

AEGIS: HERRAMIENTA DE ANÁLISIS EPIDEMIOLÓGICO EN UN SISTEMA DE INFORMACIÓN GEOGRÁFICA

V. Gómez-Rubio, A. López-Quílez y F. Verdejo

Departament d'Estadística i Investigació Operativa. Universitat de València

Introducción

Los Sistemas de Información están modificando la forma de realizar el análisis de grandes bancos de datos. Estos son recogidos de forma rutinaria y almacenados en una base de datos común. Los métodos de análisis de toda esta información son también incorporados al sistema como herramientas optimizadas para su explotación masiva. Los Sistemas de Información Geográfica (GIS) se centran en la componente espacial o geográfica de la información recogida, relacionando cada dato disponible con la localización donde fue medido o tomado. Técnicas específicas de Estadística Espacial se han desarrollado para el análisis de estos bancos de datos y se están incorporando ahora a los GIS, automatizando los estudios.

Las autoridades sanitarias, en sus tareas de vigilancia en Salud Pública, tienen en los GIS una herramienta fundamental para conocer cómo se extiende una enfermedad, estudiar su posible relación con un potencial foco de riesgo, o localizar un brote epidémico. La información necesaria para realizar este tipo de estudios proviene de muy diversas fuentes: registros de mortalidad, hospitales, facultativos, bases de datos oficiales, observatorios medioambientales o meteorológicos, proyectos específicos. Por tanto, es muy importante recopilar y tratar de forma unificada toda esta información para facilitar su acceso y análisis.

El interés creciente en el uso de GIS en Salud Pública se ha visto reflejado en la literatura con la aparición de monografías sobre estudios de Salud Pública realizados sobre un GIS (Gathrell y Löytönen, 1998; Briggs et al., 2002; Cromley y McLafferty, 2002), o sobre métodos estadísticos para la Epidemiología Espacial (Lawson et al., 1999; Elliot et al., 2000; Lawson, 2001; Haining, 2003), así como números especiales dedicados a estas materias en revistas como *Statistics in Medicine* (Falter et al., 1999; Sieber et al., 2001) o *Journal of the Royal Statistical Society* (Wakefield et al., 2001).

En este trabajo presentamos AEGIS, una nueva aplicación informática para realizar estudios epidemiológicos espaciales: cartografía de enfermedades, análisis de focos de riesgo, regresión ecológica, detección de zonas de riesgo elevado. Para ello AEGIS conecta y consulta las bases de datos, realiza los cálculos básicos de agregación y estandarización de tasas, suaviza las razones estandarizadas de mortalidad o incidencia, incorpora la información de covariables o de la ubicación de potenciales focos de riesgo, elabora mapas e informes de resultados, y conecta con otras aplicaciones para realizar

análisis más complejos. Por tanto AEGIS establece un entorno de funcionalidad GIS específicamente diseñado y adaptado para el análisis epidemiológico.

El desarrollo de AEGIS está basado en la experiencia acumulada por el proyecto EUROHEIS (Cockings y Järup, 2002), en el que se empleó la aplicación Rapid Inquiry Facility (RIF) inicialmente desarrollada por la Small Area Health Statistics Unit (Aylin et al., 1999) y posteriormente modificada para incluir el uso de covariables en los estudios (Gómez-Rubio et al., 2002). El RIF estaba basado en ArcView 3.2, Oracle Database 8i, y Oracle Forms and Reports. La necesidad de actualizar y mejorar la aplicación ha conducido a la construcción de una nueva aplicación que supera obstáculos anteriores.

¿Qué es AEGIS?

AEGIS es un sistema de información geográfico-sanitaria con una arquitectura cliente-servidor donde el servidor contiene una base de datos con información histórica sobre la estructura de población, mortalidad, morbilidad y otras variables socioeconómicas, medioambientales y meteorológicas, y el cliente es una aplicación Java que accede a la base de datos a través del JDBC o del ODBC y realiza diferentes tipos de estudio epidemiológico.

