

□

ANÀLISI DE DADES ESPACIALS EN L'ÀMBIT DE L'EPIDEMIOLOGIA

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

8, 10, 14 i 16 de setembre de 2021

Grup de Recerca en Estadística, Econometria i Salut (GRECS), Universitat de Girona
CIBER d'Epidemiologia i Salut Pública(CIBERESP)

INTRODUCCIÓ AL CURS

1. Introducció al curs
2. Introducció a l'epidemiologia i l'estadística espacial
3. Panoràmica del models mixtos
4. Panoràmica del models mixtos - Pràctiques
5. Introducció a INLA i R INLA
6. R INLA - Pràctiques

Dimecres 8

Divendres 10

INTRODUCCIÓ AL CURS

- 7. Mapes de malalties. Estandardització de raons d'incidència i mortalitat
- 8. Mapes de malalties. Suavització de raons d'incidència i de mortalitat estandarditzades
- 9. Mapes de malalties – Pràctiques
- 10. Estudis d'associació geogràfica. Regressió ecològica espacial
- 11. Regressió ecològica espacial - Pràctiques

Dimarts 14

INTRODUCCIÓ AL CURS

12. Agrupació de casos

13. Extensions: BYM2, processos puntuals, leaflet, pc priors

14. Extensions – Pràctiques

} Dijous 16

AGRUPACIÓ DE CASOS (CLUSTERING)

Mètodes per a l'anàlisi d'agregacions

- Evaluació (detecció) de la presència d'agrupacions (clústers)
- Localització d'agrupaciones (clústers)
- Agregacions al voltant d'una font

L'anàlisi d'agregacions treballa amb dades d'àrea o amb processos puntuals de l'esdeveniment de salut.

AGRUPACIÓ DE CASOS (CLUSTERING)

¿Què és un **clúster**?

- Segons Knox (1989), *‘Un clúster és un grup delimitat geogràfica i/o temporalment d’ocurrències de mida i concentració suficients perquè sigui poc probable que hagin ocorregut per casualitat’.*
- És a dir, un clúster es produeix quan hi ha una tendència general a una distribució més no aleatòria o "agrupada" de la malaltia del que s’esperaria que resultés de variacions en l’estructura de la població i fluctuacions de probabilitat.

AGRUPACIÓ DE CASOS (CLUSTERING). ANÀLISI D'HETEROGENEÏTAT DEL RISC

- Existeixen diferències entre els riscos en cadascuna de les unitats espacials analitzades
- No avaluen si existeix una estructura espacial en aquestes diferències
- **Causes d'heterogeneïtat:**
 - Presència d'una font de contaminació que provocaria l'increment de risc en el seu entorn
 - La variació espacial d'un factor de risc (més riscos relacionats amb una major exposició a aquest factor)
 - Altres fonts

AGRUPACIÓ DE CASOS (CLUSTERING) VS. IDENTIFICACIÓ DE FOCUS (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)

AGRUPACIÓ DE CASOS (CLUSTERING) VS. IDENTIFICACIÓ DE FOCUS (CLUSTER DETECTION)

L'anàlisi d'agregació espacial (clustering) en general, consisteix a avaluar (detectar) i localitzar zones on els riscos tendeixen a ser majors als esperats.

La *identificació de focus (cluster detection)* consisteix en la identificació d'agrupacions "inusuals" de casos, permet identificar hotspots.

IDENTIFICACIÓ DE FOCUS (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

IDENTIFICACIÓ DE FOCUS (CLUSTER DETECTION)

- Scan Statistics: família de mètodes
- Kulldorff
 - Establiment d'una finestra circular (window) de mida variable. Pot estar basada en fracció d'àrees o població
 - Comparar el risc (overall) de les regions dins i fora de la finestra

IDENTIFICACIÓ DE FOCUS (CLUSTER DETECTION)

- Kulldorff
 - Hipòtesi nul·la: no hi ha agregació = els dos riscos són iguals
 - Hipòtesi alternativa: el risc és més gran dins de la finestra
 - Likelihood ratio test
 - El clúster més probable és aquell on es trobi el major likelihood ratio amb la millor significació

