


Análisis espacial en epidemiología: revisión de métodos

Spatial analysis in epidemiology: methodological review

Ana Maria Valbuena-García¹, Laura Andrea Rodríguez-Villamizar¹

Forma de citar: Valbuena-García AM, Rodríguez-Villamizar LA. Análisis espacial en epidemiología: revisión de métodos. Rev Uni Ind Santander Salud. 2018; 50(4): 358-365. doi: 10.18273/revsal.v50n4-2018009 

Resumen

El análisis espacial es un conjunto de técnicas que utiliza como insumos la ubicación geográfica o espacial de los eventos que se analizan. El uso de técnicas de análisis espacial en los estudios epidemiológicos ha tenido un rápido crecimiento en las últimas décadas porque permiten incluir en los análisis la variabilidad espacial. En el presente trabajo se hace una revisión general de los métodos de análisis espacial aplicados a la epidemiología y tiene como propósitos incentivar el estudio del análisis espacial y promover la incorporación de estas técnicas en las investigaciones actuales. Se describen los métodos de análisis espacial desde el abordaje descriptivo y analítico haciendo énfasis en el uso, ventajas y limitaciones de mapas coropléticos, pruebas de hipótesis y la identificación de clústers de enfermedades.

Palabras clave: Análisis espacial, mapeo geográfico, análisis por conglomerados.

Abstract

Spatial analysis is a set of techniques that use the geographical locations of the events as input data. The use of spatial analysis techniques in epidemiological studies has been growing rapidly in the last decades because they allow geographic or spatial variability to be included in the analyses. This paper makes a general review of the methods of spatial analysis applied to epidemiology and aims to encourage the study of spatial analysis and promote the incorporation of these techniques in current research. We describe the methods of spatial analysis from the descriptive and analytical approach emphasizing the use, advantages and limitations of choropleth maps, hypothesis tests and the identification of clusters of diseases.

Keywords: Spatial analysis, geographic mapping, cluster analysis.

1. Universidad Industrial de Santander. Bucaramanga, Colombia.

Correspondencia: Ana María Valbuena. Dirección: Carrera 32 No. 29-31 Oficina 310. Bucaramanga. Teléfono: +57 3213321773. Correo electrónico: anavalbuenag@gmail.com

Introducción

Se reconoce que el lugar a donde pertenece un individuo o una población son determinantes en el proceso de salud-enfermedad; sin embargo, en la mayoría de las investigaciones en salud pública predomina el análisis de las características de las personas y el tiempo. Durante los últimos 25 años, el uso de los métodos analíticos de tipo espacial ha crecido exponencialmente en epidemiología de la mano con los avances en los Sistemas de Información Geográficos (SIG), proporcionando a los investigadores nuevas herramientas para explorar y analizar los datos¹.

El análisis espacial puede definirse como un conjunto de técnicas cuyos insumos corresponden a la localización espacial o geográfica de los objetos o eventos que se analizan, lo que requiere información tanto de la ubicación como de los atributos de los objetos². Estos análisis permiten que las dependencias espaciales entre los datos puedan ser evaluadas y se comprenda la naturaleza de estas relaciones espaciales³. Los objetivos del análisis epidemiológico espacial son la descripción de patrones espaciales, la identificación de grupos de enfermedades y la explicación o predicción del riesgo de enfermedad².

Aunque el análisis espacial es considerado un método incorporado recientemente a la epidemiología, hay dos referentes históricos en los siglos XVIII y XIX en los que se usaron exitosamente los mapas como una herramienta descriptiva útil durante la investigación de brotes. El primero fue el trabajo del Dr. Seaman en Nueva York durante un brote de fiebre amarilla en 1798, donde determinó que muchas de las muertes causadas por la enfermedad estaban ubicadas cerca a lo que denominaba “miasmas mórbidos”⁴. El segundo referente histórico fue el trabajo del Dr. Snow durante una epidemia de Cólera en Londres en 1854, en el que por medio del uso de mapas concluyó que los casos estaban relacionados con lugares específicos de provisión de agua; este trabajo es el ejemplo más famoso del análisis espacial aplicado a la epidemiología⁵.

