas en campos como la salud pública, educación, cambios demográficos y otros. La oportuna vigilancia y respuesta informada a este tipo de eventos de salud puede reducir sustancialmente los resultados de costos humanos, financieros y sociales de una población. Actualmente se tiene suficiente conocimiento de las causas de prevención del cáncer, como para reducir los casos en un 50% en las próximas décadas (Colditz et al., 2002). Sin embargo, las distribuciones espaciales de los problemas de salud han sido poco estudiadas y no se han explotado las potencialidades de análisis espacial en función de un mejor desempeño de las investigaciones y toma de decisiones en salud (Alegret, 2006).

## **1.2 Conglomerado de cáncer**

Un conglomerado es la ocurrencia real y significativa de más casos que los esperados, basados en un sustento científico (CDC, 2001). Con el fin de determinar si existe un conglomerado real, se utiliza la estadística de exploración espacio-temporal, en la cual convergen diferentes ramas de la ciencia como son la geografía, epidemiología y estadística (Dunn et al., 2001); grupo en el que se puede incluir a las ciencias de la computación como una herramienta muy valiosa para la implementación y ejecución de los métodos de almacenamiento, representación gráfica y análisis estadístico diseñados.

Para lograr detecciones y caracterizaciones precisas de los conglomerados, se debe de integrar información de múltiples fuentes de datos espaciales y temporales, con el fin de obtener un conocimiento completo y coherente de la situación (Neill and Cooper, 2010).

En caso de demostrarse la existencia de un conglomerado de cáncer, se justifica la utilización de recursos en la ejecución de una investigación de tipo epidemiológica (Santamaría-Ulloa, 2002). El empleo de este tipo de método de exploración, ha mostrado la presencia de conglomerados significativos tanto espaciales como espacio-temporales. Aunque éstos no permiten establecer las causas del cáncer, orienta sobre las áreas con mayor riesgo de enfermar y de morir, y permite la generación de hipótesis sobre posibles determinantes que ayuden a orientar nuevas investigaciones (Batista et al., 2012).

El objetivo de este trabajo fue el de realizar en la Ciudad de La Paz, B. C. S., un análisis estadístico espacial sobre el cáncer de mama.

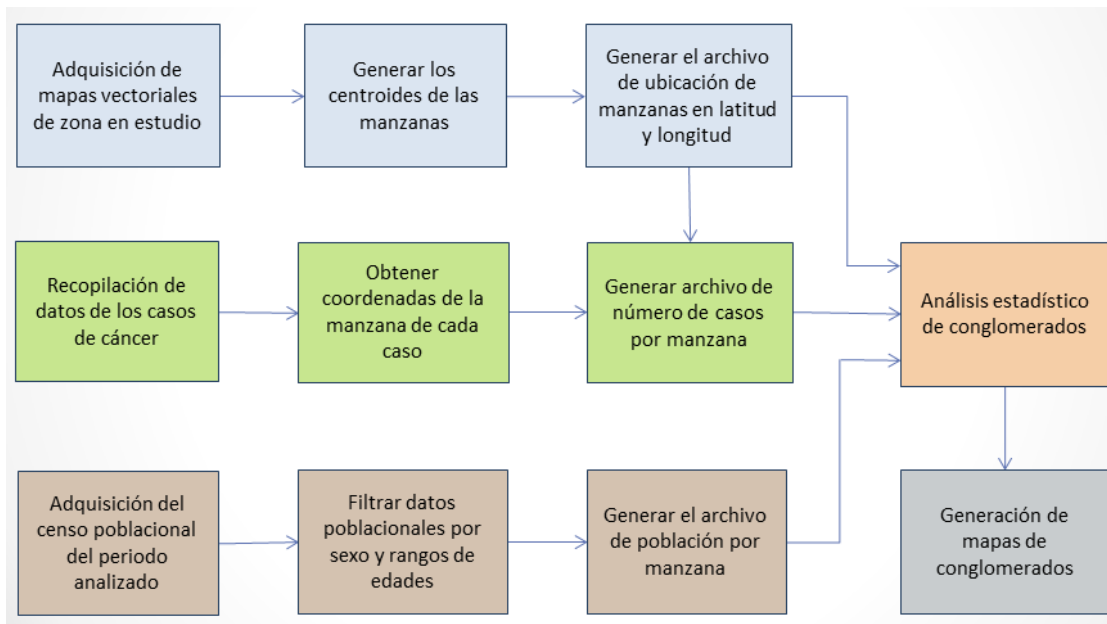


Figura 1. Flujo de actividades y procesos realizados.