IDENTIFICACIÓ DE FOCUS (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



IDENTIFICACIÓ DE FOCUS (CLUSTER DETECTION) - KULLDORF

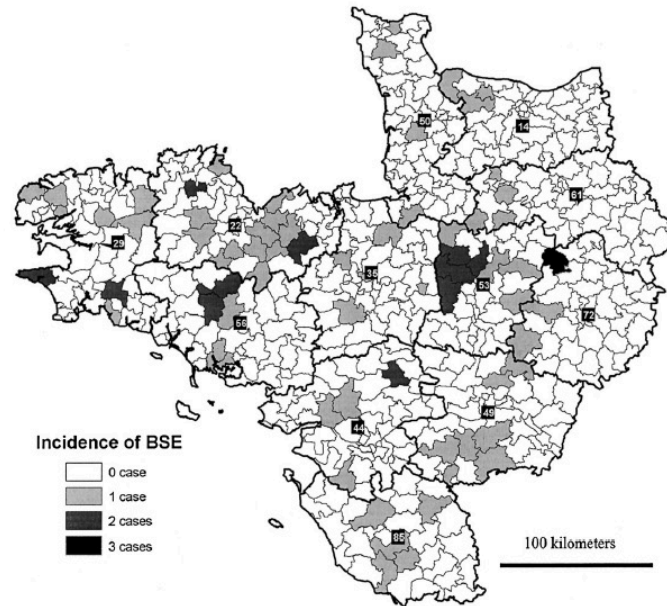


Figure 1. Incidence of bovine spongiform encephalopathy in western France (“canton” level) between August and December 2000.

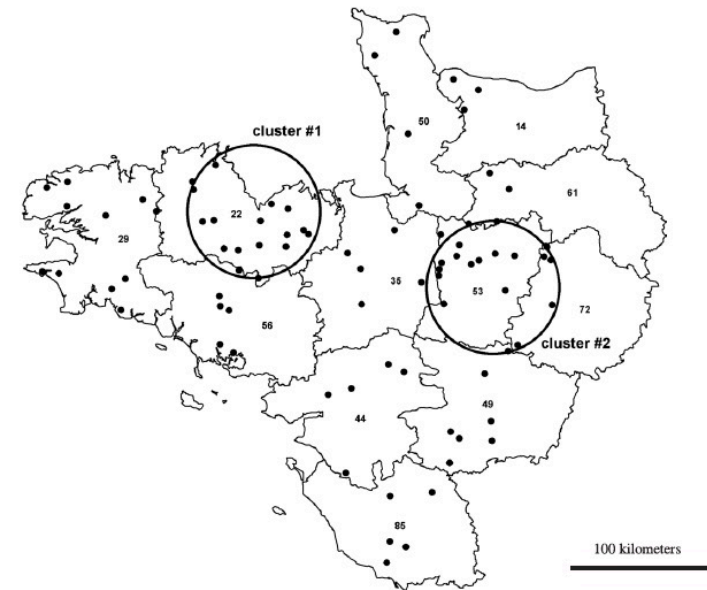


Figure 3. Spatial distribution of the 84 bovine spongiform encephalopathy cases (black dots) in western France between August and December 2000. Two clusters appeared with the method of Kulldorff: Cluster #1: 18 observed cases and 6 expected; 45 km radius. Cluster #2: 18 observed cases and 7 expected;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