Describir adecuadamente el patrón espacial de un evento puede proporcionar elementos para elaborar hipótesis de sus causas. La descripción del lugar donde las personas viven puede ser importante para identificar patrones de una enfermedad; sin embargo,

las exposiciones de los individuos se presentan no sólo en el lugar donde viven, sino también en donde trabajan, juegan o pasan tiempo³. Por esta razón, el análisis espacial es útil no solamente para identificar patrones de enfermedades, sino también para analizar la distribución espacial de estructuras sociales, patrones ocupacionales y exposiciones ambientales de una población⁵.

Debido a que la epidemiología contemporánea está dirigida al estudio de riesgos que son generalmente pequeños, se requiere de herramientas que permitan que la evaluación de la exposición sea muy detallada y de alta calidad; los métodos de análisis espacial aportan a esta necesidad al tener en cuenta las fuentes de variabilidad espacial y temporal de las exposiciones⁶. En la literatura científica, la epidemiología ambiental y del cáncer han sido los campos pioneros en el uso de métodos sofisticados de análisis espacial⁷.

El objetivo de este artículo es hacer una revisión general de los métodos de análisis espacial aplicados a la epidemiología y su propósito es incentivar el estudio del análisis espacial y promover la incorporación de estas técnicas en las investigaciones epidemiológicas en el país.

Datos espaciales

Las investigaciones que emplean métodos de análisis espacial demandan datos que estén espacialmente referenciados. El proceso por el cual se obtienen estos datos se llama geocodificación, y consiste en tomar los datos textuales, como direcciones o códigos postales, y convertirlos en información de localización espacial susceptibles de analizar posteriormente⁸. La información espacial está conformada por datos que se pueden ver o localizar en dos o tres o más dimensiones, por ejemplo, en un mapa⁵.

La cantidad de datos georreferenciados que contienen información de salud de las personas ha aumentado extraordinariamente durante las últimas décadas. Describir la localización de los eventos permite establecer relaciones espaciales entre ellos, como distancias entre dos puntos y la proximidad entre diferentes puntos⁹. No obstante, para realizar análisis espacial se requiere que los datos cuenten con características y atributos espaciales, así como el soporte de los datos⁵.

Los tipos de características espaciales son el punto, la línea, el área y el volumen. Un punto es una localización precisa espacialmente (un punto en un mapa), que puede ser georreferenciado por dos coordenadas, la latitud y la longitud. Una línea son puntos conectados secuencialmente, el área es una región cercada por líneas y el volumen es un objeto tridimensional que tiene una extensión vertical y horizontal. Cada característica tiene cierto tamaño, forma y orientación espacial específica, y en conjunto, estas propiedades forman el soporte de los datos. A su vez, los datos espaciales tienen atributos asociados que son observaciones o medidas (por ejemplo tasas de incidencia de enfermedad por área)⁵.

Mediante el uso de los SIG es posible representar los datos espaciales. De acuerdo con Lance los SIG son “un complejo software interactivo para la gestión, síntesis y visualización de datos espaciales”⁵. En SIG, sistema expresa que un SIG está compuesto de diversos componentes diferentes pero relacionados para trabajar juntos, información implica que los datos introducidos pueden organizarse de manera que faciliten la interpretación, y geográfico hace referencia a que las ubicaciones espaciales pueden especificarse por coordenada, tales como latitud y longitud⁵. Por lo tanto, los SIG son mucho más que un sistema de cartografía automatizado o computarizado⁶.

En este artículo, se propone que desde la perspectiva epidemiológica el uso de los datos espaciales se entienda y ejecute como se realiza el de los otros datos de uso cotidiano: comenzar con métodos descriptivos, desarrollar luego un análisis de carácter bivariado con pruebas de hipótesis y finalmente, de ser necesario incorporar los datos espaciales a análisis de tipo multivariable, incluso de tipo multinivel. A continuación, se revisan los métodos descriptivos y analíticos más importantes de análisis espacial aplicado a la epidemiología.

Métodos descriptivos de datos espaciales: mapas de eventos

Al igual que con el análisis de cualquier conjunto de datos, se recomienda iniciar con la exploración de los datos y los gráficos, y en este caso los mapas, son uno de los mejores aliados para ese propósito¹⁰. Los mapas son una de las herramientas más poderosas para comunicar de manera visual los datos y poder

establecer comparaciones de forma sencilla¹¹. El mapeo de eventos en salud se utiliza con fines descriptivos, para generar hipótesis etiológicas, para vigilancia epidemiológica de áreas de riesgo, ayudar a desarrollar políticas sanitarias y determinar la mejor disposición de recursos financieros y humanos^{12,13}.

