

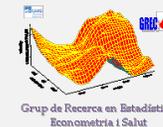
□

ESTADÍSTICA Y EPIDEMIOLOGÍA ESPACIAL

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

8 de octubre de 2020

Research Group on Statistics, Econometrics and Health (GRECS)
CIBER of Epidemiology and Public Health (CIBERESP)



Grup de Recerca en Estadística,
Econometria i Salut



ESTADÍSTICA Y EPIDEMIOLOGÍA ESPACIAL

1. Epidemiología y Epidemiología Espacial
2. Estadística Espacial
3. Mapas de Enfermedades
4. Regresión ecológica
5. Agregación espacial (clustering)
6. Identificación de focos (cluster detection)



Epidemiología Espacial

ESTADÍSTICA Y EPIDEMIOLOGÍA ESPACIAL

1. **Epidemiología y Epidemiología Espacial**
2. Estadística Espacial
3. Mapas de Enfermedades
4. Regresión ecológica
5. Agregación espacial (clustering)
6. Identificación de focos (cluster detection)



Epidemiología Espacial

1. EPIDEMIOLOGÍA Y EPIDEMIOLOGÍA ESPACIAL

Epidemiología

- La **epidemiología** se dedica al estudio de la distribución, frecuencia, causas y control de los factores relacionados con la salud y la enfermedad en poblaciones humanas bien definidas y a la aplicación de este estudio para defender y mejorar la salud de la población.
- La **epidemiología** estudia, sobre todo, la relación entre exposición y enfermedad.
- La **epidemiología** se considera la ciencia básica para la medicina preventiva y una fuente de información para la formulación de políticas de salud pública.

1. EPIDEMIOLOGÍA Y EPIDEMIOLOGÍA ESPACIAL

- Las enfermedades no se producen de forma aleatoria. Tienen causas, muchas de ellas de origen humano, que pueden evitarse.
- Los métodos epidemiológicos han sido cruciales para identificar numerosos factores etiológicos que, en su momento, han justificado la formulación de políticas sanitarias encaminadas a la prevención de enfermedades.

1. EPIDEMIOLOGÍA Y EPIDEMIOLOGÍA ESPACIAL

- La epidemiología surgió del estudio de las epidemias de las enfermedades infecciosas.
- En la actualidad, la epidemiología se interesa por el estudio demográfico de cualquier enfermedad con la ayuda de la estadística.

1. EPIDEMIOLOGÍA Y EPIDEMIOLOGÍA ESPACIAL

Epidemiología espacial

“Spatial epidemiology is the description and analysis of geographic variations in disease with respect to demographic, environmental, behavioral, socioeconomic, genetic, and infectious risk factors”.

Elliot P, et al. **Spatial epidemiology: current approaches and futures challenges.** *Environ. Health Perspect.* 2004.

1. EPIDEMIOLOGÍA Y EPIDEMIOLOGÍA ESPACIAL

- La **epidemiología espacial** está relacionada con la descripción de las enfermedades y el estudio de las causas y prevención de las mismas, utilizando diferentes perspectivas de análisis en las que la localización de los eventos es una componente fundamental, ya que estudia también las variaciones geográficas de las enfermedades.
- La **epidemiología espacial** es un subcampo de la epidemiología centrado en el estudio de la distribución espacial de los resultados en salud.
- La **epidemiología espacial** se basa en un concepto de salud, en el cual los individuos son vistos en su contexto socio-cultural-ambiental.

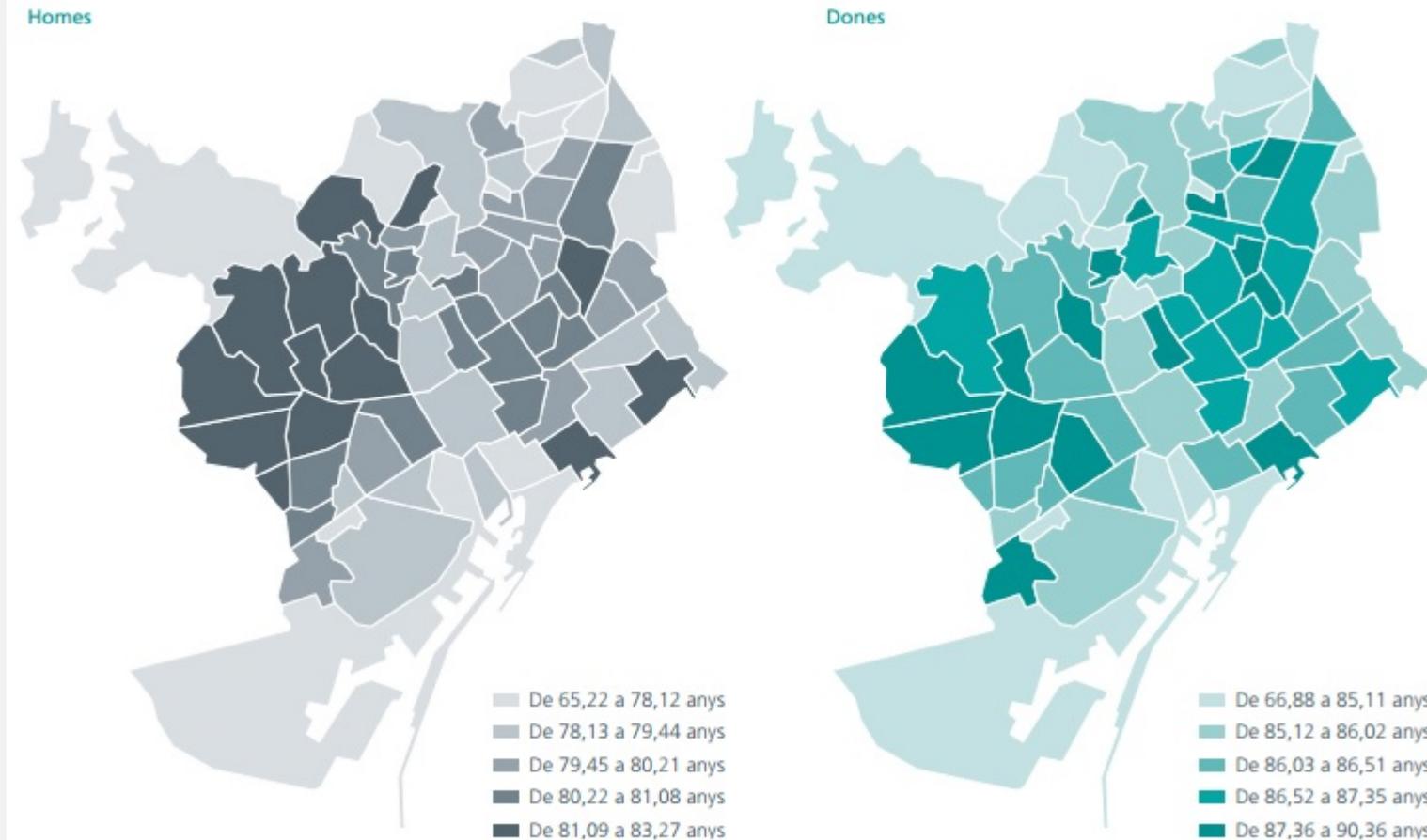
1. EPIDEMIOLOGÍA Y EPIDEMIOLOGÍA ESPACIAL

Principal objetivo de la epidemiología espacial:

- Mostrar qué parte de la variación espacial de la distribución de la ocurrencia de un evento de salud no está explicada ni por la distribución espacial de factores explicativos conocidos, ni por una variación aleatoria.
- De hecho, a menudo nos interesa encontrar pistas sobre algún factor de riesgo desconocido de una determinada enfermedad.

1. EPIDEMIOLOGÍA Y EPIDEMIOLOGÍA ESPACIAL

Figura 1. Esperança de vida en homes i dones als barris. Barcelona, 2008-2012.



Font: Registre de Mortalitat de Barcelona. Agència de Salut Pública de Barcelona.