IDENTIFICACIÓ DE FOCUS (CLUSTER DETECTION) - KULLDORF

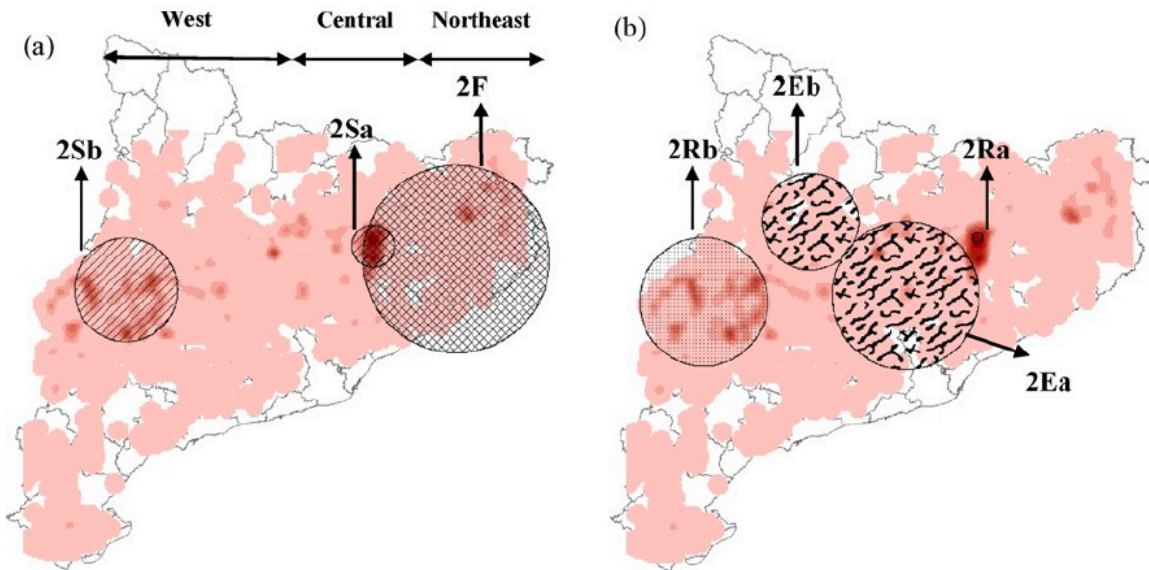
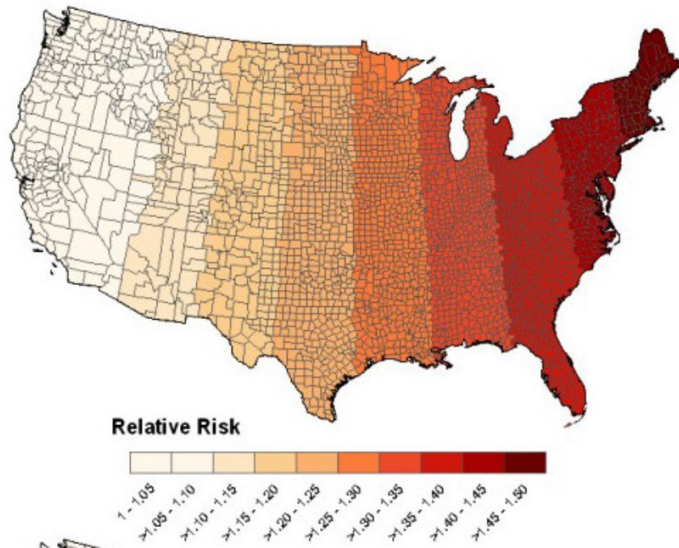


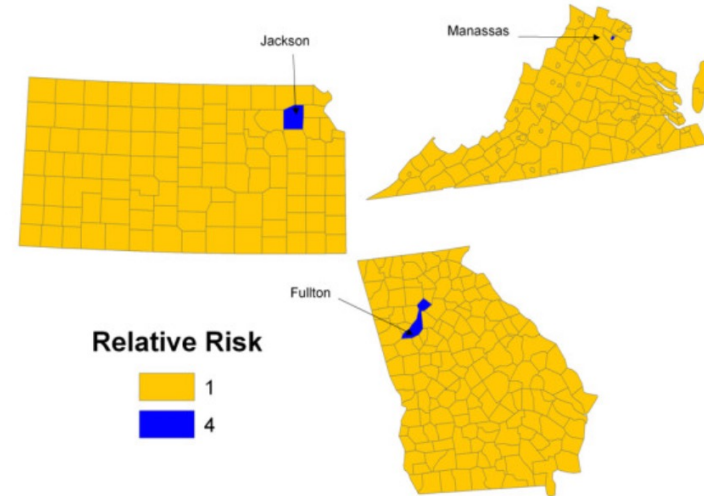
Fig. 2. Clusters identified in the second period (June 2004–May 2005) with the spatial scan statistic (Bernoulli model). (a) Clusters of positive sow farms (S) and fattening farms (F), (b) clusters of elimination (E) and reinfection (R). Clusters are represented over a Kernel density surface of pig farms.