Los mapas se pueden clasificar como topográficos, cualitativos y cuantitativos; los mapas cuantitativos son los más usados, ya que proporcionan tanto información cuantitativa del evento en estudio, como su distribución espacial⁵. Existen tres tipos básicos de mapas de enfermedades correspondientes a ciertos tipos de datos, que son: mapa de densidad de puntos o caso-evento, mapas coropléticos para datos regionales e isopléticos para datos continuos. En el análisis espacial se usan principalmente los mapas de puntos con datos de caso-evento y los mapas coropléticos que resumen datos de áreas¹⁰.

Los mapas de densidad de puntos son de tipo cualitativo, son los más sencillos y son capaces de mostrar eventos de salud con la resolución de un par de coordenadas¹³. La construcción de un mapa de puntos, donde la ubicación representa casos o agregaciones de casos de enfermedad, se realiza asignando símbolos en un mapa. Los símbolos pueden representar un cierto número de casos de enfermedad. Si bien estos mapas son útiles para proporcionar una visión general de los patrones de una enfermedad, son poco confiables porque no toman en cuenta la densidad o estructura de la población y por lo tanto no son adecuados para medir el impacto de las enfermedades o hacer comparaciones entre distintas poblaciones⁸.

Los mapas coropléticos son los más usados para visualizar datos en un área que está coloreada o sombreada con una intensidad proporcional a un valor asociado⁵. Esto es particularmente útil ya que, en diferentes escenarios de salud pública es necesario poder comparar tasas de morbilidad o mortalidad de una enfermedad entre dos o más poblaciones; sin embargo, puede haber diferencias en las distribuciones de las poblaciones que distorsionan la comparación. El uso de mapas coropléticos con tasas estandarizadas (directas o indirectas) es una herramienta que proporciona una manera útil de comparar los resultados de salud entre poblaciones o regiones que pueden tener diferente densidad poblacional o distribuciones de edad¹ (**Figura 1**).

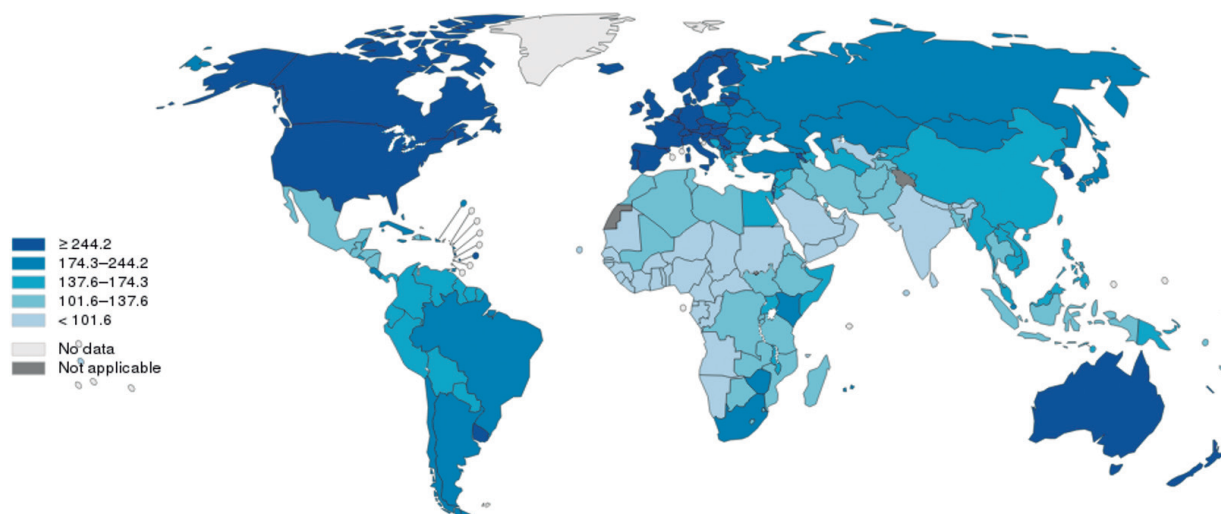


Figura 1. Mapa coroplético mundial de las Tasas estimadas estandarizadas por edad de incidencia de todos los tipos de cáncer excluyendo el cáncer de piel no melanoma, en ambos sexos, 2012.

Fuente: Globocan 2012-IARC.