## 2. Datos y métodos

El desarrollo de esta investigación se dividió en un conjunto de actividades y procesos que incluyen desde la recopilación de los datos, estructuración y procesado de los

mismos; para posteriormente realizar la interpretación y análisis de los resultados obtenidos. La Figura 1 muestra la secuencia de las actividades y procesos desarrollados.

## **2.1 Fuentes de datos**

Los datos oncológicos requeridos para este estudio fueron proporcionados por el Centro Estatal de Oncología (CEO). Se consideraron los casos de cáncer de mama en la población femenina mayor de 24 años de la Ciudad de La Paz, B.C.S., que fueron diagnosticados entre el 2006 y 2014.

Los mapas vectoriales de la zona de estudio, así como los datos del censo poblacional requeridos para el análisis estadístico, fueron proporcionados por el Instituto Nacional de Estadística y Geografía (INEGI). El censo utilizado para este estudio corresponde al año 2010.

## **2.2 Obtención de las coordenadas espaciales de los casos de cáncer.**

Uno de los datos más importantes requeridos por el software de análisis estadístico, son las coordenadas espaciales de cada uno de los casos. Con este propósito se visitó cada una de las direcciones de los casos de cáncer, y se capturó su coordenada espacial con ayuda de un dispositivo GPS.

## **2.3 Métodos para el análisis y presentación de datos**

Las herramientas de software utilizadas para el análisis y presentación de los datos son: Sistema estadístico de exploración espacio temporal *SaTScan versión 9.4.1*, y el Sistema de información geográfica *QGIS versión 2.8.1*.

El programa SaTScan recibe tres archivos como entrada de datos: archivo de casos (.cas), archivo de coordenadas (.geo) y archivo de población (.pop). Para la generación de estos archivos se realizó un procesamiento de los datos fuentes a través de diversas herramientas computacionales como: editores de texto, hojas de cálculo, funciones en MATLAB y pequeños módulos programados en C#.

A continuación se describe cada uno de estos archivos, y se explica el proceso de su creación.

### 2.3.1 Archivo de casos (.cas)

Para generar el archivo de casos se desarrolló un módulo programado en lenguaje C#. Este programa realiza el conteo de pacientes por cada manzana. Para ello, se calcula la distancia más cercana entre el domicilio de cada paciente y cada uno de los centroides de las manzanas de la ciudad., cuyas respectivas ubicaciones están en coordenadas geográficas (longitud y latitud). Para este proceso se reciben como datos de entrada un archivo con las coordenadas domicilio del paciente y el identificador del caso, y un segundo archivo con las coordenadas, y el identificador de cada manzana. Como resultado se genera un tercer archivo con el formato de entrada requerido por el programa SatScan. La Figura 2 muestra una pantalla del programa en ejecución.