Como características principales se pueden destacar las siguientes:

- Es un sistema multiplataforma, es decir, funciona bajo casi cualquier sistema operativo y bajo casi cualquier tipo de máquina.
- El sistema de gestión de base de datos puede ser cualquiera (ORACLE, MySQL, ...) ya que mediante el JDBC y el ODBC se puede acceder a casi todos los sistemas de forma totalmente transparente al usuario.
- Sin coste adicional puesto que se ha desarrollado en Java y con herramienta de libre distribución y software abierto. Es posible usar un sistema de gestión de base de datos gratuito como MySQL, y aplicaciones GIS libres como GRASS para la explotación de los mapas.
- Está diseñado y optimizado para trabajar en red, proporcionando un mayor rendimiento.
- Permite la conexión con aplicaciones estadísticas como R, y de GIS como ArcGIS o GRASS.
- Para representar los resultados se emplea XML con las ventajas que ello aporta como poder editarlo con multitud de editores (cada vez más), el poder buscar documentos con mayor facilidad y el poder colocarlo directamente en una página Web.

- El mantenimiento y la instalación son sencillas. Cada usuario sólo tiene que instalar la aplicación en su PC y el mantenimiento se realiza desde la base de datos central sin que los usuarios tengan que actualizar su aplicación.
- Facilidad de uso: toda la información necesaria para realizar un estudio se introduce desde una pantalla con menús desplegables y ayudas.

Estudios epidemiológicos implementados

La herramienta está diseñada para realizar diferentes tipos de estudios epidemiológicos, y facilita la información necesaria y la conexión con otras aplicaciones para llevar a cabo otros análisis más complejos. De forma básica, AEGIS calcula una estandarización de tasas por distintos métodos y estima una suavización de riesgos por cada unidad geográfica, en relación con las covariables o la proximidad a un foco de riesgo. Por tanto, están implementados la cartografía de enfermedades, el análisis de focos de riesgo y el análisis de covariables.

Para cada uno de los estudios se realiza primero una estandarización directa o indirecta “filtrando” los efectos de un índice de privación u otra covariable. Los riesgos relativos son suavizados con el modelo Poisson-Gamma (Clayton y Kaldor, 1987) para proporcionar una visión general de la enfermedad analizada sin distorsiones causadas por la variabilidad de las regiones más pequeñas. El resultado de cada uno de estos estudios será uno o varios mapas donde se representan los resultados obtenidos y un informe en XML donde se detallan agrupados adecuadamente los valores estimados (estandarización, tasas crudas, suavizadas, etc.), sus correspondientes intervalos de confianza y los contrastes de hipótesis pertinentes.

La *cartografía de enfermedades* se emplea para representar la distribución espacial de la enfermedad (Ferrándiz et al., 2002). Un estadístico asociado a la enfermedad se calcula para cada unidad geográfica y se emplea categorizado para colorear el mapa. El mapa obtenido permite estudiar la mortalidad o la incidencia de la enfermedad en la población de estudio que hayamos elegido.

El *análisis de focos de riesgo* establece bandas alrededor de un punto o una zona geográfica representando una potencial fuente contaminante para comparar el riesgo de una enfermedad en cada una de esas bandas (Diggle y Elliot, 1995). Las unidades geográficas son agrupadas en función de la distancia a esos puntos y los mismos estadísticos son calculados para establecer si existe mayor riesgo en la zona más próxima.

El *análisis de covariables* realiza un estudio similar al anterior pero agrupando las observaciones por los valores de la covariable (Ferrándiz et al., 2004). De esta forma podemos describir la relación entre una determinada covariable y la enfermedad elegida, en la población de estudio.

Para llevar a cabo otros análisis estadísticos más complejos decidimos emplear el programa R (Ihaka y Gentleman, 1996) pues proporciona una serie de técnicas

de análisis espacial, permite la programación de otros métodos complementarios y es de libre distribución. Entre los desarrollos realizados están los paquetes *RArcInfo* (Gómez-Rubio, 2003) para la representación de mapas y *DCluster* (Gómez-Rubio et al., 2003) para detección de zonas de riesgo elevado. Además, es posible llamar desde el R directamente al programa *WinBUGS* (Spiegelhalter et al., 2002) para realizar estudios de cartografía de enfermedades y regresión ecológica desde la metodología de los modelos jerárquicos bayesianos con una flexibilidad mayor.