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

IDENTIFICACIÓ DE FOCUS (CLUSTER DETECTION) - KULLDORF



Global clustering
From Jackson et al, 2009



Localised clusters

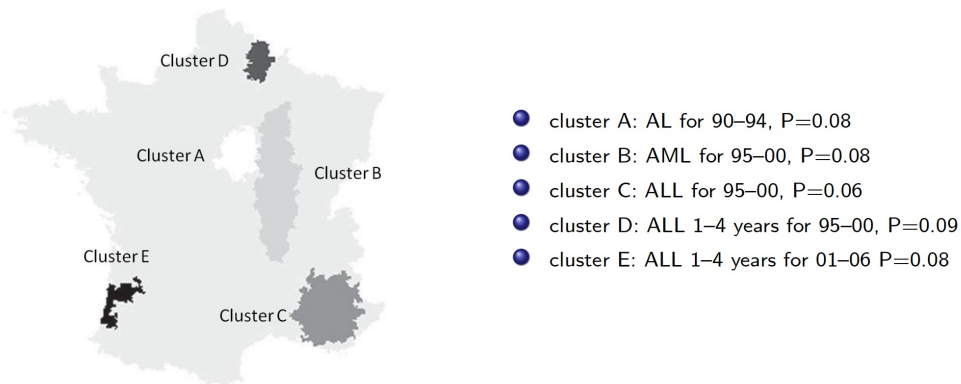
Jackson MC, Huang L, Luo J, Hackey M, Feuer E. Comparison of tests for spatial heterogeneity on data with global clustering patterns and outliers. *Int J Health Geogr.* 2009; 8:55. doi: 10.1186/1476-072X-8-55

