

Caracterización de los datos geospaciales para su integración en estudios de salud y epidemiología.

Dietrich Patricia¹, De Pietri Diana^{1,2}, Carcagno Alejandro¹, Ernesto de Titto² & Igarzabal María Adela¹

¹ Centro de Información Metropolitana, Facultad de Arquitectura, Diseño y Urbanismo, Universidad de Buenos Aires.

² Dirección Nacional de Determinantes de la Salud. Ministerio de Salud de la Nación.

pdietr@fadu.uba.ar, depietrid@hotmail.com, acarcagno@yahoo.com.ar,
edetitto@msal.gov.ar

Resumen. Se estudian diferentes indicadores geográficos para el monitoreo de un desarrollo sustentable, se genera un modelo espacial para el análisis de la exposición a ambientes degradados y se relaciona el diagnóstico territorial con un análisis de la mortalidad de la población.

Se describen los aspectos metodológicos y los conceptos asociados a los datos provenientes del campo de la salud tendientes a ser incorporados en la infraestructura nacional. Se detallan los procedimientos operativos en 4 etapas: 1-recopilación de información secundaria de fuente oficial; 2-normalización para vinculación de los datos recopilados; 3-asignación de población de referencia y cálculo de las tasas mortalidad; 4-generación del mapa de riesgo y validación con la mortalidad acumulada.

La calidad del dato espacial es fundamental para realizar el análisis de diagnósticos territoriales tendientes a contribuir en el proceso de toma de decisiones en políticas públicas. Un ejemplo de ello es la dificultad que surge de la falta de estandarización de criterios relacionados con la especificación de localidad. Esto conlleva a la definición arbitraria de tasas de mortalidad, a nivel de localidad, ya que se definen diferentes escenarios a partir de los siguientes factores: periodos temporales, disposición espacial de radios censales (aislados-contiguos), zonas urbanas y rurales.

PALABRAS CLAVES

Mortalidad. Localidad. Modelo. Riesgo. Calidad.

INTRODUCCIÓN

Uno de los ejes temáticos donde existe un importante auge en el uso, análisis y gestión de datos espaciales está referido al ámbito de las políticas públicas. Lo relevante para el decisor público es la construcción de juicios basados en instrumentos de información y valoración que ofrezcan garantías de credibilidad. Por ello, todo proceso va encaminado a satisfacer las necesidades informativas garantizando la fiabilidad de los datos, la solidez de los análisis, la credibilidad de los hallazgos, la validez de las conclusiones y la utilidad de sus recomendaciones (1).

La capacidad de relacionar variables que poseen valor alfanumérico como expresión territorial, que hace posible realizar investigaciones que especialicen el alcance e impacto de las políticas públicas, permite poner en valor el análisis espacial en las ciencias de la salud.

Los estudios epidemiológicos tienen el objetivo de determinar las relaciones persona-espacio-tiempo. Sin embargo, no siempre se le da la importancia debida al espacio como un componente primordial en los estudios sobre la salud, (2).

Determinar los riesgos existentes y presentar cartográficamente las áreas expuestas a estos peligros es parte importante de la vigilancia epidemiológica. Los sistemas de información geográfica (SIG) en salud pública son utilizados en el análisis de la situación de salud, la vigilancia de eventos, el estudio epidemiológico, la planeación y la evaluación de estrategias por zonas de salud, así como en la gestión y toma de decisiones (3).

En la actualidad existe una gran cantidad de información de salud georreferenciada, que suelen ser tratada con limitadas herramientas de análisis. Sin embargo estas herramientas son las líneas fundamentales de avance técnico metodológico de la disciplina en las últimas décadas y el desarrollo de sistemas de ayuda a las decisiones espaciales en el espectro público e intersectorial.

Este artículo forma parte de un trabajo en el cual se estudian diferentes indicadores geográficos para el monitoreo de un desarrollo sustentable, se genera un modelo espacial para el análisis exposición a ambientes degradados y se valida el diagnóstico territorial mediante un análisis de la mortalidad de la población (4).