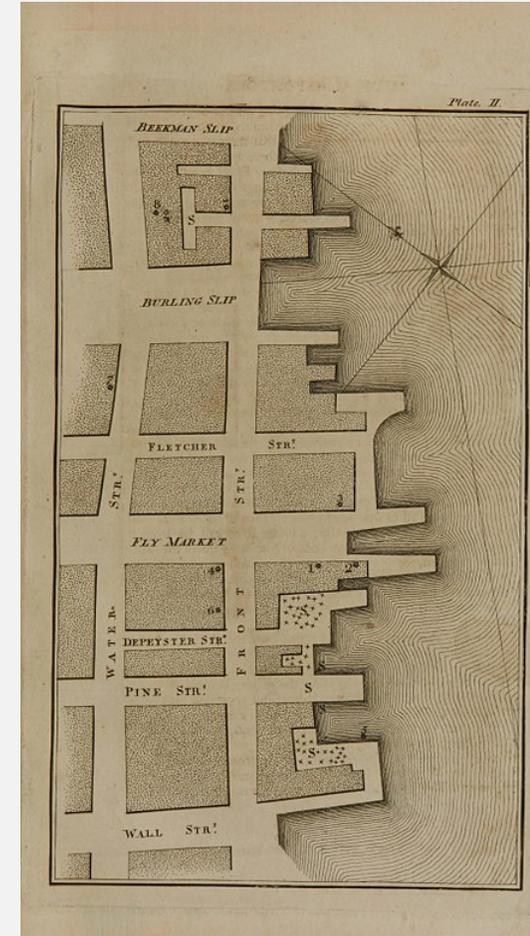
1. EPIDEMIOLOGÍA Y EPIDEMIOLOGÍA ESPACIAL

- El estudio de la distribución geográfica de los eventos de salud ha sido de gran interés entre los epidemiólogos durante las últimas décadas, aunque como veremos a continuación, tiene una historia de más de 200 años.
- Los **primeros mapas de enfermedades** se hicieron para representar la ubicación de casos de enfermedades infecciosas:
 - Fiebre amarilla en Nueva York (Seaman, 1798)
 - Cólera en Londres (Snow, 1854)

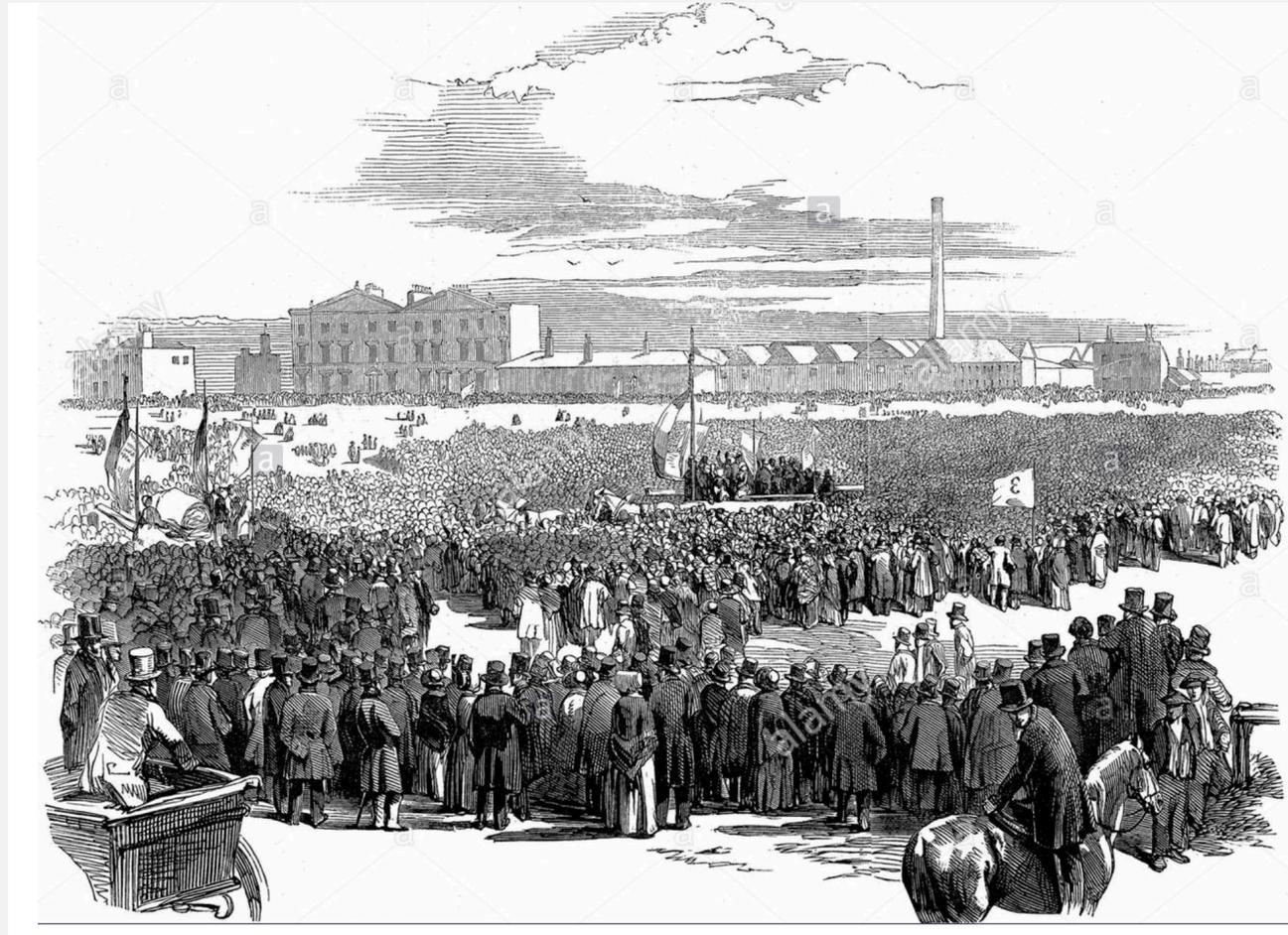
1. EPIDEMIOLOGÍA Y EPIDEMIOLOGÍA ESPACIAL

Mapa de Seaman de casos de fiebre amarilla

Los puntos representan las muertes por fiebre amarilla y las cruces, los vertederos



1. EPIDEMIOLOGÍA Y EPIDEMIOLOGÍA ESPACIAL



1. EPIDEMIOLOGÍA Y EPIDEMIOLOGÍA ESPACIAL

- Durante el otoño de 1848, se produjo una segunda epidemia de cólera en Inglaterra, causando una gran mortalidad.
- No se conocía con certeza la etiología, ni el modo de transmisión de esta enfermedad.
- Dos corrientes teóricas:
 - Contagio por contacto
 - Miasmas

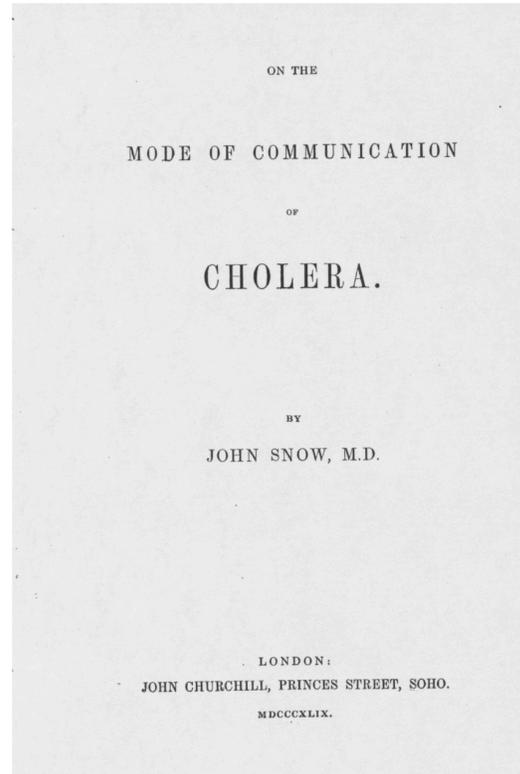
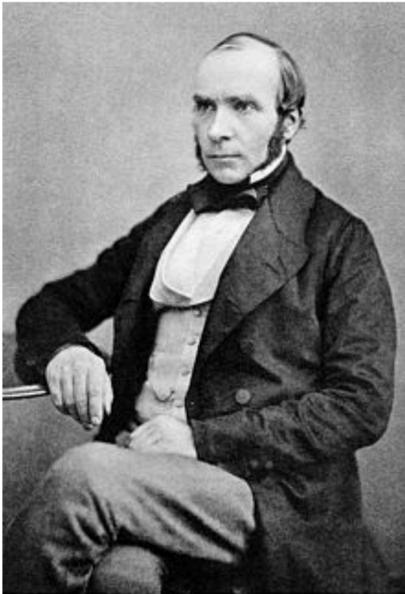
1. EPIDEMIOLOGÍA Y EPIDEMIOLOGÍA ESPACIAL