En muchas ocasiones se necesita estimar de manera óptima las exposiciones en lugares donde no se cuenta con suficientes datos, a partir de valores conocidos de la exposición de puntos geográficos cercanos; esto se logra mediante el uso de modelos de exposición. Existen diferentes modelos para estimar las exposiciones espacialmente que se han desarrollado de manera escalonada en el tiempo y que se pueden clasificar en tres tipos de modelos con complejidad ascendente en términos de la cantidad de información y el procesamiento estadístico que necesita para sus estimaciones¹⁴:

- 1) Modelos basados en distancia: constituyen la estimación más sencilla y solo necesitan información de coordenadas y distancias entre un punto y una fuente de exposición.
- 2) Modelos de interpolación espacial⁵: En este grupo se encuentran los métodos determinísticos (basados en modelos matemáticos) y probabilísticos (basados en modelos estadísticos que incluyen estimados de precisión). El método determinista más usado es la interpolación de la distancia inversa, que es un promedio ponderado de valores vecinos. El peso dado a cada observación es una función de la distancia entre la ubicación de esa observación y el punto en el que se desea la interpolación^{5, 15}. Aunque es un método rápido de calcular, tiende a producir patrones poco reales conocidos como tipo “ojo de buey” alrededor de los puntos muestreados⁵. El método probabilístico mas usado es el de Kriging,

el cual se basa en la autocorrelación espacial de las variables, donde las ponderaciones están basadas no solo en la distancia entre los puntos medidos y la ubicación de la predicción, sino también en la disposición espacial de los puntos medidos¹⁵. Tiene como ventaja sobre los métodos deterministas que además de generar una superficie de predicción, también proporcionan alguna medida de precisión de las predicciones, puesto que la producción de errores estándar (variación Kriging) en ubicaciones no muestreadas cuantifican el grado de incertidumbre en las predicciones espaciales en sitios no muestreados, ilustrando dónde la interpolación es menos confiable¹⁴.

- 3) Modelos de uso de suelo: son modelos estadísticos que permiten estimar la exposición en diferentes puntos geográficos de una superficie basados en la integración de diferentes variables topográficas y de uso de suelo relacionadas con la exposición por medio de un modelo lineal multivariable de predicción. Las variables que se incluyen en el modulo frecuentemente se relacionan con la distancia a diferentes fuentes de emisión de exposiciones. En modelos de predicción de concentraciones de contaminantes del aire, por ejemplo, se utilizan como medidas básicas la medición de los contaminantes en diferentes puntos de la ciudad y se combinan con otras variables como la distancia a vías de alto y mediano tráfico y la distancia al centro de la ciudad, entre otras.

Para la estimación de exposiciones atmosféricas se han desarrollado además modelos de dispersión lineal y modelos integrados de emisiones con variables meteorológicas¹⁴ en los que además del uso de datos de emisiones o concentraciones de las exposiciones en puntos específicos, se adiciona a los modelos información de condiciones meteorológicas y topográfica relacionadas con la exposición. Estos modelos se basan en modelos deterministas con distribución Gaussiana. También se describe el uso de modelos híbridos, que consisten en la combinación de dos o más de los modelos anteriores para la estimación de exposiciones.

Finalmente, es importante recordar que representar en mapas de una exposición o un evento es considerado un análisis exploratorio, que es utilizado para obtener una visión de la distribución geográfica o espacial de estas condiciones^{6,10}. Por lo anterior, se requiere cautela en la interpretación porque los aparentes patrones observados pueden ser creados o eliminados artificialmente dependiendo de cómo se represente o interpole la variable asignada y la escala o resolución geográfica. Adicionalmente, la variación de las tasas en el mapa puede reflejar diferencias en la calidad de los datos, en el diagnóstico, la clasificación o notificación de la enfermedad, en lugar de diferencias reales en las tasas de enfermedad¹⁵.

Métodos analíticos de datos espaciales: identificación de clústers

Un clúster espacial o conglomerado espacial, hace referencia a un aumento o exceso de casos en ubicaciones específicas o un patrón inusual en un área de estudio⁶. Por ejemplo, un clúster de cáncer se define como un número mayor al esperado de casos de cáncer que ocurre en un grupo de personas dentro de un área geográfica o durante un período de tiempo específico. Un número mayor al esperado quiere decir que el número observado de casos es mayor al que normalmente se observaría en una situación similar, es decir en un grupo con población, edad, raza o sexo similares¹⁶.