En el presente trabajo se describen los aspectos metodológicos y los conceptos asociados al proceso de recolección y revisión de los datos oficiales existentes

tendientes a ser incorporados en la infraestructura nacional de datos. Se describen los principales conjuntos de datos seleccionados, la generación de un campo ID común, la asignación de datos de defunción y las herramientas utilizadas para su integración espacial. La finalidad es generar procesos metodológicos para la obtención de evidencia geográfica para la priorización de decisiones en salud pública.

ÁREA DE ESTUDIO

Se tomó como estudio de caso a la provincia de Neuquén (figura 1) dado el libre acceso a la información espacial a través del IDE provincial (5). En este sitio web están disponibles los datos de las actividades antrópicas que se desarrollan en el territorio provincial.

PROCEDIMIENTOS OPERATIVOS

1.- Recopilación de información secundaria de fuente oficial.

La información secundaria de organismos oficiales está constituida por datos no espaciales a través de tablas (herramienta de organización de información que se utiliza en bases de datos específicas) y datos espaciales (elemento ubicado en el espacio mediante un sistema predefinido de coordenadas y el cual puede ser descrito por sus atributos y su relación con otros elementos en el mismo plano); que vinculan distintas entidades (polígono, línea, puntos) en el espacio geográfico. Para ello es necesario contar con un registro sistemático de cada una de estas variables a fin de definir los aspectos formales, funcionales y estructurales de la provincia.

1.1.-Tablas de datos

Mortalidad

La mortalidad es la cantidad de muertes ocurridas por todas las causas, en todos los grupos de edad y para ambos sexos. La mortalidad expresa la dinámica de las muertes acaecidas en las poblaciones a través del tiempo y el espacio.

El número de defunciones de la provincia a nivel de localidad correspondiente al periodo 2000 al 2012 fue provisto por la Dirección de Estadística e Información en Salud, Ministerio de salud de la Nación.

La tabla de defunciones de la provincia presenta por año las siguientes columnas con códigos numéricos: provincia, departamento, localidad, sexo, grupo de edad, causa de defunción y código CIE10.

Localidad

El listado de lugares utilizados por la Secretaria de Salud de la provincia de Neuquén

1.2.- Conjuntos de datos espaciales.

Población

Fuente: Instituto Nacional de Estadística y Censos. Censo Nacional de Población, Hogares y Viviendas 2001 y 2010 Serie 2. Argentina. Acceso 24/07/2011. Unidad de análisis censal (número de habitantes totales).

Radios censales 2001: Los datos se encuentra bajo el sistema de coordenadas Planas, Proyección origen Gauss Krüger Faja 2.

Radios censales 2010: Los datos se encuentra bajo el sistema de coordenadas Planas, Proyección origen Gauss Krüger Faja 2 y sistema de coordenadas geográficas, proyección EPSG-WGS 84

La representación geométrica es mediante polígonos, caracterizados a través de sus coordenadas geográficas, un identificador único, área, perímetro y campos temáticos específicos (denominación y código de provincia, departamento y localidad; número de fracción y radio; y datos poblacionales.

Localidad

Fuente: Base de asentamientos humanos de la República Argentina (BAHRA), trabajo conjunto entre el Ministerio de Educación, a través del Programa Nacional Mapa Educativo, el Instituto Nacional de Estadística y Censos (INDEC) y el Instituto Geográfico Nacional (IGN).Proyección: EPSG: 4326 - WGS 84. Los datos se encuentran bajo el sistema de coordenadas geográficas, proyección EPSG- WGS 84

Datos espaciales representados geoméricamente a través de puntos y caracterizados mediante sus coordenadas geográficas, un identificador único y campos temáticos específicos (Localidad: código y denominación de provincia, código y denominación de departamento, código BAHARA, y nombre de la entidad)

Ambiente. Usos del suelo.

Fuente: COPADE Unidad de Sistemas de Información Territorial del Ministerio de Desarrollo Territorial. Provincia de Neuquén. Dicha información fue adquirida en septiembre del 2014.

