

□

ANÁLISIS DE DATOS ESPACIALES EN EL ÁMBITO DE LA EPIDEMIOLOGÍA

Prof. Dr. Maria A Barceló y Prof. Dr. Marc Saez

8, 10, 14 y 16 de septiembre de 2021

Grupo de Investigación en Estadística, Econometría y Salud (GRECS), Universidad de Girona
CIBER de Epidemiología y Salud Pública (CIBERESP)

INTRODUCCIÓN AL CURSO

1. Introducción al curso
2. Introducción a la epidemiología y la estadística espacial
3. Panorámica de los modelos mixtos
4. Panorámica de los modelos mixtos - Prácticas
5. Introducción a INLA y R INLA
6. R INLA - Prácticas

Miércoles 8

Viernes 10

INTRODUCCIÓN AL CURSO

- 7. Mapas de enfermedades. Estandarización de razones de incidencia y mortalidad
- 8. Mapas de enfermedades. Suavización de razones de incidencia y de mortalidad estandarizadas
- 9. Mapas de enfermedades – Prácticas
- 10. Estudios de asociación geográfica. Regresión ecológica espacial
- 11. Regresión ecológica espacial - Prácticas

Martes 14

INTRODUCCIÓN AL CURSO

12. Agrupación de casos

13. Extensiones: BYM2, procesos puntuales, leaflet, pc priors

14. Extensiones – Prácticas

} Jueves 16

AGRUPACIÓN DE CASOS (CLUSTERING)

Métodos para el análisis de agregaciones

- Evaluación (detección) de la presencia de agrupaciones (clústeres)
- Localización de agrupaciones (clústeres)
- Agregaciones alrededor de una fuente

El análisis de agregaciones trabaja con datos de área o con procesos puntuales del suceso de salud.

AGRUPACIÓN DE CASOS (CLUSTERING)

¿Qué es un **clúster**?

- Según Knox (1989), *‘Un clúster es un grupo delimitado geográfica y/o temporalmente de ocurrencias de tamaño y concentración suficientes para que sea poco probable que hayan ocurrido por casualidad’.*
- Es decir, un clúster se produce cuando existe una tendencia general a una distribución más no aleatoria o "agrupada" de la enfermedad de lo que se esperaría que resultase de variaciones en la estructura de la población y fluctuaciones de probabilidad.

AGRUPACIÓN DE CASOS (CLUSTERING). ANÁLISIS DE HETEROGENEIDAD DEL RIESGO

- Existen diferencias entre los riesgos en cada una de las unidades espaciales analizadas
- No evalúan si existe una estructura espacial en estas diferencias
- **Causas de heterogeneidad:**
 - Presencia de una fuente de contaminación que provocaría el incremento de riesgo en su entorno
 - La variación espacial de un factor de riesgo (más riesgos relacionados con una mayor exposición a este factor)
 - Otras fuentes

AGRUPACIÓN DE CASOS (CLUSTERING) VS. IDENTIFICACIÓN DE FOCOS (CLUSTER DETECTION)

Disease Clustering and Cluster detection: conclusion

- **Clustering**

- ▶ We have defined a number of statistics (Moran, Tango) to determine the level of clustering in a set of data
- ▶ In the context of count data in spatial epidemiology these methods have some drawbacks due to non-constant variance of the response
 - when the response is the SMR, variance of the SMR depends on the expected numbers
- ▶ These methods are useful in an *exploratory* step in an analysis
- ▶ Hierarchical modelling provides greater information...

- **Cluster detection**

- ▶ The most popular method for cluster detection is the scan statistic
- ▶ Crucial choice: maximum size of the cluster (population)

AGRUPACIÓN DE CASOS (CLUSTERING) VS. IDENTIFICACIÓN DE FOCOS (CLUSTER DETECTION)

El análisis de agregación espacial (clustering) en general, consiste en evaluar (detectar) y localizar zonas donde los riesgos tiendan a ser mayores a los esperados.

La *identificación de focos (cluster detection)* consiste en la identificación de agrupaciones "inusuales" de casos, permite identificar hotspots.