Snow observó:

- **Miasmas no podían causar la enfermedad:** los pacientes deberían presentar síntomas respiratorios producto de la inhalación de los “miasmas” y no el síndrome diarreico agudo presente en el cólera.
- **Defunciones por cólera** ocurridas entre 1848-49 (segunda epidemia) estaban **concentradas** en los distritos de la **zona sur de Londres**. Las tasas de mortalidad observadas en esta zona eran muy superiores a la del resto de la ciudad (8,0 y 2,4 defunciones por 1.000 habitantes, respectivamente).

1. EPIDEMIOLOGÍA Y EPIDEMIOLOGÍA ESPACIAL

John Snow
1813-1858



Deaths from Cholera in London, registered from September 23d, 1848, to August 25th, 1849.

Districts of London.	Population in 1841.	Deaths from Cholera.	Deaths to each 1,000 inhabitants.
West . .	300,711	533	1.77
North . .	375,971	415	1.10
Central . .	373,605	920	2.48
East . . .	392,444	1,597	4.06
South . .	502,548	4,001	7.95
Total . .	1,948,369	7,466	3.83

On the Mode of Communication of Cholera. John Snow. 1849

1. EPIDEMIOLOGÍA Y EPIDEMIOLOGÍA ESPACIAL

Además, Snow observó que:

- Los habitantes de la zona sur de Londres obtenían agua para beber río abajo del Támesis (aguas altamente contaminadas).
- Las demás zonas de Londres obtenían el agua para beber de sectores menos contaminados del río (aguas arriba del mismo o de afluentes).

1. EPIDEMIOLOGÍA Y EPIDEMIOLOGÍA ESPACIAL

- En 1853-1854, se da la tercera epidemia de cólera en Londres.
- Los habitantes de ciertos distritos del sur de la ciudad extraían el agua de pequeños afluentes del río Támesis o bien a partir de numerosas bombas de agua de uso público.
- Los desechos humanos eran vertidos en improvisadas alcantarillas o directamente al río.
- Había dos compañías responsables del abastecimiento: Sothwark and Vauxhall Company y Lambeth Water Company.
- Durante la segunda epidemia de cólera de 1848-49, ambas compañías extraían el agua de sectores contaminados del Támesis, presentando similar número de muertes los distritos abastecidos por una y otra compañía.

1. EPIDEMIOLOGÍA Y EPIDEMIOLOGÍA ESPACIAL

- En 1853:
 - Lambeth Water Company había trasladado sus instalaciones río arriba (aguas no contaminadas).
 - Southwark and Vauxhall Water Company seguía aguas abajo.
- Snow observó que la tasa de mortalidad por cólera en hogares abastecidos por la segunda compañía era 8,5 veces mayor a la de hogares abastecidos por la primera.

1. EPIDEMIOLOGÍA Y EPIDEMIOLOGÍA ESPACIAL

- A principios de septiembre de 1854:
 - En el sector de Londres “Golden Square” (en el Soho), se produjo un brote epidémico de cólera de inusual intensidad (500 muertos en tan solo 10 días).
 - La mayoría de los residentes del sector extraían el agua a partir de una bomba de uso público ubicada en Broad Street.

1. EPIDEMIOLOGÍA Y EPIDEMIOLOGÍA ESPACIAL

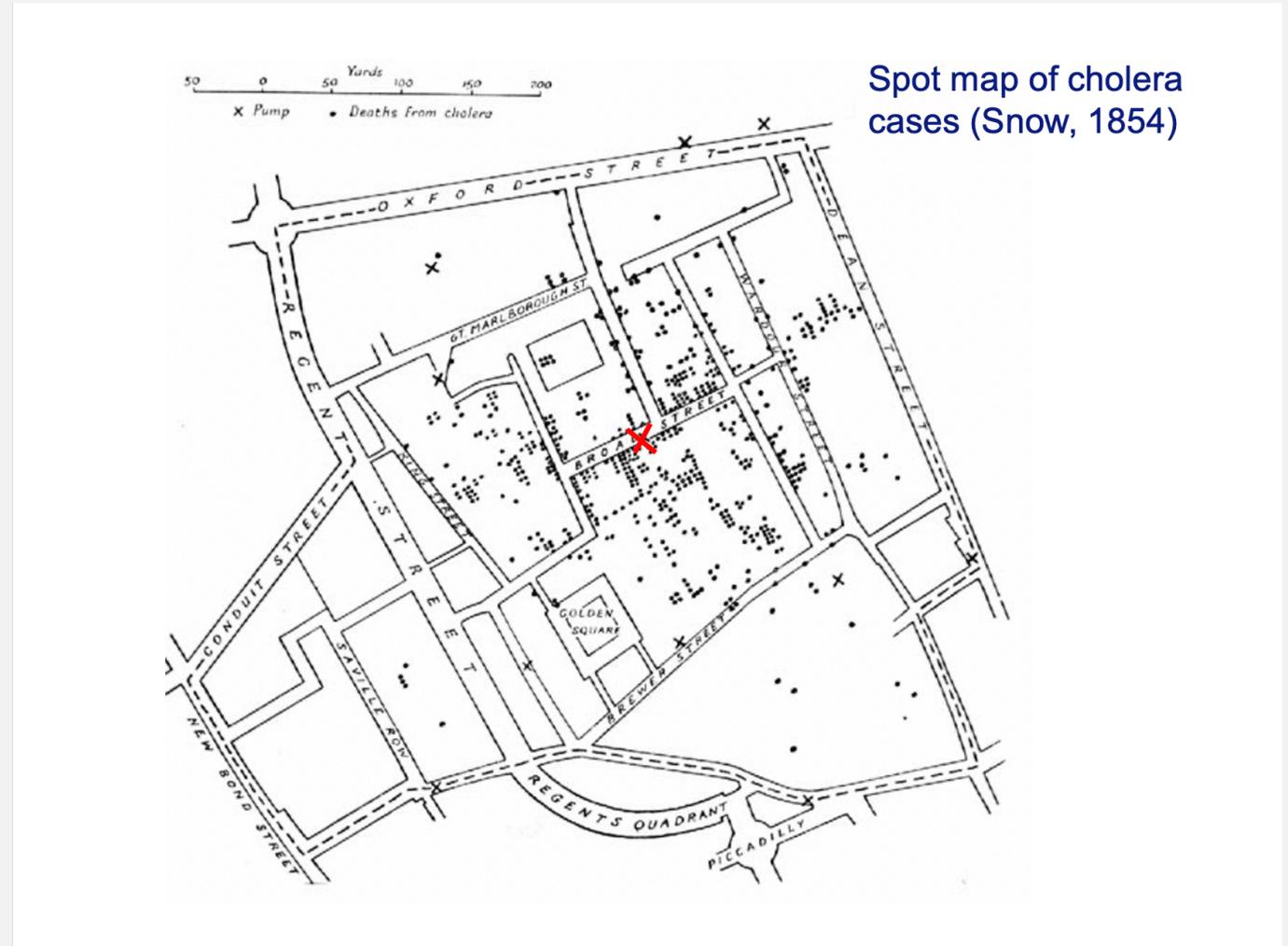


1. EPIDEMIOLOGÍA Y EPIDEMIOLOGÍA ESPACIAL

- realizó un **mapa del sector**, marcando los puntos correspondientes a defunciones por cólera y las distintas bombas de agua potable existentes.
- planteó que este brote se debía a la ingestión de aguas contaminadas procedentes de la bomba de Broad Street.