IDENTIFICACIÓ DE FOCUS (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

IDENTIFICACIÓ DE FOCUS (CLUSTER DETECTION) - KULLDORF

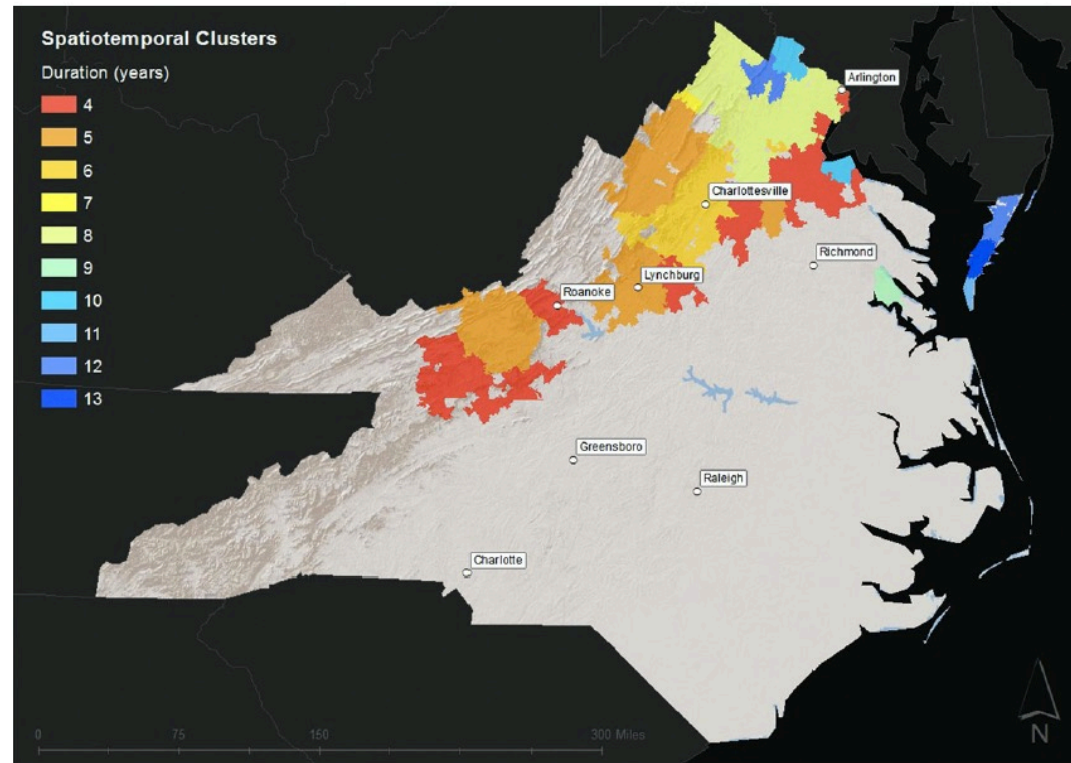


Figure 3. Spatiotemporal scanning statistic cluster analysis. This analysis was limited to spatial clusters of at least 3 years duration and a P value of $\leq .0001$, ensuring that only significant and durable trends were identified. Clusters of longest duration were in northern and eastern Virginia,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