IDENTIFICACIÓN DE FOCOS (CLUSTER DETECTION)

Cluster detection

- Methods discussed so far evaluate the tendency for global disease clustering, but no information on the location of the clusters
- Different methods available for cluster detection, depending on the nature of the data
- Methods using 'windows' to investigate spatial patterns:
 - ▶ Superimpose a number of circular windows on the region of interest
 - ▶ Determine whether the number of cases in each window is larger than expected
- Different methods define the circles in terms of:
 - ▶ distance (Openshaw)
 - ▶ number of cases (Besag and Newell)
 - ▶ population size (Scan statistic)
- These methods may be used as particular areas may be highlighted and subsequently investigated

IDENTIFICACIÓN DE FOCOS (CLUSTER DETECTION)

- Scan Statistics: familia de métodos
- Kulldorff
 - Establecimiento de una ventana circular (window) de tamaño variable. Puede estar basada en fracción de áreas o población
 - Comparar el riesgo (overall) de las regiones dentro y fuera de la ventana

IDENTIFICACIÓN DE FOCOS (CLUSTER DETECTION)

- Kulldorff
 - Hipótesis nula: no existe agregación = los dos riesgos son iguales
 - Hipótesis alternativa: el riesgo es más grande dentro de la ventana
 - Likelihood ratio test
 - El clúster más probable es aquel donde se encuentre el mayor likelihood ratio con la mejor significación

IDENTIFICACIÓN DE FOCOS (CLUSTER DETECTION) - KULLDORF

Scan statistics (Kulldorff)

- spatial scan statistic implemented in the software Satscan (<http://www.satscan.org/>)
- drawing all possible circles/ellipses in the study region with areas' centroids as the centers
- size of the circles/ellipses based on population or number of areas in the cluster
- determining whether there were more cases inside the circle than expected
- the most likely cluster is found by a likelihood ratio test



IDENTIFICACIÓN DE FOCOS (CLUSTER DETECTION) - KULLDORF

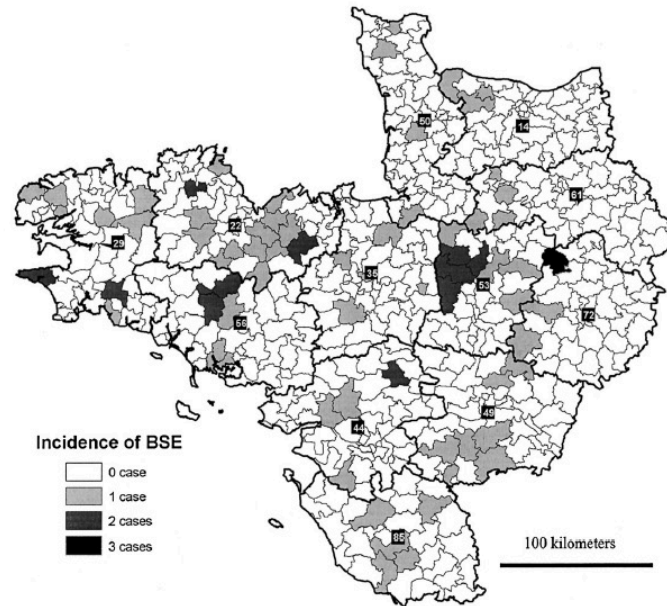


Figure 1. Incidence of bovine spongiform encephalopathy in western France (“canton” level) between August and December 2000.