1. EPIDEMIOLOGÍA Y EPIDEMIOLOGÍA ESPACIAL

Snow identificó la agregación espacial de los casos de cólera en Londres en 1857.



1. EPIDEMIOLOGÍA Y EPIDEMIOLOGÍA ESPACIAL

- tomó muestras de la bomba de Broad Street y otras 4 bombas cercanas (había diferencias en relación con la claridad del agua).
- calculó la distancia entre la residencia de cada difunto y la bomba de agua más cercana.
- observó que en 73 de 83 casos la bomba de Broad Street era la más cercana y comunicó esto a las autoridades sanitarias y estas cerraron la bomba.

1. EPIDEMIOLOGÍA Y EPIDEMIOLOGÍA ESPACIAL

Evolución de los mapas en epidemiología espacial

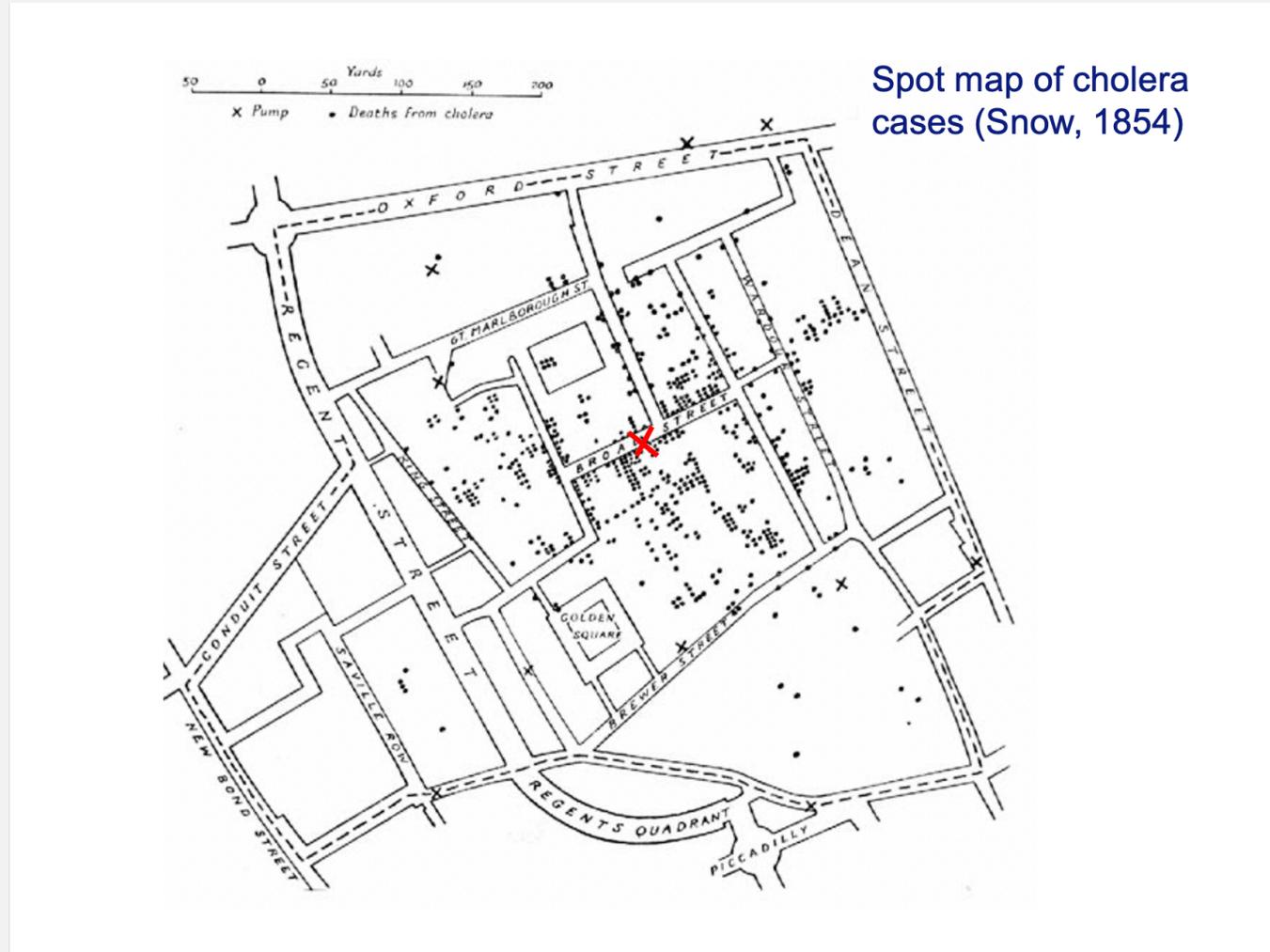
- Mapas de puntos
- Mapas de coropletas
- Atlas de enfermedades, a nivel nacional e internacional

1. EPIDEMIOLOGÍA Y EPIDEMIOLOGÍA ESPACIAL

Evolución de los mapas en epidemiología espacial

- Mapas de puntos
 - Fiebre amarilla en Nueva York (Seaman, 1798)
 - Cólera en Londres (Snow, 1854)

1. EPIDEMIOLOGÍA Y EPIDEMIOLOGÍA ESPACIAL



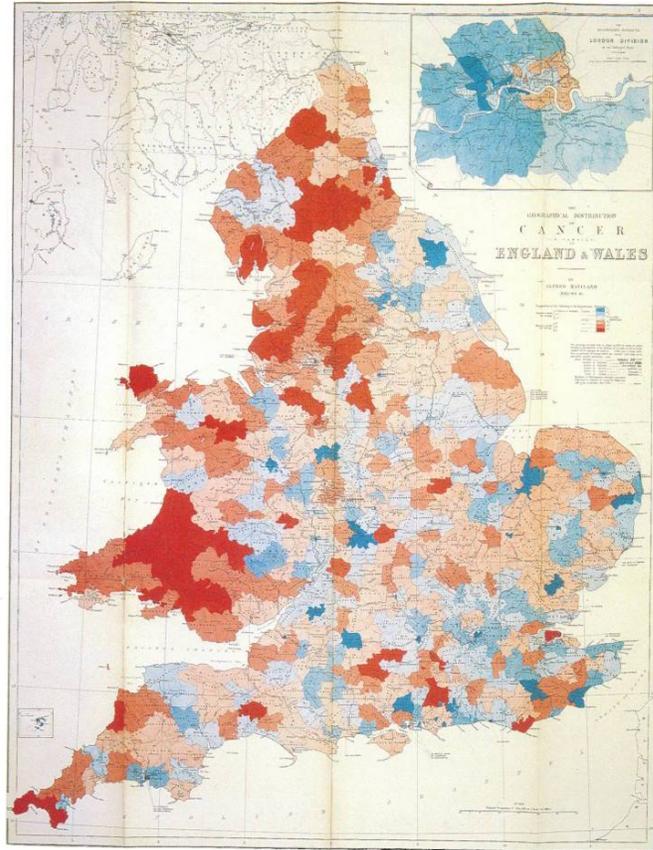
1. EPIDEMIOLOGÍA Y EPIDEMIOLOGÍA ESPACIAL

Evolución de los mapas en epidemiología espacial

- Mapas de coropleta
 - Distribución geográfica de la mortalidad por enfermedad cardíaca, cáncer y tuberculosis en Inglaterra y Gales (Haviland, 1878)
 - Mortalidad por cáncer por condado en Inglaterra y Gales, ajustada por edad y sexo (Stocks, 1936,1937,1939)