Descripción de la aplicación

El principal problema de este tipo de estudios es la gran cantidad de información requerida y la forma en la que se puede acceder a ella. Por eso, esta herramienta está diseñada para que la selección de los datos del estudio se realice de forma sencilla y rápida. Para elegir la población o área de estudio tenemos un árbol de municipios estructurado en comunidades autónomas y provincias, por orden alfabético. Podemos guardar o abrir el área seleccionada si vamos a trabajar con ella posteriormente e incluso podemos seleccionar el conjunto de municipios que tengamos en un fichero o en una base de datos local.

La herramienta permite seleccionar de forma cómoda el tipo de estandarización, directa o indirecta. En el caso de que sea directa, si la población de referencia será la población europea o mundial y en el caso de que sea indirecta podemos seleccionar una población de referencia de forma análoga a la población de estudio o podemos indicar que el área de referencia es igual a la de estudio.

También podemos añadir/eliminar una o varias covariables y modificarlas. Podemos seleccionarlas del conjunto de covariables disponibles en la base de datos central o de un fichero de texto o de una base de datos que tengamos en nuestro ordenador. De esta forma podemos realizar estudios sin esperar a que la covariable sea almacenada en la base de datos central. Para que sea más sencillo encontrar la covariable deseada tenemos un índice de covariables y una herramienta de búsqueda por palabras clave.

Al igual que para las covariables, para introducir el código de la enfermedad disponemos de una herramienta de búsqueda. Según los años elegidos, la clasificación CIE que aparecerá será la CIE 9 para estudios hasta 1998, la CIE 10 a partir de 1999 o unas agrupaciones CIE 9 - CIE 10 para los estudios en periodos en los que se combinan ambas clasificaciones. Además para cada uno de las codificaciones se pueden seleccionar grupos predefinidos de enfermedades.

La aplicación permite seleccionar de forma rápida el periodo de años, los grupos de edad y el sexo sobre los que se realizará el estudio. También debemos seleccionar la base de datos a emplear: mortalidad, morbilidad u otros registros disponibles. Adicionalmente podremos configurar la conexión a la base de datos, guardar o recuperar un estudio, conectarnos con el R, ArcGIS o programas similares para una mejor ma-

nipulación de los datos y guardar los mapas en formato shapefile o en formato imagen jpg o gif. También dispone de una ayuda para aprender a manejar la herramienta con varios ejemplos.

Agradecimientos

Este trabajo ha sido parcialmente financiado por la Dirección General de Salud Pública de la Generalitat Valenciana mediante un acuerdo de colaboración con la Universitat de València, y por la Comisión Europea a través del Proyecto EUROHEIS SI2.329122 (2001CVG2-604).

Referencias

- Aylin, P., Maheswaran, R., Wakefield, J., Cockings, S., Jarup, L., Arnold, R., Wheeler, G. y Elliot, P. (1999). A national facility for small area disease mapping and rapid initial assessment of apparent disease clusters around a point source: The U.K. Small Area Health Statistics Unit. *Journal of Public Health Medicine* 21(3):289–298.
- Briggs, D. J., Forer, P., Järup, L. y Stern, R. (eds.) (2002). *GIS for Emergency Preparedness and Health Risk Reduction*. Kluwer Academic Publishers.
- Clayton, D. y Kaldor, J. (1987). Empirical bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* 43:671–681.
- Cockings, S. y Järup, L. (2002). A European Project and Environment Information System for exposure and disease mapping and risk assessment (EUROHEIS). D. J. Briggs, P. Forer, L. Järup y R. Stern (eds.), *GIS for Emergency Preparedness and Health Risk Reduction*, volumen 11 de *NATO Sciences Series*, capítulo 11, 201–226. Kluwer Academic Publishers.
- Cromley, E. K. y McLafferty, S. L. (2002). *GIS and Public Health*. The Guilford Press.
- Diggle, P. y Elliot, P. (1995). Disease risk near point sources: Statistical issues for analyses using individual or spatially aggregated data. *Journal of Epidemiology and Community Health* 49:S20–S27. Suppl. 2.
- Elliot, P., Wakefield, J. C., Best, N. G. y Briggs, D. J. (2000). *Spatial Epidemiology: Methods and Applications*. Oxford University Press.
- Falter, K. H., Betts, D. R., Rolka, D. B., Rolka, H. R. y Sieber, W. K. (eds.) (1999). *Statistics in Medicine*, volumen 18 (23). Wiley and Sons, Inc. Special Issue: Symposium on statistical bases for Public Health decision making: from exploration to modelling.