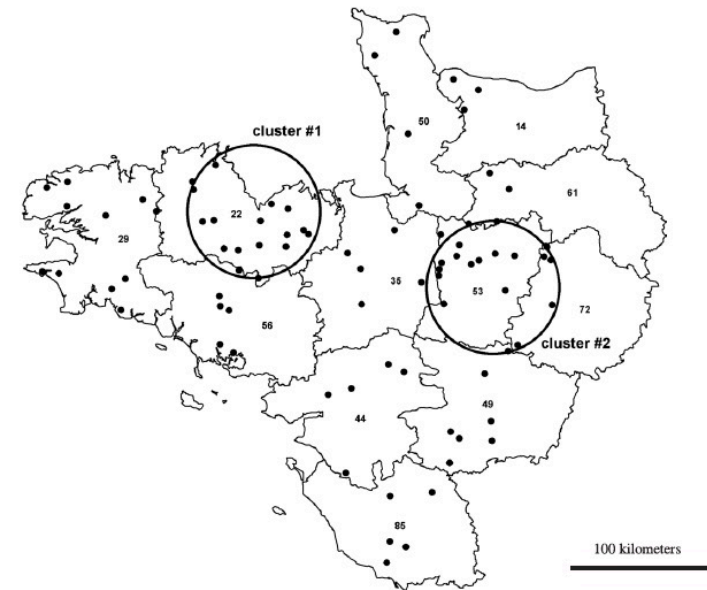


Figure 3. Spatial distribution of the 84 bovine spongiform encephalopathy cases (black dots) in western France between August and December 2000. Two clusters appeared with the method of Kulldorff: Cluster #1: 18 observed cases and 6 expected; 45 km radius. Cluster #2: 18 observed cases and 7 expected; 44 km radius.

Abrial D, Calavas D, Lauvergne N, Morignat E, Ducrot C. Descriptive spatial analysis of BSE in western France. *Vet Res.* 2003; 34(6):749-760. doi: 10.1051/vetres:2003032

IDENTIFICACIÓN DE FOCOS (CLUSTER DETECTION) - KULLDORF

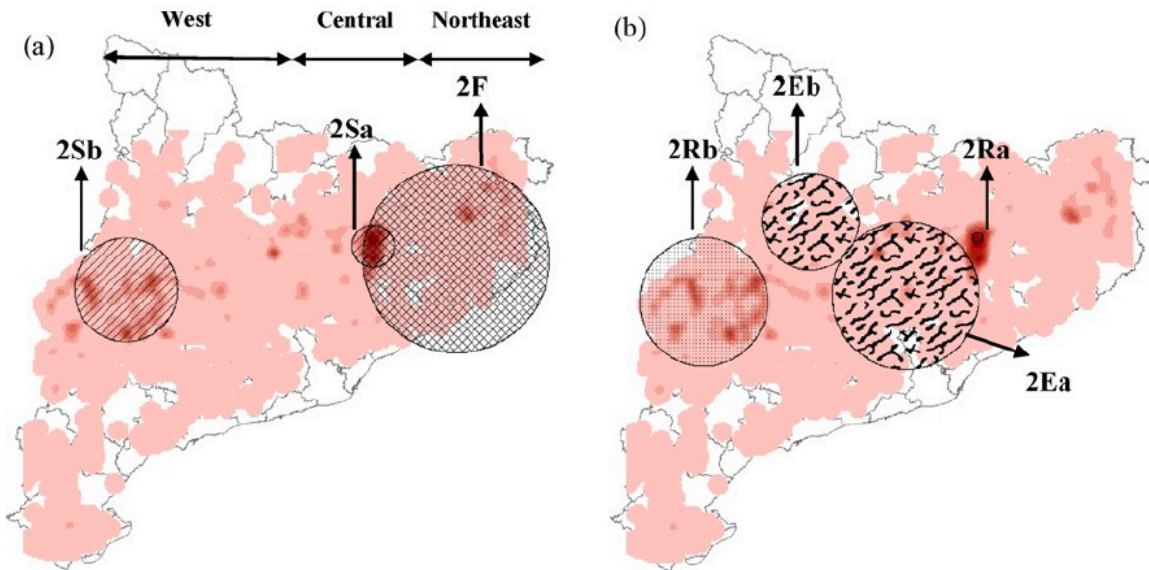
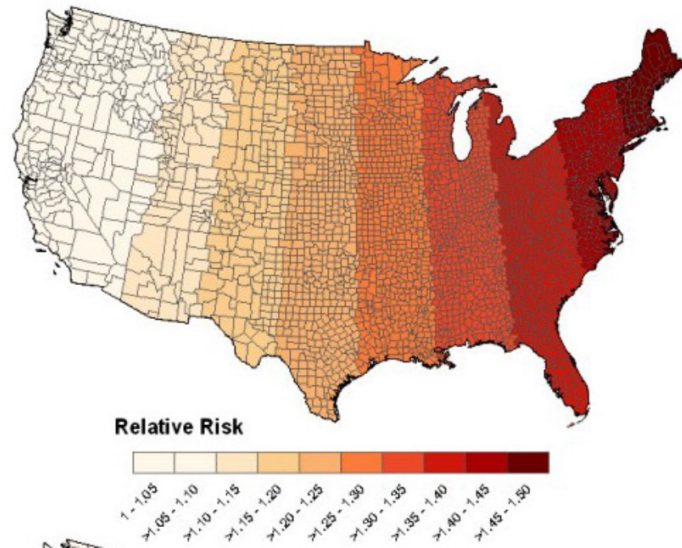