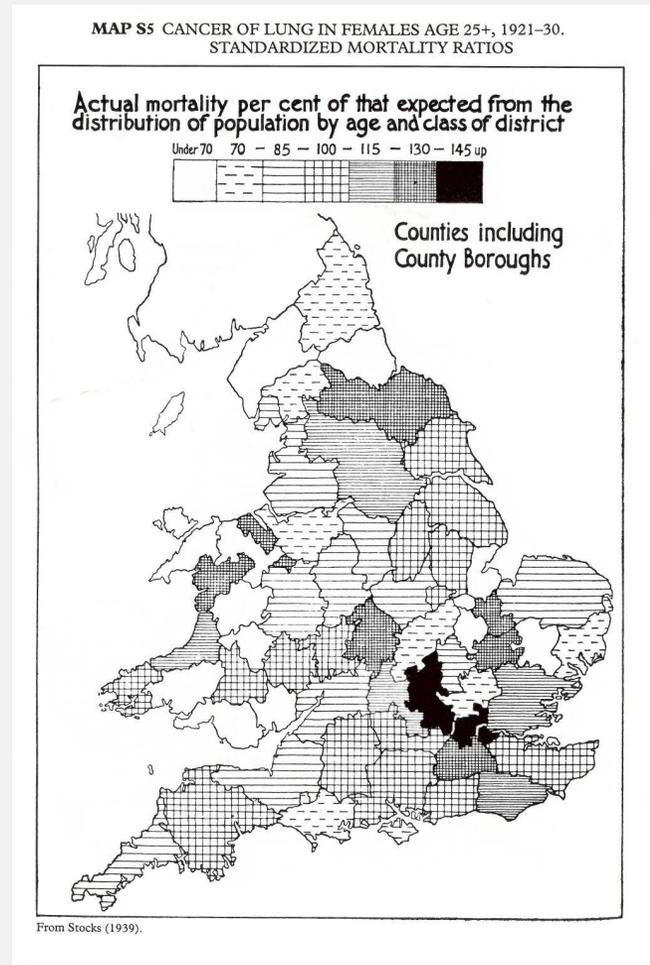
1. EPIDEMIOLOGÍA Y EPIDEMIOLOGÍA ESPACIAL

Cáncer en mujeres,
1851-1860
(Haviland, 1878)



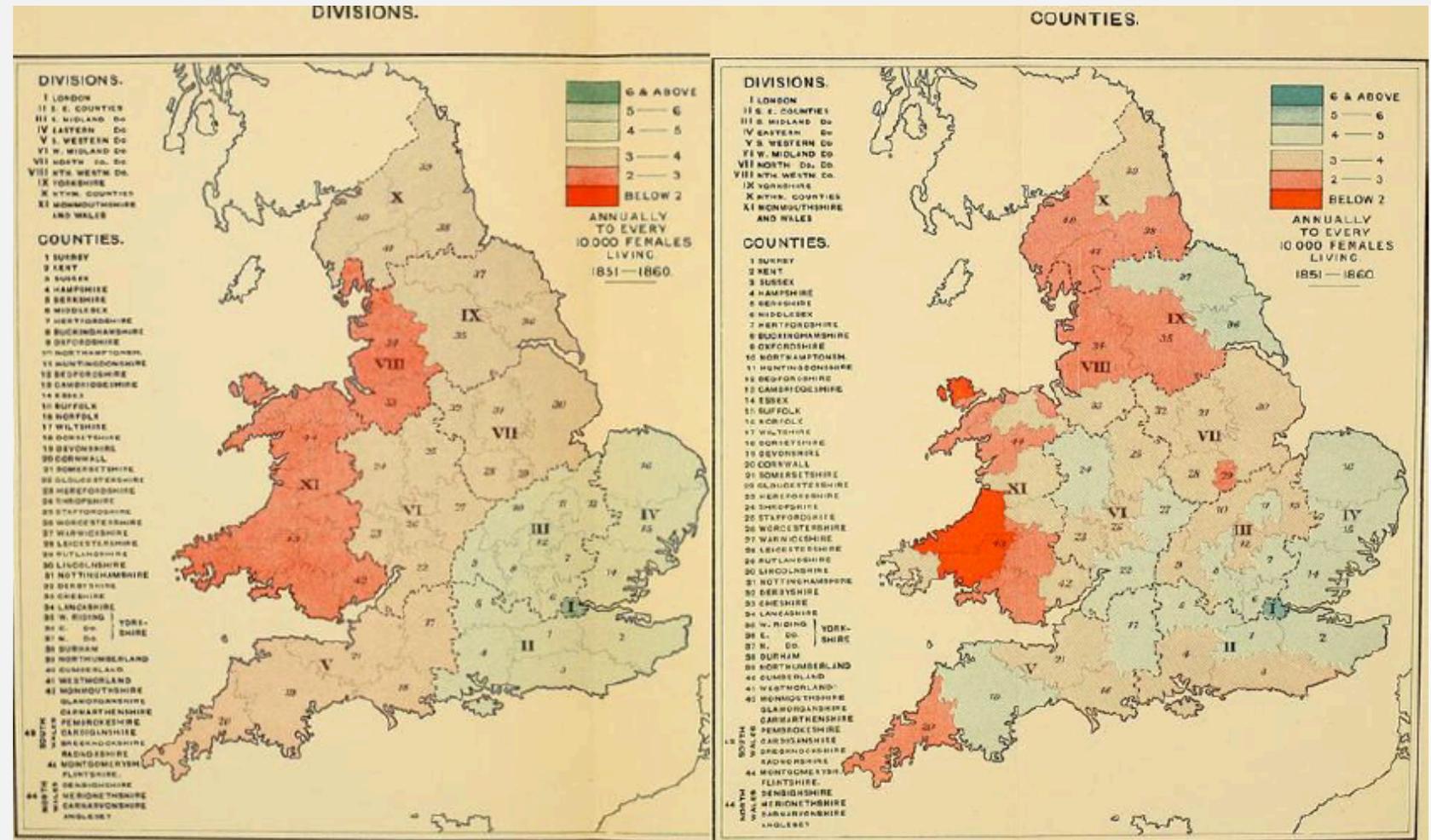
The geographical distribution of cancer in females in England and Wales, 1851-60. From Haviland (1878).

Cáncer de pulmón en mujeres,
1921-1930
(Stocks, 1939)



1. EPIDEMIOLOGÍA Y EPIDEMIOLOGÍA ESPACIAL

Cáncer en mujeres, 1851-1860
(Haviland, 1878)