- Ferrándiz, J., Abellán, J. J., Gómez-Rubio, V., López-Quílez, A., Sanmartín, P., Abellán, C., Martínez-Beneito, M. A., Melchor, I., Vanaclocha, H., Zurriaga, O., Ballester, F., Gil, J. M., Pérez-Hoyos, S. y Ocaña, R. (2004). Spatial analysis of the relationship between cardiovascular mortality and drinking water hardness. *Environmental Health Perspectives* 112(9).
- Ferrándiz, J., Abellán, J. J., López, A., Sanmartín, P., Vanaclocha, H., Zurriaga, O., Martínez-Beneito, M. A., Melchor, I. y Calabuig, J. (2002). Geographical distribution of cardiovascular mortality in Comunidad Valenciana (Spain). D. J. Briggs, P. Forer, L. Järup y R. Stern (eds.), *GIS for Emergency Preparedness and Health Risk Reduction*, volumen 11 de *NATO Sciences Series*, capítulo 15, 267–282. Kluwer Academic Publishers.
- Gathrell, A. y Löytönen, M. (eds.) (1998). *GIS and Health*. 6 GISDATA. Taylor & Francis.
- Gómez-Rubio, V. (2003). RArcInfo: Using G.I.S. data with R. K. Hornik, F. Leisch y A. Zeileis (eds.), *Proceedings of the 3rd International Workshop on Distributed Statistical Computing, March 20-22, 2003, Technische Universität Wien, Vienna, Austria*. ISSN 1609-395X.
URL <http://www.ci.tuwien.ac.at/Conferences/DSC-2003/Proceedings/>
- Gómez-Rubio, V., Ferrándiz, J. y López, A. (2002). Tr15-2002 - Implementation of covariate studies in an epidemiological G.I.S. T. Rep., Department of Statistics and Operational Research. University of Valencia.
- Gómez-Rubio, V., Ferrándiz, J. y López, A. (2003). DCluster: Detecting clusters of disease with R. K. Hornik, F. Leisch y A. Zeileis (eds.), *Proceedings of the 3rd International Workshop on Distributed Statistical Computing, March 20-22, 2003, Technische Universität Wien, Vienna, Austria*. ISSN 1609-395X.
URL <http://www.ci.tuwien.ac.at/Conferences/DSC-2003/Proceedings/>
- Haining, R. (2003). *Spatial Data Analysis. Theory and Practice*. Cambridge University Press.
- Ihaka, R. y Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 5(3):299–314.
- Lawson, A., Biggeri, A., Böhning, D., Lesaffre, E. y Viel, J. F. (eds.) (1999). *Disease Mapping and Risk Assessment for Public Health*. John Wiley and Sons Ltd.
- Lawson, A. B. (2001). *Statistical Methods in Spatial Epidemiology*. Wiley.
- Sieber, W. K., Green, T. A., Haugh, G. S., Kresnow, M., Luman, E. T. y Wilson, H. G. (eds.) (2001). *Statistics in Medicine*, volumen 20 (9-10). Wiley and Sons,

Inc. Special Issue: Symposium on Emerging Statistical Issues in Public Health for the 21st Century.

Spiegelhalter, D. J., Thomas, A., Best, N. G. y Lunn, D. (2002). *WinBUGS Version 1.4 User Manual*. MRC Biostatistics Unit.

URL <http://www.mrc-bsu.cam.ac.uk/bugs>

Wakefield, J., Quinn, M. y Raab, G. (eds.) (2001). *Journal of the Royal Statistical Society (Series A)*, volumen 164 (1). Blackwell.