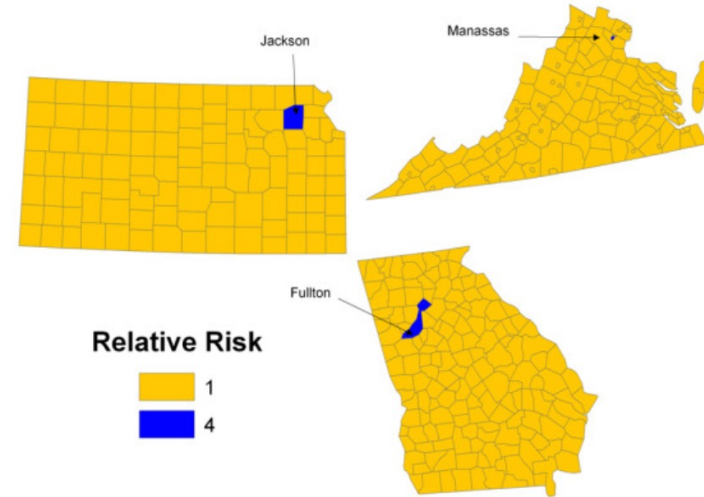
Fig. 2. Clusters identified in the second period (June 2004–May 2005) with the spatial scan statistic (Bernoulli model). (a) Clusters of positive sow farms (S) and fattening farms (F), (b) clusters of elimination (E) and reinfection (R). Clusters are represented over a Kernel density surface of pig farms.

Allepuz A, Saez M, Alba A, Napp S, Casal J. Exploratory spatial analysis of Aujeszky's disease during four phases of the eradication programme in Catalonia, Spain (2003-2007). *Prev Vet Med.* 2008; 86(1-2):164-75. doi: 10.1016/j.prevetmed.2008.04.005

IDENTIFICACIÓN DE FOCOS (CLUSTER DETECTION) - KULLDORF



Global clustering
From Jackson et al, 2009



Localised clusters

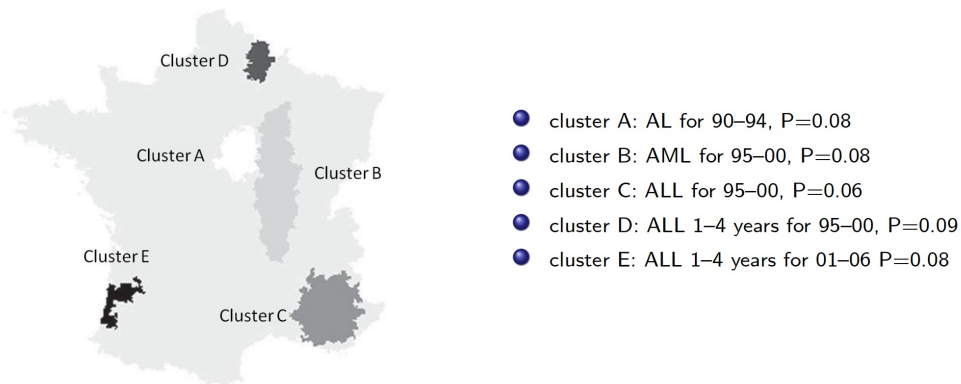
Jackson MC, Huang L, Luo J, Hackey M, Feuer E. Comparison of tests for spatial heterogeneity on data with global clustering patterns and outliers. *Int J Health Geogr.* 2009; 8:55. doi: 10.1186/1476-072X-8-55