La identificación espacial de la ocurrencia de clúster permite también ayudar a identificar posibles causas relacionadas, por ejemplo, exposiciones ambientales. Dentro de las razones para realizar este tipo de investigaciones esta establecer si observaciones aisladas de aumento de casos en un área específica son realmente anormales y evaluar si hay posibles causas que puedan ser intervenidas para prevenir exposiciones

en la población¹⁷. El estudio de clústers ha permitido el descubrimiento de enfermedades, como la enfermedad de Lyme. Pese a ello, existe un debate alrededor de la efectividad de las estrategias de investigación para identificar aquellos clúster que se deben a una variación normal de la incidencia de una enfermedad, de los que se les puede atribuir una causa¹⁷.

Las pruebas de hipótesis de clúster espacial se utilizan como pruebas de identificación de clústers y tienen cinco componentes¹⁸: 1) *La hipótesis nula*, que describe el patrón esperado cuando la hipótesis alternativa es falsa, por ejemplo un riesgo uniforme en un área.; 2) *La hipótesis alternativa*, que describe el patrón espacial que la prueba está diseñada para detectar; 3) *El modelo espacial nulo*, que es un mecanismo para generar la distribución de referencia; 4) *La distribución de referencia*, que es la distribución de la prueba estadística cuando la hipótesis nula es verdadera; y 5) *La prueba estadística*, que son pruebas de hipótesis que se detallan a continuación.

Pruebas de hipótesis espaciales

Para la detección de clúster espacial de una enfermedad se pueden utilizar pruebas focalizadas y no focalizadas, que a su vez se clasifican según la pregunta estadística que se desea resolver con respecto al clúster⁶. (**Figura 2**).

Las pruebas globales mediante una sola prueba establecen patrones generales y autocorrelación espacial; su resultado permite rechazar o aceptar la hipótesis nula, la cual plantea que los casos observados se distribuyen al azar, según lo esperado para el área geográfica. Las estadísticas globales no identifican dónde están los clústeres, ni cuantifican cómo varía la dependencia espacial de un lugar a otro. Las pruebas globales pueden usar los estadísticos de Moran I o Tango, entre otros⁶.

Las pruebas focales pueden ser localizadas o no localizadas. Las pruebas no localizadas evalúan la presencia de clústers de manera general en un área geográfica y las localizadas evalúan la presencia de clústers alrededor de una ubicación específica (por ejemplo una fuente de emisión de contaminación)⁵.

Dentro de las pruebas focales no localizadas, están las desarrolladas por Turnbull, Openshaw, Besag and Newell, y la de Kulldorff, siendo ésta última la más ampliamente utilizada⁶. La prueba de Kulldorff es una prueba de escaneo circular, la cual crea ventanas circulares sucesivas en cada región y el radio del círculo varía de cero a una distancia máxima preespecificada o

hasta un número máximo preespecificado de regiones a incluir en el clúster. Para cada círculo se calcula una estadística de razón de probabilidad con base al número de casos observados y esperados dentro y fuera del círculo. La hipótesis nula de esta prueba es que el riesgo de la enfermedad es igual dentro y fuera del

círculo escaneado¹⁹. Una de las aplicaciones que se han realizado de este método se realizó en el Nordeste de los Estados Unidos, donde se detectó un clúster espacial principal y cuatro clústeres secundarios de mortalidad por cáncer de seno en mujeres²⁰.