1. EPIDEMIOLOGÍA Y EPIDEMIOLOGÍA ESPACIAL

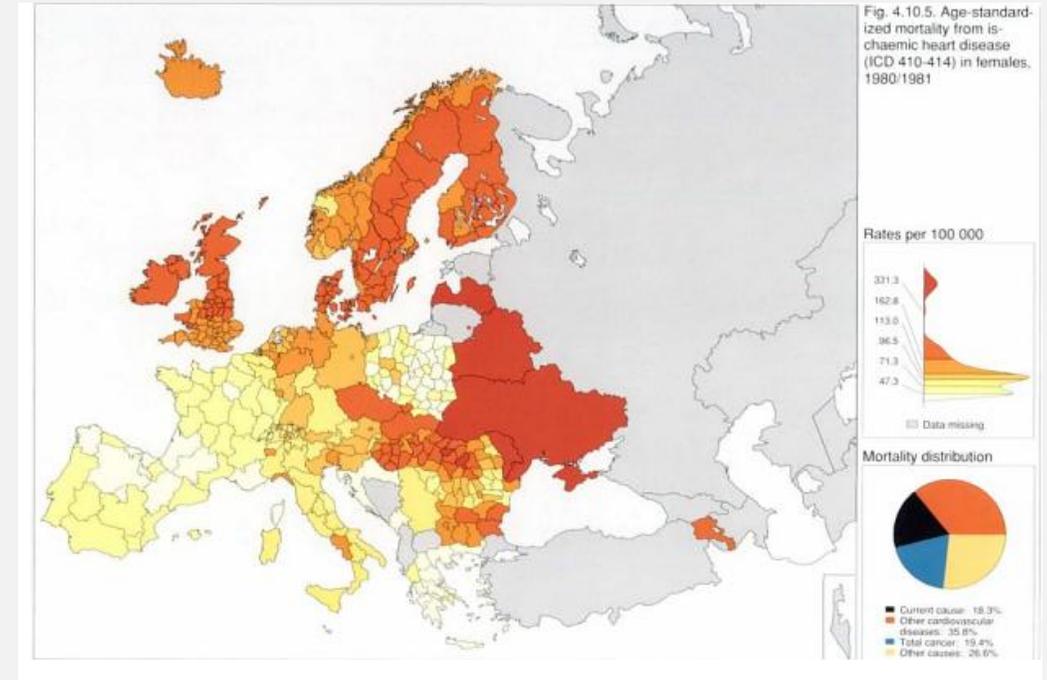
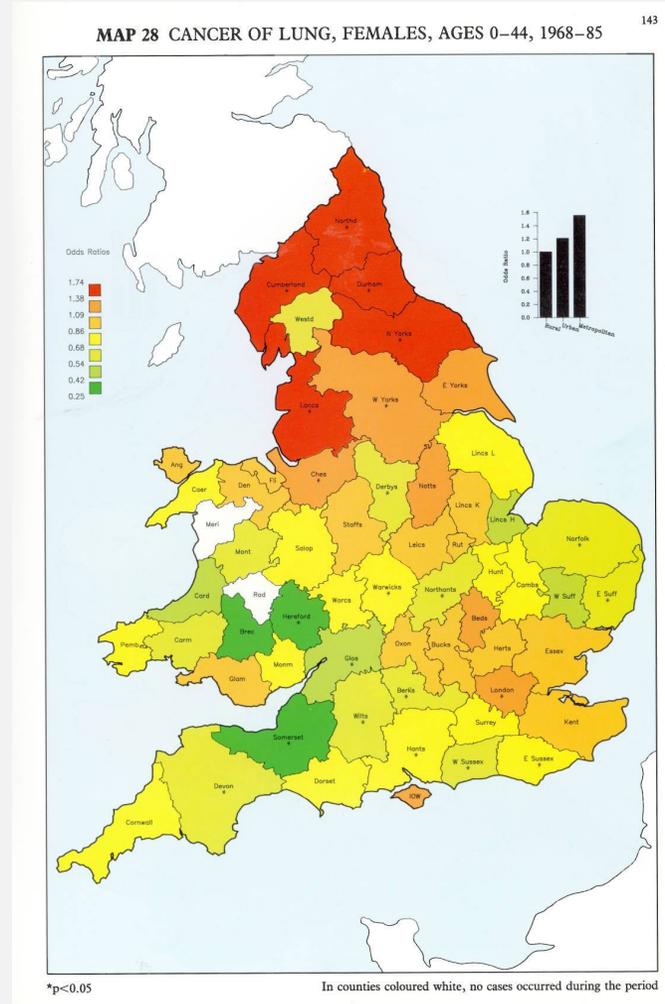
Evolución de los mapas en epidemiología espacial

- Atlas de enfermedades, a nivel nacional e internacional
 - Atlas de incidencia del cáncer en Inglaterra y Gales 1968-85 (Swerdlow y dos Santos Silva, 1993)
 - Atlas de mortalidad en Europa 1980-81 y 1990-91 (OMS, 1997)

1. EPIDEMIOLOGÍA Y EPIDEMIOLOGÍA ESPACIAL

Incidencia de cáncer de pulmón en mujeres 1968-1985

(Swerdlow and dos Santos Silva, 1993)



Mortalidad por Infarto de miocardio estandarizada por edad, 1980-81

(OMS)

1. EPIDEMIOLOGÍA Y EPIDEMIOLOGÍA ESPACIAL

- Como podemos ver, la epidemiología espacial partió de la representación de la distribución espacial de eventos en salud con el doble objetivo de caracterizar su extensión y establecer hipótesis de sus posibles causas.
- Posteriormente, la epidemiología espacial creció mucho en complejidad (métodos de análisis, unidades pequeñas) y utilidad.
- La confluencia entre la epidemiología, la estadística y la informática, junto con el gran desarrollo tecnológico, han permitido que la epidemiología espacial se desarrolle a pasos agigantados.

1. EPIDEMIOLOGÍA Y EPIDEMIOLOGÍA ESPACIAL

Desarrollos a partir de 1996

- Sistemas de Información Geográfica (SIG)
- Software para representar mapas y analizar datos espaciales
- Mayor disponibilidad de datos georeferenciados (GPS, etc.)
- Desarrollo de métodos estadísticos
 - Técnicas sofisticadas para separar la señal del ruido
 - Métodos para controlar la dependencia espacial y temporal
 - Métodos para detectar aglomeraciones de casos (clusters)

1. EPIDEMIOLOGÍA Y EPIDEMIOLOGÍA ESPACIAL

Problemas en epidemiología espacial:

- Diferentes escalas (por ejemplo, comunidades autónomas, provincia, municipio, distrito, barrio, código postal, sección censal, ...)
- Cambios en los límites de algunas de estas unidades
- Pueden existir errores en la georeferenciación (debido a direcciones erróneas o inexistentes, etc.)
- Desalineación (missalignement)

ESTADÍSTICA Y EPIDEMIOLOGÍA ESPACIAL

1. Epidemiología y Epidemiología Espacial
2. **Estadística Espacial**
3. Mapas de Enfermedades
4. Regresión ecológica
5. Agregación espacial (clustering)
6. Identificación de focos (cluster detection)



Epidemiología Espacial

2. ESTADÍSTICA ESPACIAL

- La **estadística espacial** se ocupa de la exploración, descripción, visualización y análisis de los datos, considerando sus características de distribución en el espacio, que suelen expresarse a través del uso de coordenadas geográficas.
- La **estadística espacial** es la rama de la estadística que analiza datos georeferenciados, es decir, datos de los que se dispone de sus coordenadas espaciales (datos espaciales).
- Por **datos espaciales** se entiende aquellas mediciones u observaciones realizadas en localizaciones o en áreas específicas. Además del valor de la medición, incorporan la localización/posición de los valores observados.

2. ESTADÍSTICA ESPACIAL

Características de los datos espaciales:

1. Heterogeneidad espacial

- Las observaciones no son homogéneas en el espacio

2. Dependencia espacial

- Las observaciones de una localización dependen de otras observaciones en otras localizaciones (generalmente próximas).

2. ESTADÍSTICA ESPACIAL

Características de los datos espaciales:

1. Heterogeneidad espacial

- Las observaciones no son homogéneas en el espacio

2. ESTADÍSTICA ESPACIAL

Existen dos tipos de **heterogeneidad espacial**:

➤ **Heterocedasticidad:**

- *Causa estructural*: debido a la utilización de datos procedentes de unidades espaciales arbitrarias
- *Causa muestral*: existencia de outliers, omisión de variables relevantes, errores de medida y otros errores de especificación.

➤ **Cambio-inestabilidad estructural:** los individuos no son homogéneos en el espacio (por ejemplo, norte-sur; centro-periferia, ...).

2. ESTADÍSTICA ESPACIAL

Características de los datos espaciales:

2. Dependencia espacial

- Las observaciones de una localización dependen de otras observaciones en otras localizaciones (generalmente próximas).

2. ESTADÍSTICA ESPACIAL

Existen dos tipos de **dependencia espacial**:

- ***Interacción espacial (dependencia espacial sustantiva)***: es un efecto spillover de un individuo sobre otro individuo.
- ***Autocorrelación en los residuos (dependencia espacial residual)***: causada por errores de medida espacialmente relacionados.

2. ESTADÍSTICA ESPACIAL

Tipos datos espaciales:

Los datos espaciales se han clasificado históricamente en tres grandes grupos (Cressie, 1993):

1. Datos en rejilla o **datos de área** ('lattice data' o 'areal data')
2. Datos referenciados puntualmente: **Procesos puntuales** ('point processes')
3. Datos referenciados puntualmente: **Datos geoestadísticos** ('geoestatistical data')

2. ESTADÍSTICA ESPACIAL

1. Datos en rejilla o datos de área ('lattice data' o 'areal data')

- Los **datos de área** o en rejilla se corresponden con variables aleatorias discretas (recuentos) ("count data").
- En los datos de área ***no se conoce la localización exacta del caso.***
- ***Las localizaciones son áreas*** con límites geográficos bien definidos, habitualmente unidades administrativas (ciudades, barrios, secciones censales, etc.).