IDENTIFICACIÓN DE FOCOS (CLUSTER DETECTION) - KULLDORF

Example: Childhood acute leukaemia in France, 1990-2006

Demoury et al, 2012

- This study did not find evidence of any global spatial heterogeneity of AL incidence rates in France (Potthoff-Whittinghill test)
- Although no significant spatial cluster was detected over the whole period, the study identified a few significant spatial clusters in specific periods
- Even though the significance levels of those clusters do not strongly support the existence of local risk factors, the clusters may still reflect a slight impact of shared risk factors



Demoury C, Goujoun-Bellec S, Guyot-Goubin A, Hémon D, Clavel J. Spatial variations of childhood acute leukaemia in France, 1990-2006: global spatial heterogeneity and cluster detection at 'living-zone' level. *Eur J Cancer Prev.* 2012; 21(4):367-374. doi: 10.1097/CEJ.0b013e32834e31d8

IDENTIFICACIÓN DE FOCOS (CLUSTER DETECTION) - KULLDORF

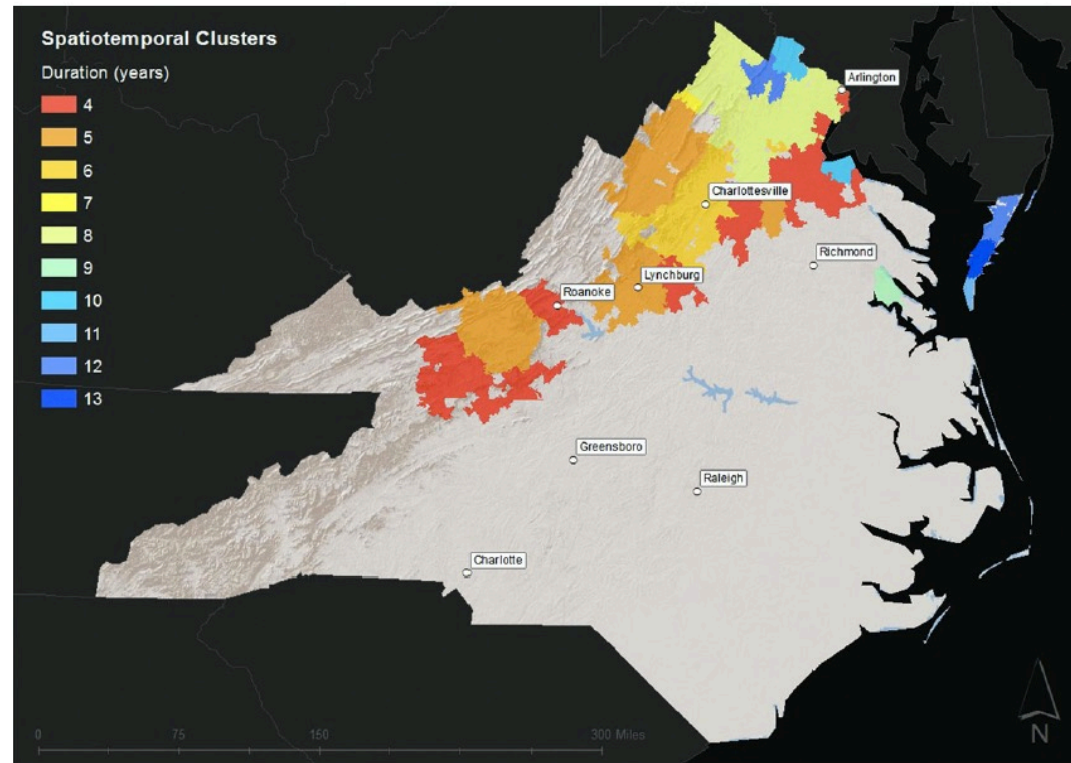


Figure 3. Spatiotemporal scanning statistic cluster analysis. This analysis was limited to spatial clusters of at least 3 years duration and a P value of $\leq .0001$, ensuring that only significant and durable trends were identified. Clusters of longest duration were in northern and eastern Virginia, and the most recently emergent clusters extended along the Appalachian Mountains towards southwest Virginia.