Proyección: Los datos se encuentran bajo el sistema de coordenadas planas, proyección origen Gauss Krüger Faja 2

Datos espaciales representados geoméricamente a través de puntos y caracterizados mediante sus coordenadas geográficas, un identificador único y campos temáticos específicos:

- Centrales/usinas hidroeléctricas, aserraderos, industrias, tanques de combustibles planta de tratamiento, planta de bombeo, tanques de almacenamiento, estaciones geotérmicas; nombre de cada entidad, tipo, entre otros.
- Pozos de extracción de gas y petróleo del subsuelo, nombre de cada pozo, el tipo, la provincia y departamento entre otros.
- Estaciones de servicio, nombre de la empresa, razón social, CUIT, domicilio, bandera, tipo de boca, localidad, partido entre otros.

Datos espaciales representados geoméricamente a través de líneas y caracterizados mediante sus coordenadas geográficas, un identificador único, longitud de línea, nodo izquierdo, nodo derecho y campos temáticos específicos:

- Gasoductos, oleoductos, acueductos; nombre de cada entidad, jurisdicción, entre otros
- Sistema vial; nombre Jurisdicción, el tipo, clase, tránsito, entre otros.

Datos espaciales representados geoméricamente a través de polígonos y caracterizados mediante sus coordenadas geográficas, un identificador único, área, perímetro y campos temáticos específicos:

- Canteras, minas; nombre de cada entidad, expediente, titular, sustancia explotada y categoría entre otros.
- Agroindustria y producción primaria; nombre y descripción.
- Zonas de bosque nativo; tipo, entre otros.
- Parques nacionales y provinciales; la clase, entre otros.

2.- Normalización para vinculación de los datos recopilados

Para poder vincular los datos espaciales con los de tablas, a estas últimas se les crea un identificador común (ID), concatenando los códigos existentes de provincia, departamento y localidad.

Se establece relaciones entre tablas. Se une mediante el ID cada una de las tablas de defunción del periodo 2000 – 2012 con los datos de localidad usados por salud de la provincia;

Los datos 2000-2010 se vinculan con los datos provenientes de la provincia para asignarle la localización espacial y así obtener los mapas de defunción anual por localidad. De igual modo se procedió con las defunciones 2011-2012 con las localidades de BAHRA. Este procedimiento permitió visualizar mapas de puntos de defunción.

3.- Asignación de población de referencia y cálculo de las tasas mortalidad

Las defunciones anuales del total de la provincia de Neuquén fueron georreferenciadas en el mayor detalle espacial referido. Es decir que la defunción fue referenciada como punto en un mapa según fuera una localidad, comuna, barrio, estancia y/o paraje. Para poder realizar comparaciones se asignaron a los datos de defunción la población de referencia.

Los datos de población provienen de los censos nacionales del 2001 y 2010, razón por la cual al conjunto de datos defunción se lo subdividió en dos periodos: 2000-2006 y 2007-2012. En este caso la unidad de referencia espacial fue el radio censal cuya delimitación en el territorio difiere entre los censos nacionales por el simple crecimiento poblacional.

Para el cálculo de las tasas de mortalidad se superpusieron las defunciones (localidades con defunción) sobre la población (polígonos- número de habitantes totales por radio censal). De esta integración surgieron diferentes escenarios, por

ejemplo a.- una localidad por radio censal; b.- varias localidades por radio censal, y c.- una localidad con varios radios censales. A lo anterior se suma la diversidad de tamaños de los polígonos que representan las zonas urbanas (alto número de habitantes en superficies reducidas, polígonos pequeños) frente a las zonas rurales (bajo número de habitantes en superficies extensas, polígonos grandes).

En razón de la complejidad existente en el tipo y límite de la unidad espacial de ambas fuentes de información (puntos-localidad de las defunciones; polígonos-radios censales de la población de referencia) se establecieron diferentes criterios siguiendo las pautas descritas en la bibliografía (6-8): a.-se trabajó con tasas acumuladas de mortalidad por periodo, dado los bajos valores de defunción a nivel de localidad. b.-se utiliza la densidad de población a nivel de radio censal para establecer una unidad de referencia espacial

Se diferenciaron los sitios según la densidad poblacional del radio censal. Este procedimiento se realizó para cada censo nacional. Los radios censales que presentaron 10 o más habitantes por km^2 conformaron la zona urbana diferenciándose de aquellos con menor densidad poblacional (rural). En estas zonas urbanas, se delimitaron sub-zonas cuando los radios censales con 10 o más hab/ km^2 eran contiguos, obteniendo un único valor de población total. Detalle en tabla 1.