2. ESTADÍSTICA ESPACIAL

1. **Datos en rejilla o datos de área ('lattice data' o 'areal data')**
 - La ***variable respuesta es el número agregado de casos en dicha área*** en un periodo de tiempo determinado.
 - Este era el tipo de dato más comunmente utilizado al principio de la estadística espacial.

2. ESTADÍSTICA ESPACIAL

1. **Datos en rejilla o datos de área ('lattice data' o 'areal data')**
 - Observaciones procedentes de un proceso aleatorio sobre un conjunto de regiones espaciales distribuidas espacialmente de forma regular o irregular.
 - Quedan definidos matemáticamente como un conjunto de índices de localizaciones con un conjunto asociado de vecinos.
 - **Vecinos:** áreas vecinas (proximidad, contigüidad, etc.) a una región concreta.

2. ESTADÍSTICA ESPACIAL

1. Datos en rejilla o datos de área ('lattice data' o 'areal data')

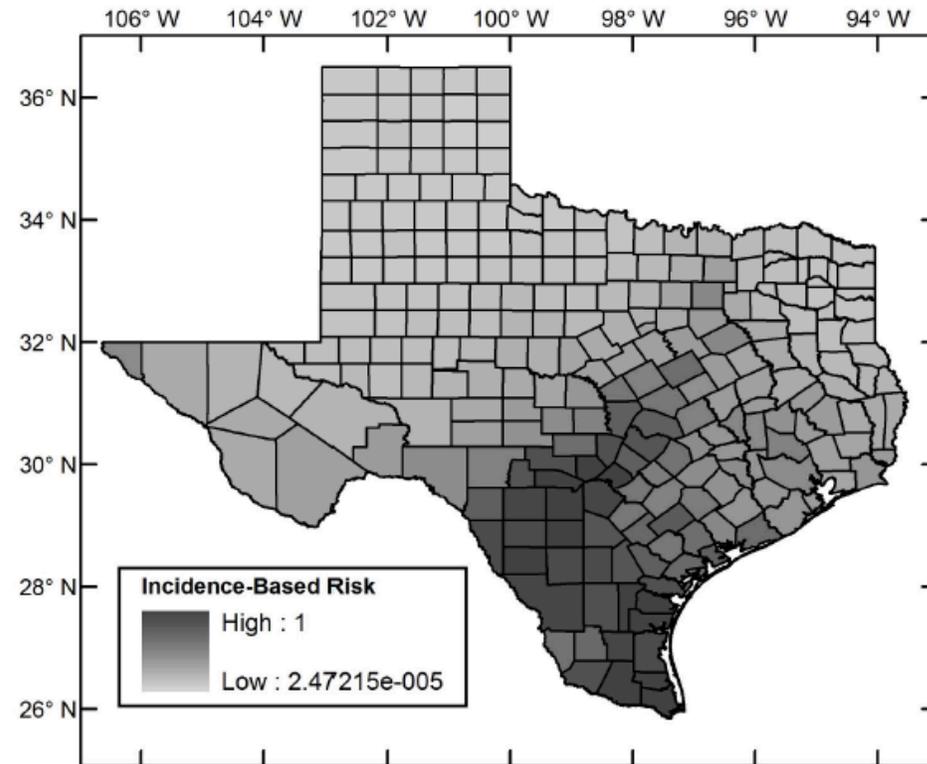


Figure 5. Incidence-based risk map for Chagas Disease in Texas.

2. ESTADÍSTICA ESPACIAL

2. Datos referenciados puntualmente: Procesos puntuales ('point processes')

- Los **procesos puntuales** se corresponden con variables aleatorias Bernouilli.
- En los procesos puntuales ***se conoce la localización exacta del caso y ésta es aleatoria.***
- Las ***localizaciones*** del evento de interés ***se observan en una región determinada.*** Por ejemplo, las coordenadas de los domicilios de los casos de ELA en Cataluña.
- La ***información*** sobre estos datos ***no es pública.*** Se recogen en estudios de casos y controles o en estudios de cohortes.
- Se pueden agregar por unidades espaciales, pudiéndose crear datos de área.

2. ESTADÍSTICA ESPACIAL

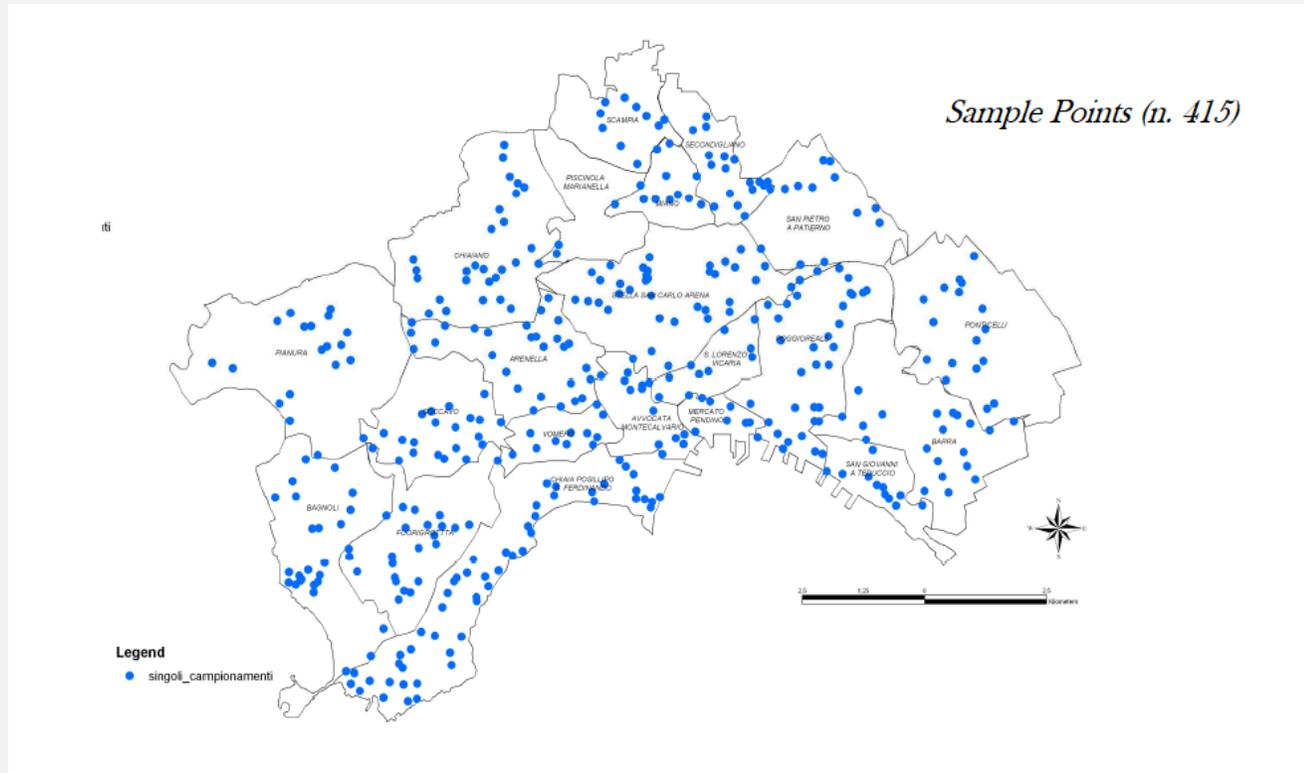
2. Datos referenciados puntualmente: Procesos puntuales ('point processes')

➤ En este caso, nos puede interesar:

- Valorar si los eventos siguen un patrón espacial determinado (agregación, forma regular, etc.)
- Estudiar si un patrón observado se asocia con alguna variable (exposición a alguna variable ambiental, como la contaminación atmosférica; cercanía a focos contaminantes; contexto socioeconómico, etc.)