Lantos PM, Nigrovic LE, Auwaerter PG, Fowler VG, Ruffin F, Brinkerhoff RJ, Reber J, Williams C, Broyhill J, Pan WK, Gaines DN. Geographic Expansion of Lyme Disease in the Southeastern United States, 2000-2014. *Open Forum Infect Dis.* 2015; 2(4):ofv143. doi: 10.1093/ofid/ofv143

IDENTIFICACIÓN DE FOCOS (CLUSTER DETECTION) - KULLDORF

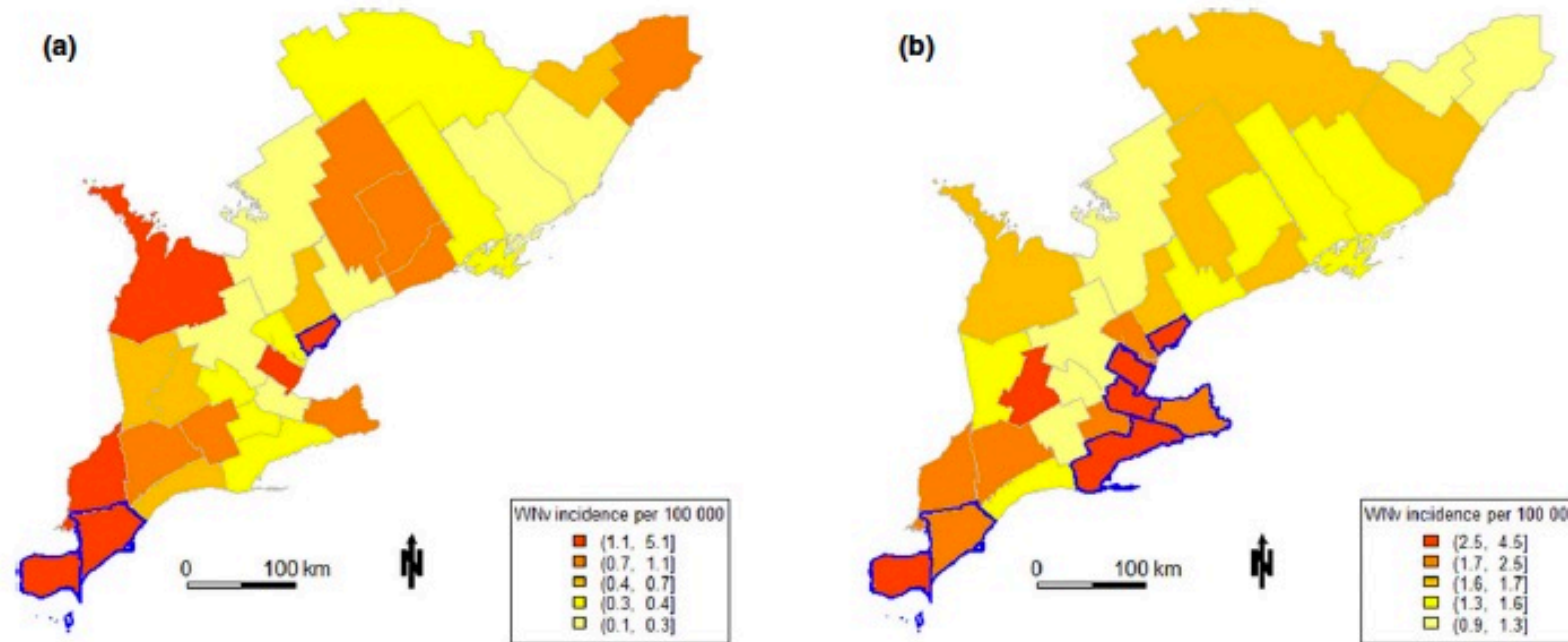
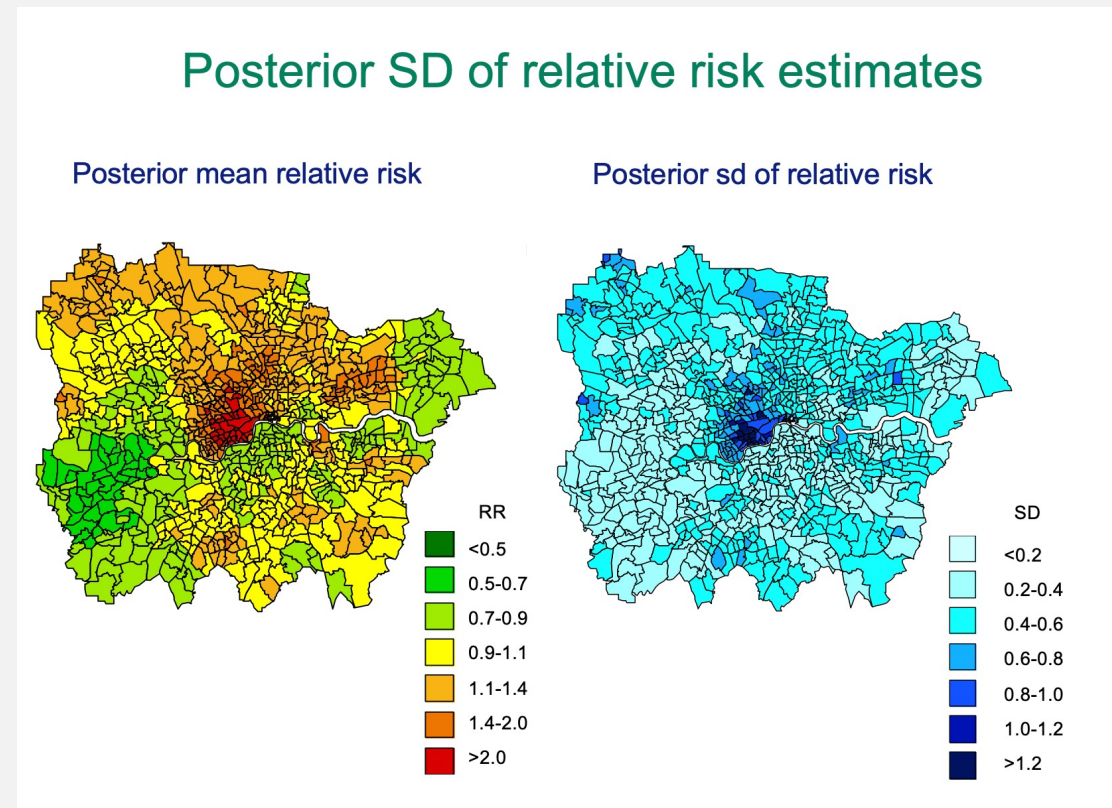