El cálculo de la tasa de mortalidad se realizó con la información georreferenciada del número de defunciones en función de la población total. Las tasas acumuladas (de 6 años) se expresaron por 1000 habitantes en riesgo y las tasas específicas (por causa de defunción) por 10000 habitantes en riesgo. Se asume que en el periodo de tiempo la tasa se ha mantenido constante.

4.- Generación del mapa de riesgo y validación con la mortalidad acumulada

Se aplicaron técnicas de evaluación multicriterio (EMC) en el entorno SIG para modelar las variables ambientales ya descritas bajo una perspectiva sanitaria con la finalidad de estimar el riesgo de exposición de la población por habitar en sitios degradados con potencial convergencia sobre la salud humana. Se describen ejemplos en la bibliografía (9).

La EMC comprende una estandarización y transformación de cada uso de la tierra a una escala de impacto ambiental siguiendo los procedimientos de la lógica difusa, resultando en un único valor que sintetiza condiciones ambientales del área de estudio. El mapa resultante muestra un gradiente de alteración ambiental utilizado como base para la estratificación epidemiológica de riesgo para la salud.

El modelo espacial es validado con la tasas de mortalidad general y específicas. El análisis estadístico se lleva a cabo con un nivel de significación de $\alpha = 0,05$; se empleó el programa EPIDAT 3.0 (10).

La validación del modelo de riesgo se pudo llevar a cabo por la unificación del sistema de coordenadas, se decidió transformar todos los datos del sistema de coordenadas Planas, Proyección origen Gauss Krüger Faja 2 al sistema de coordenadas geográficas, proyección EPSG- WGS 84.

CONSIDERACIONES FINALES

Es fundamental considerar la calidad del dato espacial para la construcción de un diagnóstico territorial que contribuya en el proceso de toma de decisión. En este marco las interpretaciones realizadas sobre el modelo espacial construido dependen de la calidad del dato de defunción, la tasa de mortalidad y el riesgo ambiental.

Calidad de los datos de defunción y tasa de mortalidad.

La asignación de la referencia espacial a nivel de localidad de los datos de defunción cambia a partir del 2011, lo que constituyó un obstáculo para un análisis temporal dado la falta de protocolo para su integración.

El código de localidad asignado a los datos históricos de defunción (previo al 2011) fue sólo numérico (sin denominación), lo que dificultó su vinculación y reconocimientos con otras tablas. La definición de localidad, no tuvo un criterio estandarizado, presentó la misma jerarquía un barrio en una ciudad que un paraje o estancia en la zona rural. Además a un porcentaje de los datos de defunción no se le asignó la localidad.

A la incertidumbre de la calidad del dato de defunción, generada por la falta de protocolo a nivel de localidad y/o su falta de referencia, se adicionó un contexto variable en el ámbito geográfico espacial y temporal ya que para el cálculo de la mortalidad se requirió la asignación de población de referencia.

Pese a que los datos de población son de la misma fuente (INDEC), cambió la proyección geográfica, la gráfica y la estructura de la bases de datos entre censos. Esta disparidad en la representación de los datos, generó una incertidumbre a escala desagregada del real crecimiento poblacional.

En relación a la temporalidad de los datos en cuanto a su actualización, se generó un desajuste para la integración de los datos (defunción con periodos de 1 año vs población de referencia con periodos de 10 años).

Sumado a lo anterior se propone incorporar la densidad poblacional promedio como factor para delimitar al territorio de una localidad de su entorno y para que las tasas estén ajustadas a la población de referencia. De tal forma de evitar la falta de criterio estandarizado en relación a los periodos temporales, a la disposición espacial de los radios censales (aislada o contigua), y a las zonas urbanas y rurales.

Calidad del modelo espacial de riesgo a la exposición ambiental

El valor predictivo del modelo espacial debe ser constatado a fin de disminuir el nivel de incertidumbre. Como ya se mencionó la mortalidad expresa la dinámica de las defunciones acaecidas en las poblaciones a través del tiempo y el espacio. El análisis de la variabilidad de las tasas de mortalidad por un periodo de 12 años, brinda información fundamental para realizar la validación del modelo espacial a escala local. Para ello, el diagnóstico ambiental que surge de una evaluación de indicadores multidimensionales es confrontado con los lugares y las causas dominantes de muerte de la población.