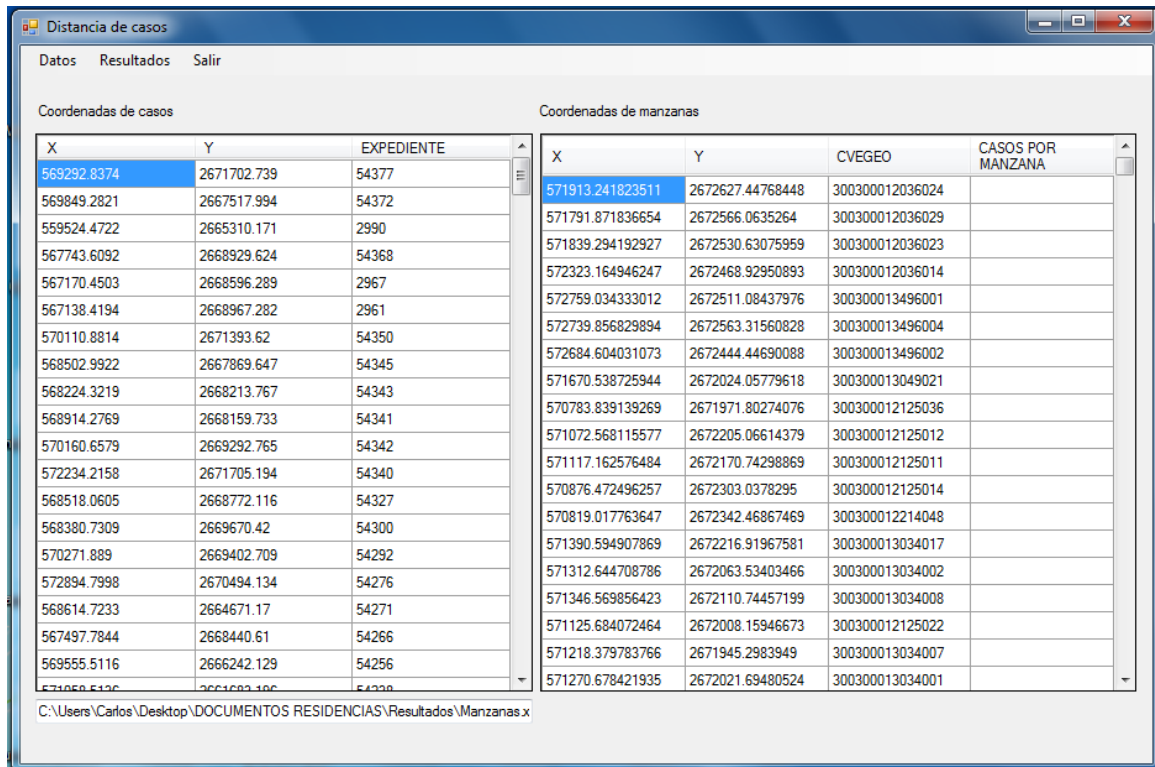


Figura 2. Programa para generar el total de casos por manzana.

La función para el cálculo de la distancia, entre la ubicación de un paciente y el centroide de cada manzana, recibe como parámetros las coordenadas de los casos y de las manzanas, y devuelve la distancia euclidiana entre ambas. Para ello fue necesario proyectar las coordenadas al sistema de coordenadas métrico y plano UTM (Universal Transverse of Mercator).

El Algoritmo 1 corresponde a la función principal que realiza el conteo de los pacientes por manzana. Esta función recorre cada caso, y por cada uno, recorre a su vez cada manzana, buscando aquella cuyo centroide resulte más cercano al caso correspondiente, a la cual se le acumula dicho caso.

---

**Algoritmo 1.** Cálculo del total de casos por manzana.

---

**Entrada:** Coordenada y número de expediente de los casos; coordenada y clave de las manzanas.

**Salida:** Coordenada y clave de las manzanas con el número de casos acumulados.

1. **Recorre** con  $i$  cada caso
2.      $distanciaMenor = calculoDist (coordenada\_caso\ i, coordenada\_manzana\ 1)$
3.      $manzanaCercana = 1$
4.     **Recorre** con  $j$  cada manzana
5.          $distancia = calculoDist (coordenada\_caso\ i, coordenada\_manzana\ j)$
6.         **Si**  $distancia < distanciaMenor$
7.              $distanciaMenor = distancia$
8.              $manzanaCercana = j$
9.     **Fin Si**
10.  **Fin Recorre**
11.  Acumula caso  $i$  a la manzana  $manzanaCercana$
12. **Fin Recorre**
13. Devuelve datos de las manzanas con el total de casos acumulados



### 2.3.2 Archivo de coordenadas (.geo)

Para la creación del archivo de coordenadas se utilizó el mapa vectorial de la ciudad. Se requirió realizar una conversión del sistema de coordenadas geográficas al sistema UTM, utilizando el QGIS.



**Figura 3. Porción del mapa de la ciudad con los centroides generados en QGIS.**

Para generar los centroides de las manzanas sobre los mapas vectoriales, también se utilizó el programa QGIS (Figura 3).

### 2.3.3 Archivo de poblaciones (.pop)

Este archivo se generó a partir de los datos del censo del 2010 proporcionados por el INEGI. Fue necesario procesar los datos, con la finalidad de filtrar únicamente a las pacientes del sexo femenino mayores de 24 años, con domicilio en la ciudad de La Paz, B.C.S.

## 2.4 Análisis estadístico espacial

El análisis estadístico para la identificación de los conglomerados se realizó a través del programa SaTScan (Figura 4). Para los cálculos se utilizó el modelo de distribución de Poisson, se solicitó únicamente análisis espacial y se realizaron diversas ejecuciones del programa probando con diferentes porcentajes para el tamaño límite de los

conglomerados. Se realizaron también diferentes pruebas con tasas altas y bajas de riesgo relativo.