IDENTIFICACIÓ DE FOCUS (CLUSTER DETECTION) - KULLDORF

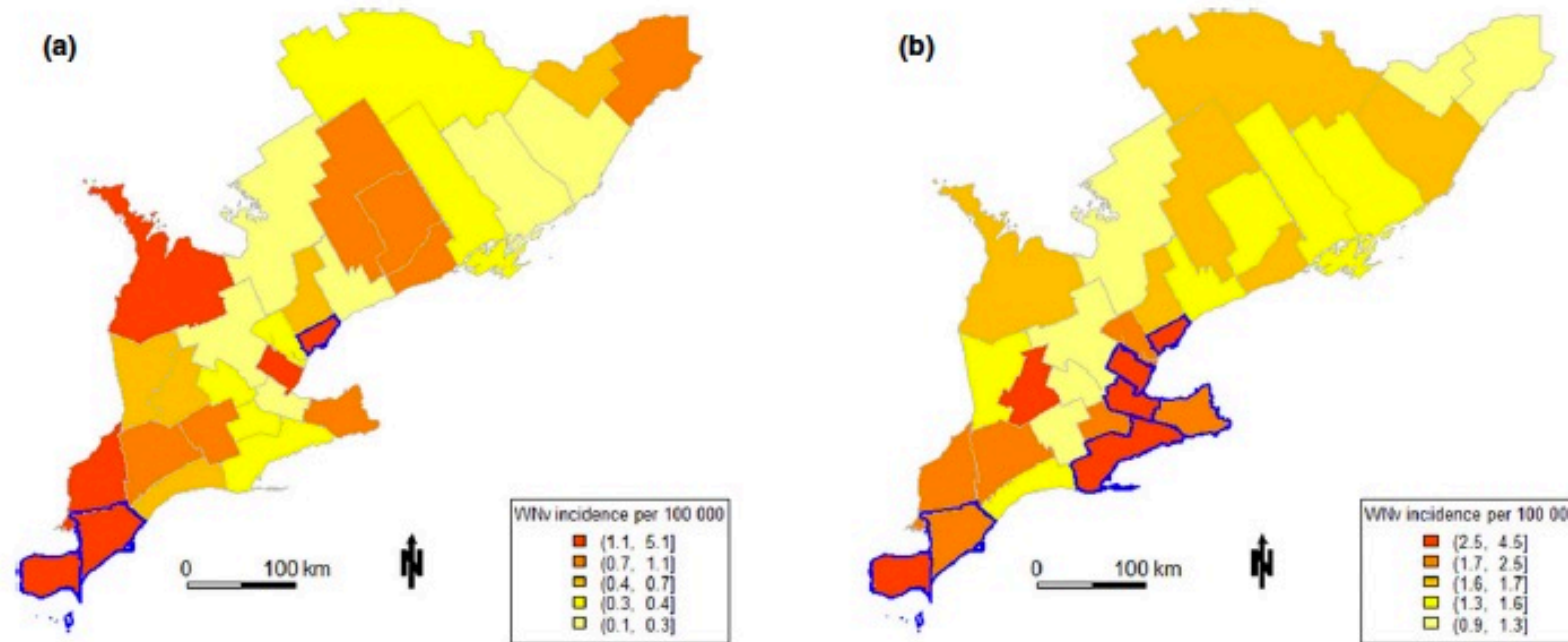
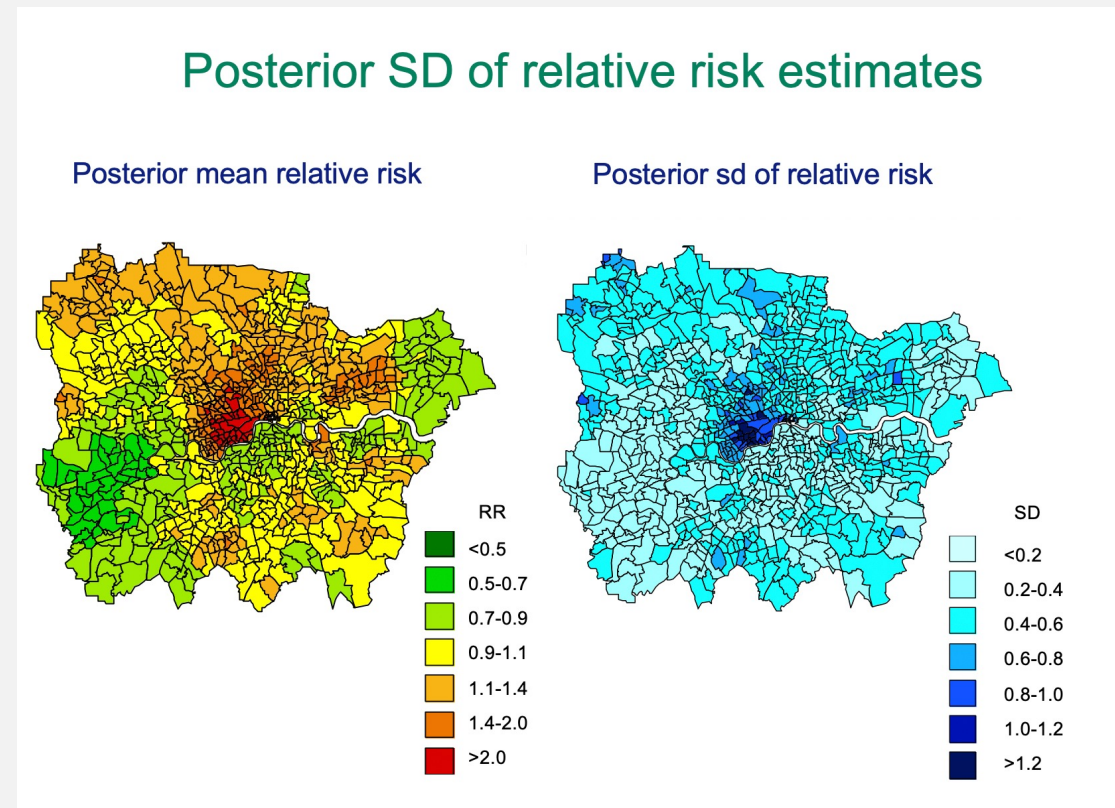


Fig. 1. Choropleth maps of empirical Bayesian smoothed annual human risk incidence estimates of WNV disease per 100 000 population for the 29 health units of southern Ontario for the years a) 2005, b) 2012 and c) 2012 adjusted. Outlined areas represent health units where clusters of disease were identified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