Habitualmente, las tasas de enfermedad y/o mortalidad de un área son obtenidas por la relación entre la agregación de casos y/o defunciones y la población de referencia de dicha área. En este trabajo la unidad de referencia espacial corresponde a una unidad ad-hoc a los objetivos de investigación desde la perspectiva de áreas con riesgo ambiental y no a unidades político-administrativa (departamentos / municipios). Al utilizar unidades espaciales no tradicionales se debe hacer un pre-procesamiento de los datos (depuración y normalización) para ajustar la población de referencia a las unidades ambientales. De esta manera las inferencias que resulten del análisis de mortalidad, indican un comportamiento diferencial acerca de las variaciones en las condiciones ambientales.

Por ello es necesario exigir exactitud del dato espacial y la adopción de un rigor en su calidad que debe estar presente desde la fase de planeamiento y construcción de la base cartográfica para evitar vacío de información o superposición de unidades. Por ejemplo un mismo punto (que representa por ejemplo un evento de salud) puede estar dentro de diferentes tipos de "polígonos" que describen una condición contrapuesta de un territorio.

Los datos históricos de los eventos de salud a nivel de localidad o equivalente son necesarios para monitorear fenómenos que ocurren en el territorio particularmente cuando no se ajustan a las unidades de análisis tradicional.

La construcción de las tasas de un evento de salud en una serie de tiempo o en una comparación entre distintos periodos tiene la incertidumbre asociada a la disposición espacial de las unidades espaciales y/o a la delimitación de zonas urbanas y rurales.

De tal forma entre las consideraciones a ser tratadas para mejorar la calidad del dato espacial están:

- Estandarizar el procedimiento de integración de series temporales de un evento de salud a nivel desagregado con datos provenientes de otras fuentes (ej datos demográficos)
- Establecer un criterio para asignar la población de referencia en eventos de salud desagregados tanto en zonas urbanas como rurales.

PROGRAMAS

Para la elaboración de las bases gráficas se ha utilizado el programa Arcgis10.2 [ESRI, USA; de las bases temáticas se han utilizado Dbase (vers.3) y Excel (vers.5.0a). El análisis estadístico se llevó a cabo a través del programa EPIDAT 3.0

FIGURAS Y TABLAS

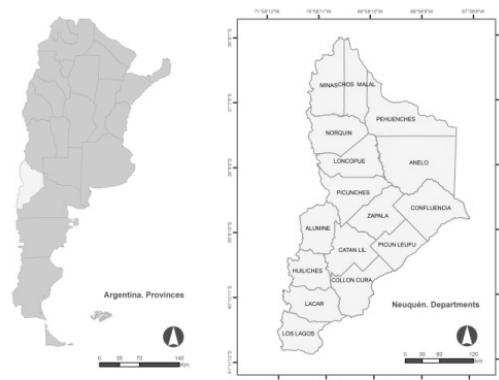


Figura 1. Localización del área de estudio

Tabla 1. Procedimiento metodológico para el cálculo de tasas de mortalidad a escala desagregada.

Defunciones acumuladas en periodo 2000 - 2006		Tasa de mortalidad x 1000 hab
Densidad poblacional por radio censal 2001		
$\geq 10 \text{ hab/km}^2$ → Radios censales contiguos → Agrupamiento subzonas		[<u>Núm. defunciones localizada en subzona</u> Núm. habitantes residentes en subzona
	↳ Radio censal aislados →	
$< 10 \text{ hab/km}^2$ → Radio censal	→	[<u>Núm. defunciones localizada en radio censal</u> Núm. habitantes residentes en radio censal
<hr/>		
Defunciones acumuladas en periodo 2007 - 2012		Tasa de mortalidad x 1000 hab
Densidad poblacional por radio censal 2010		
$\geq 10 \text{ hab/km}^2$ → Radios censales contiguos → Agrupamiento subzonas		[<u>Núm. defunciones localizada en subzona</u> Núm. habitantes residentes en subzona
	↳ Radio censal aislados →	
$< 10 \text{ hab/km}^2$ → Radio censal	→	[<u>Núm. defunciones localizada en radio censal</u> Núm. habitantes residentes en radio censal