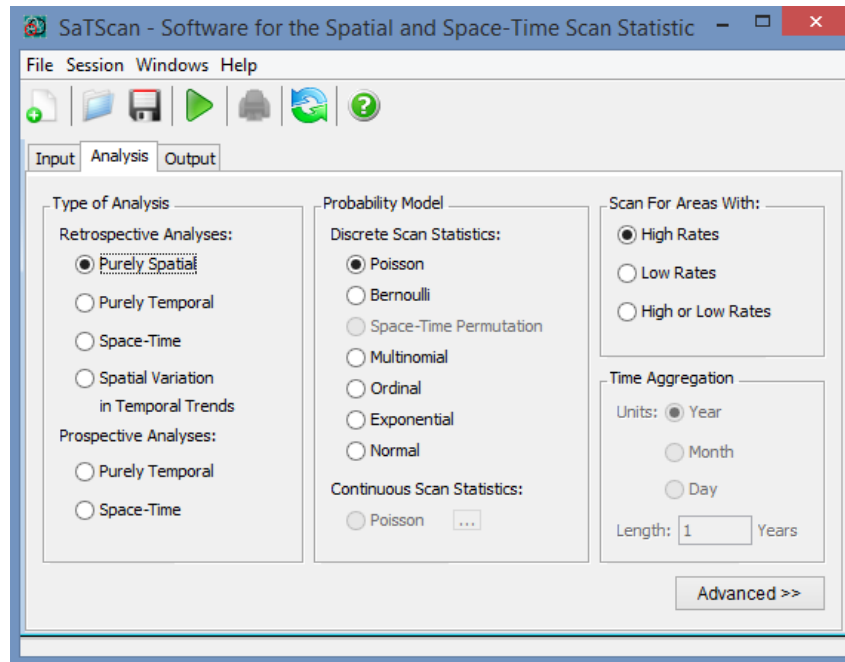


Figura 4. Ejecución del software SaTSan.

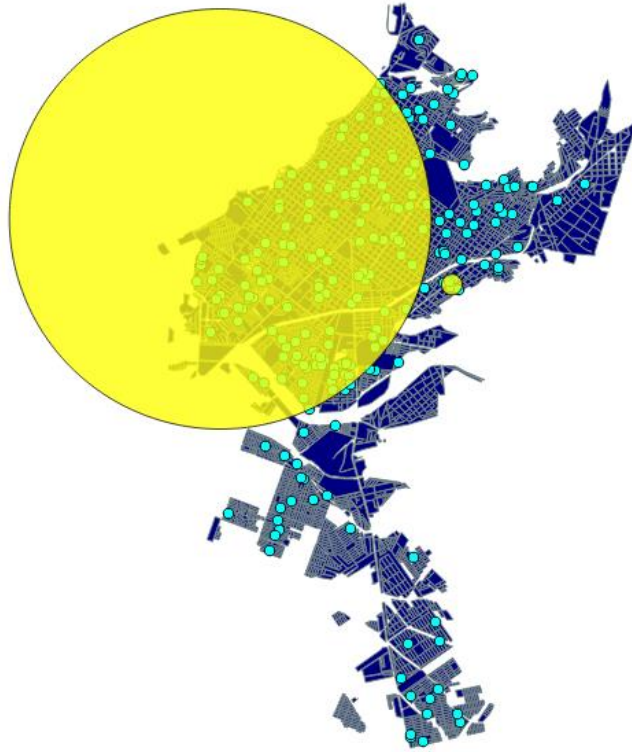
### 3. Resultados

Una vez realizado el análisis estadístico espacial y su correspondiente representación a través del sistema QGIS, se obtuvieron los siguientes resultados.

En la Figura 5 se presenta el mapa de la ciudad de La Paz, con el conglomerado descubierto de mayor riesgo relativo, y con mayor significancia estadística, marcado con un círculo de color amarillo. Este conglomerado se obtuvo utilizando un tamaño máximo de los conglomerados analizados de un 50% respecto al total de la población, bajo la opción de no traslape de conglomerados.

Este conglomerado corresponde a una zona de la ciudad donde de 29,747 mujeres en el rango de edad de estudio, 157 enfermaron de cáncer de mama. El factor de riesgo

relativo de esta zona es de 1.96, con un p-valor de 0.0047, lo cual corresponde a un porcentaje de confianza del 99.53%. La Figura 6 resume las estadísticas obtenidas por el programa SaTScan para este conglomerado.