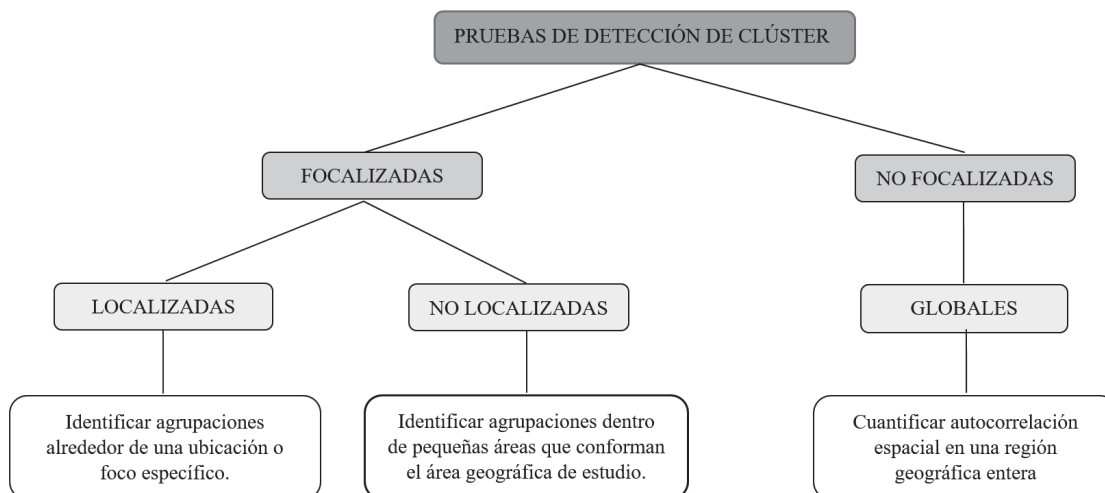


Figura 2. Clasificación de las pruebas de detección de clúster espaciales.
Fuente: Autores.

La prueba de Kulldorff también puede convertirse en una prueba focal localizada al incorporar un punto de referencia, por ejemplo, la localización de una fuente de exposición en estudios ambientales⁶. Esta prueba considera ventanas de tamaño variable con un centro en un foco que puede ser una fuente de exposición. Cada ventana define un grupo esperado de casos. En este caso la hipótesis nula es que el riesgo de la enfermedad o evento es igual dentro y fuera de la ventana que tiene como centro un foco preestablecido¹⁹. Kulldorff desarrolló un software libre llamado SatScanTM, diseñado para analizar datos espaciales, temporales y espacio-temporales utilizando las estadísticas de escaneo espacial, temporal o espacial-temporal²¹.

Otra prueba focal localizada es la prueba de Stone, en la que las subregiones se ordenan en términos de distancia del punto de referencia, lo que permite evaluar si el riesgo de una enfermedad se altera con la distancia a la fuente putativa. La hipótesis nula de esta prueba es que los riesgos relativos son constantes en todas las áreas geográficas y la hipótesis alterna es que hay una tendencia decreciente en el riesgo de la enfermedad a medida que la distancia a la fuente de exposición aumenta²². Esta prueba puede implementarse usando el paquete de programación “DCluster” codificado en el software R²³.

La prueba de Lawson es también una prueba focal localizada que considera los efectos direccionales que se espera que tengan las emisiones de una fuente putativa. Esta prueba usa los casos observados y esperados en cada unidad geográfica y el ángulo entre la fuente de contaminación y el centroide (centro del polígono geográfico) de las unidades geográficas. La hipótesis nula de esta prueba es que el riesgo de la enfermedad es igual en todas las direcciones desde la fuente putativa²⁴.

Alcances y limitaciones de los análisis de clústeres en salud

Algunas pruebas de hipótesis para detección de clúster se han hecho computacionalmente muy sencillas por lo cual son usadas frecuentemente. No obstante, el centro de control y prevención de enfermedades CDC por sus siglas en inglés de Estados Unidos, ha reportado que la probabilidad de establecer una relación de causa y efecto definitiva entre los clústeres de enfermedad y una exposición no es muy alta; por tal motivo, estos estudios se utilizan principalmente para dirigir algunas actividades de control de enfermedades, generar hipótesis de investigación y orientar el inicio estudios epidemiológicos para identificar posibles factores asociados o causales²⁵.

También es importante tener en cuenta que los patrones observados reflejan la influencia de una compleja red de factores demográficos, sociales, económicos, culturales y ambientales que probablemente cambian de forma continua a través del tiempo y el espacio²⁶. De igual manera, la elección de un área geográfica demasiado pequeña o demasiado grande, o de un período de tiempo demasiado corto o demasiado largo, puede dar lugar a insuficiente potencia estadística para identificar un clúster²⁵.

Otros métodos analíticos de datos espaciales

Tradicionalmente, el análisis de regresión multivariable en epidemiología se utiliza para evaluar los efectos de una variable de exposición sobre una variable de resultados de salud controlando por el efecto de otras variables que son potencialmente confusoras. Los análisis de regresión multivariable pueden modelarse de muchas formas, siendo los más familiares los modelos lineales generalizados (GLM, por sus siglas en inglés), tales como la regresión lineal, Poisson y logística²⁷. El mismo raciocinio analítico multivariable se pueden utilizar en el análisis espacial incorporando datos sobre las relaciones espaciales y los términos de error espacialmente correlacionados⁵. El uso de modelos de regresión espacial ayudan a refinar más las conclusiones sobre análisis de clúster espaciales o espacio-temporales de eventos por cuanto pueden incluir el ajuste por otras variables que pueden estar relacionadas con las exposiciones y resultados de interés en el análisis.