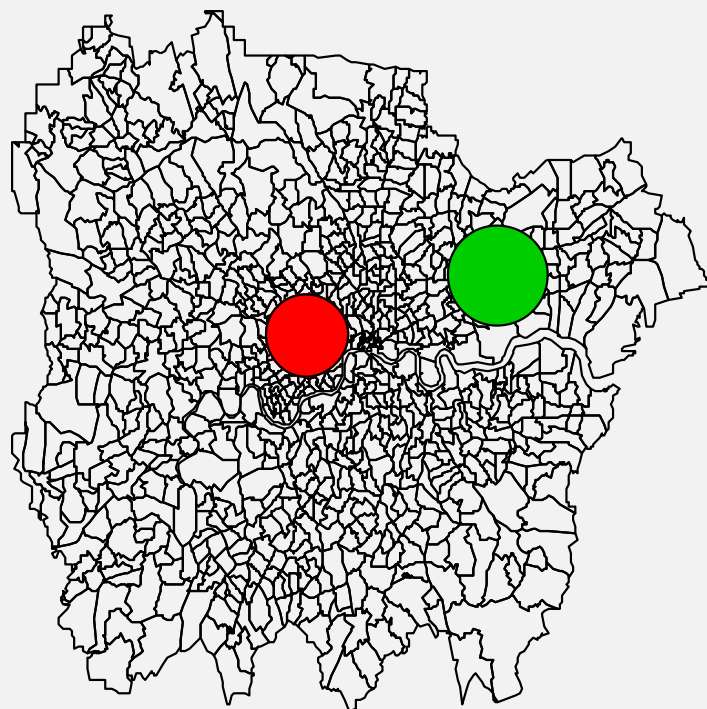
SUAVITZACIÓ DE RAONS D'INCIDÈNCIA I MORTALITAT ESTANDARDITZADES

Exemple: Childhood leukaemia incidence in London, 1986-1998

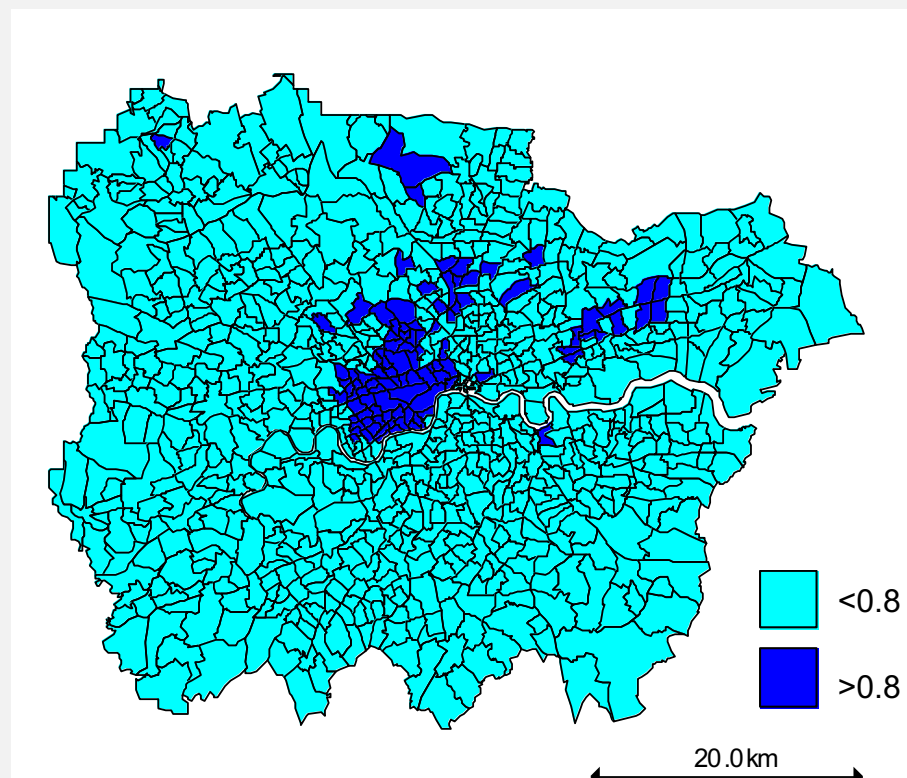


IDENTIFICACIÓ DE FOCUS (CLUSTER DETECTION) - KULLDORF

Exemple: Childhood leukaemia incidence in London, 1986-1998



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



Teaching statistics



Doing Statistics