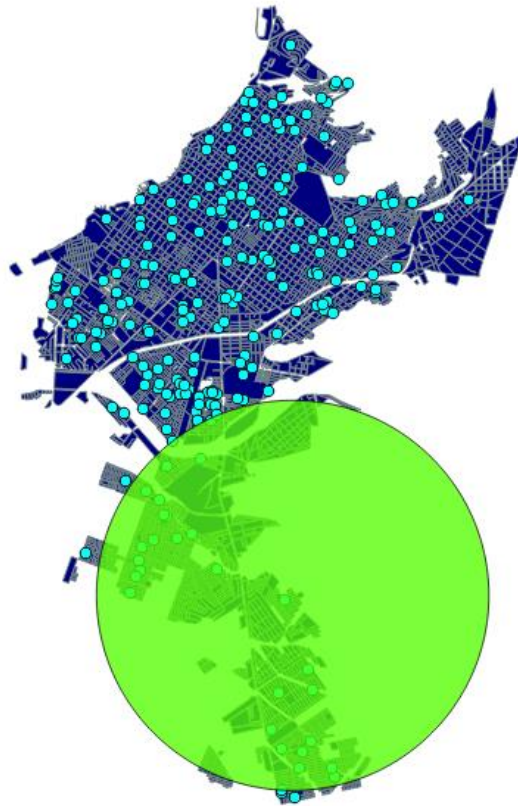
**Figura 5. Conglomerado de mayor riesgo relativo descubierto.**

Coordinates / radius..:	(24.146121 N, 110.339201 W) / 4.40 km
Population.....:	29747
Number of cases.....:	157
Expected cases.....:	117.09
Annual cases / 100000.:	528.1
Observed / expected....:	1.34
Relative risk.....:	1.96
P-value.....:	0.0047

**Figura 6. Estadística del conglomerado de mayor riesgo relativo.**

En la Figura 7 se presenta el mapa de la ciudad con el conglomerado de menor riesgo relativo descubierto por el sistema, marcado con un círculo de color verde. Este

conglomerado se localiza en las coordenadas 24.082 N, 110.298 W, y corresponde a una zona de la ciudad donde de 15,981 mujeres en el rango de edad de estudio, 26 enfermaron de cáncer de mama. El factor de riesgo relativo de esta zona es de 0.34, con un p-valor de 0.000060, lo cual corresponde a un porcentaje de confianza del 99.994%.



**Figura 7. Conglomerado de menor riesgo relativo descubierto.**

#### **4. Discusión**

Existen trabajos previos en la realización de estudios sobre la distribución espacial, apoyados en Sistemas de Información Geográfica y herramientas de análisis espacial para identificar posibles conglomerados de cáncer. En el trabajo titulado: *Aplicación de*

*técnicas de análisis espacial a la mortalidad por cáncer en Madrid* (Lopez-Abente and Ibañez, 2002), se identificaron dos conglomerados de casos de cáncer de pleura-peritoneo, así como también un exceso de riesgo de muerte por mesotelioma. En el trabajo realizado en Cuba: *Análisis espacio-temporal de eventos asociados al cáncer: una herramienta para apoyar estudios epidemiológicos* (Batista et al., 2012), se identificó un conglomerado que indicaba que los residentes de algunos municipios de la isla, presentaron un riesgo exacerbado en 16 veces para padecer cáncer de mama y cérvix en la población femenina, y cáncer de próstata en la población masculina. En cuanto a la mortalidad, se detectaron también varios conglomerados significativos.

En el presente trabajo realizado en la Ciudad de La Paz, B. C. S., se detectó un conglomerado de mayor riesgo relativo que corresponde a la parte más antigua de la ciudad de La Paz, B.C.S. Según el análisis estadístico, la población que vive en esta zona tiene un factor de riesgo de contraer la enfermedad de 1.96 veces mayor que la población que vive en otra parte de la ciudad. Esto con un valor de confianza del 99.53%.

Tratar de explicar este hecho, lleva a la generación de varias hipótesis. La hipótesis que consideramos más plausible, es la posibilidad de que por ser la parte más antigua de la ciudad, la edad promedio de la población que vive en esa zona sea mayor que la edad promedio del resto de la población que vive en colonias de más reciente creación. Es ampliamente conocido que la edad de las personas es uno de los principales factores de riesgo del cáncer. Sin embargo, esta y otras hipótesis tienen que ser validadas a través de su estudio correspondiente.