Predecir los resultados de salud, mientras se controla por otros factores de riesgo conocidos conduce a evidencia sugestiva de asociaciones estadísticas (y potencialmente causales). Cuando las investigaciones epidemiológicas usan datos de salud de áreas geográficas contiguas o cercanas, los datos pueden no proporcionar estimaciones independientes de la variable dependiente (por ejemplo, riesgo relativo en morbilidad o mortalidad). En un modelo general sin información espacial el análisis de la asociación entre una exposición y resultado de salud estaría sesgado por esta falta de independencia de las covariables que se autocorrelacionan espacialmente y por tanto violan los supuestos de los GLM. En los modelos multivariables que incluyen datos espaciales la correlación espacial de las variables se modela dentro del componente sistemático del modelo y por tanto el sesgo y la subestimación de la variabilidad estadística se reducen y producen estimados válidos porque los términos de error de estos modelos tienden a no estar correlacionados²⁸.

Para la predicción espacial, lo usual es realizar el proceso de predicción de una variable a la vez; pero con la estimación de modelos de regresión multivariable se logra la evaluación de efectos combinados de variables espaciales (como distancia y dirección) que son de particular importancia para el análisis de clústers alrededor de fuentes putativas de exposición⁶. Los GLM con distribución Poisson son modelos de regresión multivariable que han sido utilizados en estudios de análisis de distancia (o análisis de proximidad) en estudios de efectos de exposiciones ambientales y que en su análisis incorporan el análisis de correlación espacial de los datos⁶. Recientemente, se publicaron los resultados de una investigación en Epidemiología geográfica de Cáncer Hepatocelular en la ciudad de Nueva York, donde aplico un modelo de regresión Poisson incluyendo variables geográficas y hallaron que el Cáncer Hepatocelular y la hepatitis viral están relacionados con medidas de posición socioeconómica²⁹. Alternativamente, los métodos de análisis multivariable espacial pueden usar como variable independiente la inclusión o no de un caso o área específica dentro de un radio de distancia de interés para un punto de exposición (ejm. Casos ubicados dentro o fuera de un radio de 5 Km alrededor de una fuente de radiación) y de esta manera se pueden también modelar efectos de exposiciones con una dimensión espacial usando variables dicotomizadas u ordinales.

Conclusiones

La epidemiológica se basa en el análisis de la triada epidemiológica de tiempo, lugar y persona. El análisis espacial es una poderosa herramienta para el análisis de la variable de lugar que brinda a las investigaciones en salud una perspectiva más amplia de la ocurrencia de eventos de salud y enfermedad. El conocimiento y entrenamiento en el uso de métodos de análisis espacial es necesario para alcanzar las habilidades mínimas que se requieren para su ejecución y también, para comprender adecuadamente los resultados que genera y sus limitaciones.

Referencias

1. Rezaeian M, Dunn G, St Leger S, Appleby L. Geographical epidemiology, spatial analysis and geographical information systems: a multidisciplinary glossary. *J Epidemiol Community Heal.* 2007;61(2): 98-102. doi: 10.1136/jech.2005.043117.
2. Goodchild M, Haining R, Wise S, Arbia G, Anselin L, Bossard E, et al. Integrating GIS and spatial data