Fig. 1. Choropleth maps of empirical Bayesian smoothed annual human risk incidence estimates of WNV disease per 100 000 population for the 29 health units of southern Ontario for the years a) 2005, b) 2012 and c) 2012 adjusted. Outlined areas represent health units where clusters of disease were identified by the spatial scan statistic.

Thompson M, Berke O. Evaluation of the control of West Nile virus in Ontario: Did risk patterns change from 2005 to 2012?. *Zoonoses Public Health*. 2017; 64(2):100-105. doi: 10.1111/zph.12285

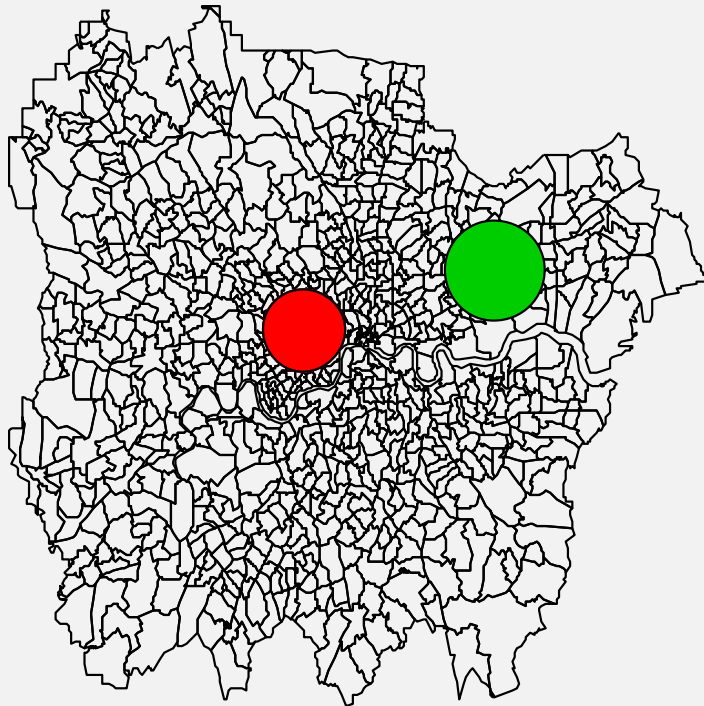
IDENTIFICACIÓN DE FOCOS (CLUSTER DETECTION) - KULLDORF

Ejemplo: Childhood leukaemia incidence in London, 1986-1998

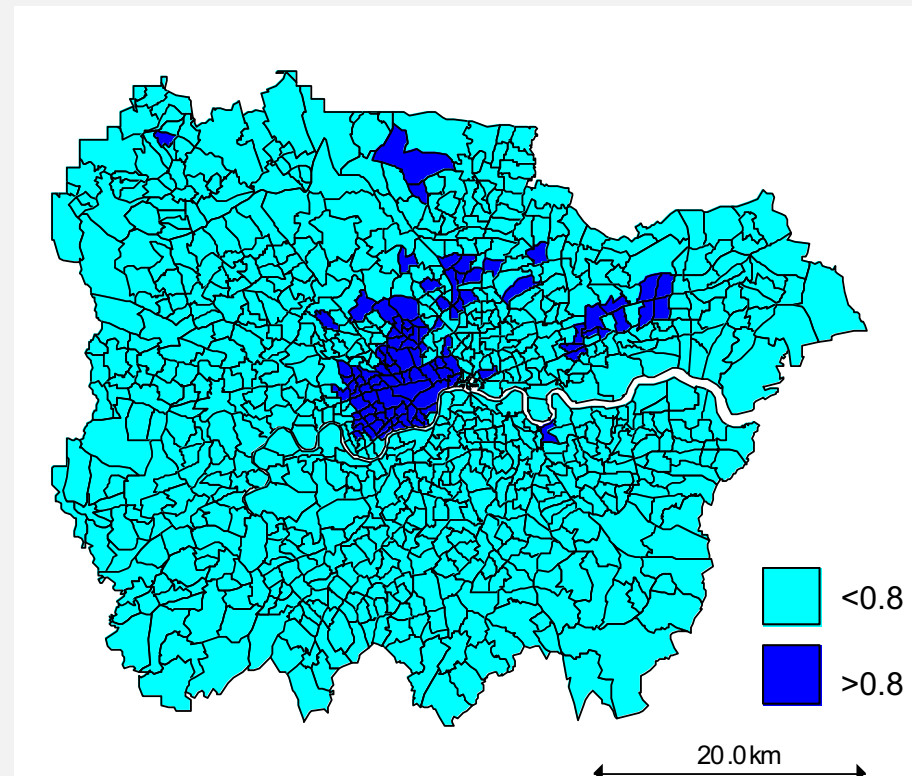


IDENTIFICACIÓN DE FOCOS (CLUSTER DETECTION) - KULLDORF

Ejemplo: Childhood leukaemia incidence in London, 1986-1998



- Most likely cluster; $p < 0.001$
- 2nd most likely cluster; $p = 0.2$



Teaching statistics



Doing Statistics