El conglomerado más significativo detectado con un bajo índice de riesgo relativo, corresponde a una zona donde el riesgo de contraer la enfermedad es apenas de 0.15 veces con relación al resto de la ciudad. Este valor fue detectado con un porcentaje de confianza mayor al 99.99%. Esta zona corresponde a las colonias más recientes de la ciudad, y la explicación de este hecho lo relacionamos con la hipótesis del promedio de edad, explicada anteriormente para los conglomerados de mayor riesgo.

Para corroborar o descartar la hipótesis asociada con el promedio de edad, se propone como trabajo próximo realizar el análisis de conglomerados por rangos de edades de 5 años. De esta manera se comparará el riesgo relativo que tiene la población de un mismo rango de edad. Si los factores de riesgo obtenidos anteriormente están relacionados principalmente con la edad, se espera que para un mismo rango de edad, el riesgo relativo de contraer la enfermedad sea similar, sin importar la zona de la ciudad donde se habite.

En caso de que el promedio de edad no resulte ser el factor dominante en la detección de estos conglomerados, y estos mantengan riesgos relativos muy elevados, se justificaría realizar una investigación del tipo epidemiológica donde se consideren factores de riesgo ambientales y conductuales como: contaminación de la zona, calidad del agua, cercanía a industrias, hábitos alimenticios, actividad física, entre otros.

## 5. Agradecimientos

Nuestro agradecimiento al Instituto Nacional de Estadística y Geografía (INEGI) y al Centro Estatal de Oncología (CEO) de La Paz, B. C. S., por las facilidades de información otorgadas para el desarrollo de esta investigación.

## 6. Bibliografía

- [1] Alegret, M. (2006). *Propuestas metodológicas para la incorporación más efectiva del análisis espacial en Ciencias de la Salud*. PhD thesis, Escuela Nacional de Salud Pública, 134 p.
- [2] Batista, N, E., Anton, O., Alegret, C. M. (2012). *Análisis espacio-temporal de asociados al cáncer: una herramienta para apoyar estudios epidemiológicos*. *MediSur*, 10(2):93-103.

- [3] CDC (Centers for Disease Control). (2001). *Guidelines form investigating cluster of health events*. Sitio Web:<http://www.cdc.gov/mmvr/preview/mmwrhtml/00001797.htm> (Accedido Julio 2013).
- [5] Colditz, G. A., Samplin-Salgado, M., Ryan, C.T., Dart, H., Fisher, L., Tokuda, A., and Rockhill, B. (2002). *Harvard report on cáncer prevention, volumen 5 fulfilling the potential for cáncer prevention: policy approaches*. *Cancer Causes and Control*, 13:199-212.
- [6] Dunn, C., S. Kingham, Rowlingson, B., Bhopal, R., Cockings, S., Foy, C., Acquilla, S., Halpin, J., Diggle, P., and Walker, D. (2001). *Analysing spatially referenced public health data: a comparison of three methodological approaches*. *Health Place*, 7(1):1-12.
- [7] Gray, J., Evans, N., Taylor, B., Rizzo, J., and Walker, M., (2009). *State of the evidence, the connection between breast cáncer and the environment*. *Int J Occup Environ Health*, 15:43-78.
- [8] INEGI (Instituto Nacional de Estadística y Geografía), (2014). *Estadísticas a propósito del día internacional del cáncer de mama*. Sitio Web:<http://www.inegi.org.mx/inegi/contenidos/espanol/prensa/Contenidos/estadisticas/2014/mama0.pdf> (Accedido Septiembre 2015).
- [9] Lopez-Abente, G. and Ibañez, C. (2002). *Aplicación de técnicas de análisis espacial a la mortalidad por cáncer en Madrid*. Technical Report 66, Instituto de Salud Carlos III.
- [10] NCI (National Cancer Institute), (2013). *Understanding Cancer Series*. Sitio Web: <http://www.cancer.gov/cancertopics/understandingcancer/environment/allpages> (Accedido Julio 2013).
- [11] Neil, D. B., and Cooper, G. F., (2010). *A multivariate bayesian scan statistic for early event detection and characterization*. *Mach Learn*, 79-261-282.

- [12] Santamaría-Ulloa, C., (2002). *Análisis para la determinación de conglomerados espacio temporales en la incidencia de cáncer, costa rica, 1990-1997*, Master's thesis, Universidad de Costa Rica.
- [13] Santamaría-Ulloa, C., (2009). *El impacto de la exposición a plaguicidas sobre la incidencia de cáncer de mama, evidencia de costa rica. Población y Salud en Mesoamérica*, 7(1):1-43.