- analysis: problems and possibilities. *Int J Geogr Inf Syst.* 1992; 6(5): 407-423.
3. Duncan C. T. *Statistical methods in environmental epidemiology.* 2nd edition. New York. Oxford University Press. 2009.
 4. Koch T. *Disease maps: epidemics on the ground.* 1st ed. Chicago, IL. University Of Chicago Press; 2011.
 5. Waller LA, Gotway CA. *Applied spatial statistics for public health data.* 1 st ed. New York. Wiley-Interscience; 2004.
 6. Lawson AB, Banerjee S, Haining R, Ugarte L. *Handbook of spatial epidemiology.* New York. CRC Press; 2016.
 7. Auchincloss AH, Gebreab SY, Mair C, Diez Roux AV. A review of spatial methods in epidemiology, 2000-2010. *Annu Rev Public Health.* 2012; 33(1): 107-122.
 8. Boscoe FP. *Geographic Health Data: Fundamental Techniques for Analysis.* Wallingford. CABI. International. 2013.
 9. Rytkönen MJ. Not all maps are equal: GIS and spatial analysis in epidemiology. 63, *Int J Circumpolar Health.* 2004; 63(1): 9-24.
 10. Berke O. Exploratory disease mapping: kriging the spatial risk function from regional count data. *Int J Health Geogr.* 2004; 3(1): 18.
 11. Elliott P, Wartenberg D. Spatial epidemiology: current approaches and future challenges. *Environ Health Perspect.* 2004; 112(9): 998-1006. doi: 10.1289/ehp.6735.
 12. Lawson AB, Biggeri AB, Boehning D, Lesaffre E, Viel JF, Clark A, et al. Disease mapping models: an empirical evaluation. *Disease Mapping Collaborative Group. Stat Med.* 2000; 19(17-18): 2217-2241.
 13. Lawson AB, Biggeri A, Bohning D, Lesaffre E, Jean-Franpis Viel, Roberto Bertollini. *Disease Mapping and Risk Assessment for Public Health.* 1st ed. Public Health. Wiley; 1999. 502 p.
 14. Jerrett M, Arain A, Kanaroglou P, Beckerman B, Potoglou D, Sahsuvaroglu T, et al. A review and evaluation of intraurban air pollution exposure models. *J Expo Anal Environ Epidemiol.* 2005; 15(2): 185-204.
 15. Lai PC, So FM, Chan KW. *Spatial Epidemiological Approaches in Disease Mapping and Analysis.* Public Health. 1st ed. New York. CRC Press. 2009.
 16. National Center for Environmental Health. CDC - Cancer Clusters - Home.
 17. Baker DB, Nieuwenhuijsen MJ. *Environmental epidemiology: study methods and application.* 1st ed. Oxford University Press; 2008.
 18. Waller LA, Jacquez GM. Disease models implicit in statistical tests of disease clustering. *Epidemiology.* 1995; 6(6): 584-590.
 19. Kulldorff M. A spatial scan statistic. *Commun Stat - Theory Methods.* 1997; 26(6): 1481-1496.
 20. Kulldorff M, Feuer EJ, Miller BA, Freedman LS. Breast cancer clusters in the northeast United States: a geographic analysis. *Am J Epidemiol.* 1997; 146(2): 161-170.
 21. Kulldorff M. SaTScan™ v8.0: software for the spatial and space-time scan statistics. Satscan.org. 2009.
 22. Stone RA. Investigations of excess environmental risks around putative sources: Statistical problems and a proposed test. *Stat Med.* 1988; 7(6): 649-660.
 23. R Core Team. R Software. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2013.
 24. Lawson A. *Statistical Methods in Spatial Epidemiology.* Wiley Series in Probability and Statistics. 2nd edition. John Wiley & Sons. 2006. 424 p.
 25. Center for disease control and prevention. Guidelines for investigating clusters of health events. Vol. 30, *Mortality and Morbidity Weekly Report.* 1990; 39(RR-11):1-16.
 26. Jacquez GM. Cluster morphology analysis. *Spat Spatiotemporal Epidemiol.* 2009; 1(1): 19-29.
 27. Kleinbaum DG, Kupper LL, Muller KE, Nizam A. *Applied regression analysis and other multivariable methods.* 4th ed. CA. Wadsworth Publishing Co Inc. 2007.
 28. Jerrett M, Burnett RT, Goldberg MS, Sears M, Krewski D, Catalan R, et al. Spatial analysis for environmental health research: concepts, methods, and examples. *J Toxicol Environ Health.* 2003; 66: 1783-1810. doi: 10.1080/15287390306446.
 29. Ford MM, Ivanina E, Desai P, Highfield L, Qiao B, Schymura MJ, et al. Geographic epidemiology of hepatocellular carcinoma, viral hepatitis, and socioeconomic position in New York City. *Cancer Causes Control.* 2017; 28(7): 779-789. doi: 10.1007/s10552-017-0897-8.

“Todo está relacionado con todo lo demás, pero las cosas cercanas están más relacionadas que las distantes” Waldo Tobler