REFERENCIAS BIBLIOGRÁFICAS

1. Agencia Estatal de Evaluación de las Políticas Públicas y la Calidad de los Servicios. (2010). Fundamentos de Evaluación de Políticas Públicas. Ministerio de Política Territorial y Administración Pública. NIPO:012-10-015-9. Madrid. http://www.aeval.es/comun/pdf/evaluaciones/Fundamentos_de_evaluacion.pdf
2. González Polanco, Liset & Pérez Betancourt, Guillermo. (2013). La minería de datos espaciales y su aplicación en los estudios de salud y epidemiología. *Revista Cubana de Información en Ciencias de la Salud*, 24(4), 482-489. Recuperado en 21 de junio de 2016, de http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2307-21132013000400010&lng=es&tlng=es
3. Universidad Nacional de Luján, Argentina. (2013) Análisis exploratorio de datos espaciales aplicado a mp10. Luján, Año 5, Número 5, Sección I: Artículos. pp. 109-128. <http://www.gesig-proeg.com.ar/documentos/revista-geosig/2013/ARTICULO-06-FUENZALIDA.pdf>.
4. De Pietri D., Dietrich P., Carcagno, A. & De Titto E. (2015). A Spatial Model of Qualitative Exposure, Province of Neuquén, Argentina. *J Environ Health Sci* 1(4): 1-9.
5. Consejo de Planificación y Acción para el Desarrollo, COPADE GIS. Sistema de Información Territorial - Visor Web. Gobierno de la Provincia de Neuquén. (2014). <http://copadesvr02.copade.neuquen.gov.ar/gis/inicio.html>
6. Naciones Unidas. (2000). Manual de sistemas de información geográfica y cartografía digital. Departamento de Asuntos Económicos y Sociales División de Estadística. Estudios de Métodos ST/ESA/STAT/SER.F/79. 228p. <http://unstats.un.org/unsd/publication/SeriesF/SeriesF79s.pdf>
7. Comisión Económica para América Latina y el Caribe, CEPAL. Naciones Unidas. (2012). Población, territorio y desarrollo sostenible. Ecuador. 243p. Disponible <http://www.cepal.org/celade/noticias/paginas/0/46070/2012-96-poblacion-web.pdf>

8. Dirven, Martine - Echeverri, Rafael - Sabalain, Cristina - Candia Baeza, David - Faiguenbaum, Sergio - Rodríguez, Adrián G. - Peña, Carolina (2011) Propuesta metodológica para una definición funcional de ruralidad. Hacia una nueva definición de "rural" con fines estadísticos en América Latina, Documentos de proyectos, N° 397 (LC/W.397).109 p.: grafs., tabs.Santiago de Chile. Comisión Económica para América Latina y el Caribe (CEPAL), <http://www.cepal.org/es/publicaciones/3858-nueva-definicion-rural-fines-estadisticos-america-latina>
9. Propuesta metodológica para una definición funcional de ruralidad. Hacia una nueva definición de "rural" con fines estadísticos en América Latina, Documentos de proyectos, N° 397 (LC/W.397). Santiago de Chile. Comisión Económica para América Latina y el Caribe (CEPAL), mayo 2011
10. De Pietri, Diana, Dietrich, Patricia, Mayo, Patricia, & Carcagno, Alejandro. (2011). Evaluación multicriterio de la exposición al riesgo ambiental mediante un sistema de información geográfica en Argentina. *Revista Panamericana de Salud Pública*, 30(4),377-387. <https://dx.doi.org/10.1590/S1020-49892011001000012>
11. Hervada Vidal, Xurxo, Santiago Pérez, M Isolina, Vázquez Fernández, Enrique, Castillo Salgado, Carlos, Loyola Elizondo, Enrique, & Silva Aycaguer, Luis Carlos. (2004). Epidat 3.0 programa para análisis epidemiológico de datos tabulados. *Revista Española de Salud Pública*, 78(2),277-280. <https://dx.doi.org/10.1590/S1135-57272004000200013>