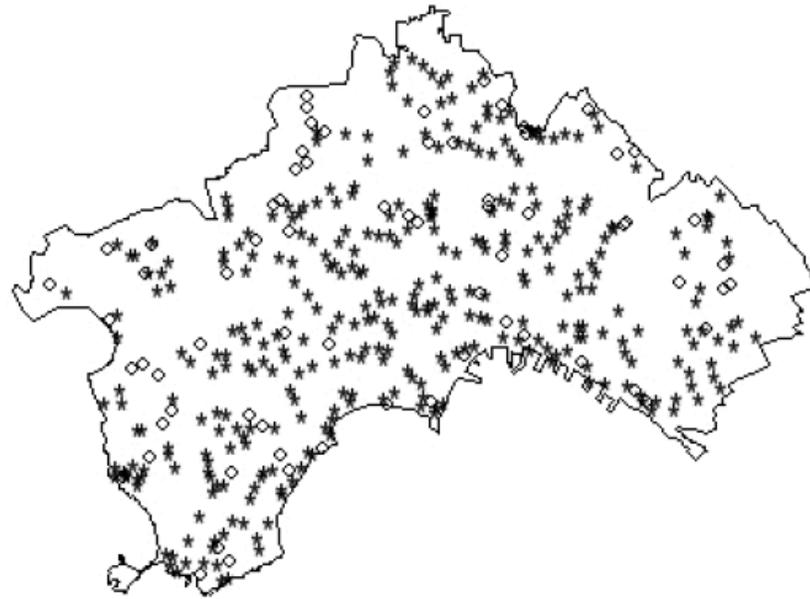
2. ESTADÍSTICA ESPACIAL

2. Datos referenciados puntualmente: Procesos puntuales ('point processes')



2. ESTADÍSTICA ESPACIAL

2. Datos referenciados puntualmente: Procesos puntuales ('point processes')



Distribution of cases (positives: circle) and controls (negatives: stars). Naples, February - May 2005

2. ESTADÍSTICA ESPACIAL

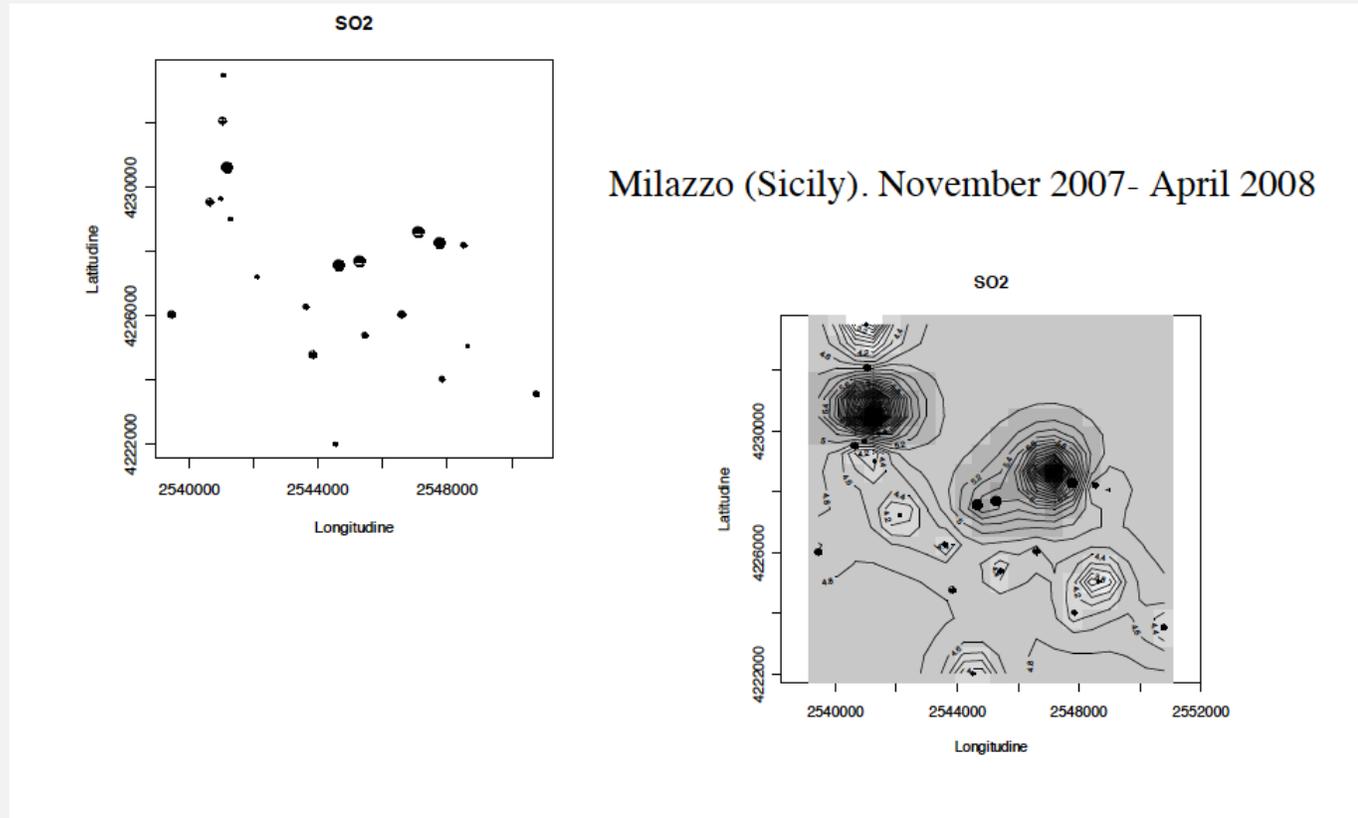
3. **Datos referenciados puntualmente: Datos geoestadísticos ('geo-statistical data')**
 - Los **datos geoestadísticos** se corresponden con variables aleatorias continuas.
 - En los datos geoestadísticos ***se conoce la localización exacta del caso y ésta es fija*** (estaciones captadoras de contaminación atmosférica, vetas de un determinado metal, etc.)
 - ***La variable respuesta se mide en cada localización*** (por ejemplo, medidas de contaminación atmosférica, medidas de composición química del suelo, temperatura, etc.).

2. ESTADÍSTICA ESPACIAL

- 3. Datos referenciados puntualmente: Datos geoestadísticos ('geoestatistical data')**
 - Es decir, disponemos de medidas tomadas en puntos fijos, definidas en cualquier lugar del espacio, por lo que sus ***localizaciones definen una superficie espacialmente continua***.
 - Se suele ***extender la distribución espacial*** de los valores de un atributo ***sobre la región de estudio*** mediante modelos matemáticos (por ejemplo, kriging).

2. ESTADÍSTICA ESPACIAL

3. Datos referenciados puntualmente: Datos geoestadísticos ('geoestatistical data')



2. ESTADÍSTICA ESPACIAL

Pero, los tipos de diseño espacial, ¿qué son en realidad? ¿procesos? ¿modelos? ¿métodos?

- En 2012, Diggle propone un cambio de paradigma y re-define la estadística espacial como ***'un conjunto de modelos y métodos estadísticos que pretenden ayudar a los científicos a entender fenómenos espaciales, que no pueden ser observados directamente, sino de forma indirecta, con información incompleta'***, en forma de datos en rejilla, procesos puntuales y datos geoestadísticos. El modelo estadístico que propuso, como modelo único base, es el **modelo log Cox**.
- Es decir, unifica la estadística espacial, tal y como McCullagh y Nelder (1989) unificaron los modelos lineales generalizados.

ESTADÍSTICA Y EPIDEMIOLOGÍA ESPACIAL

1. Epidemiología y Epidemiología Espacial
2. Estadística Espacial
- 3. Mapas de Enfermedades**
- 4. Regresión ecológica**
- 5. Agregación espacial (clustering)**
- 6. Identificación de focos (cluster detection)**



Epidemiología Espacial

1. EPIDEMIOLOGÍA Y EPIDEMIOLOGÍA ESPACIAL

Actualmente, existe una gran necesidad de **métodos espaciales** ya que:

- la mayoría de estudios epidemiológicos tienen un **componente espacial**
- estamos interesados en tener en cuenta el **componente espacial**
- muy habitualmente, el **área de estudio es pequeña** y/o existe gran cantidad de **información a nivel individual con información de la localización** geográfica

1. EPIDEMIOLOGÍA Y EPIDEMIOLOGÍA ESPACIAL

¿Por qué estamos interesados en tener en cuenta el componente espacial?

- Porque nos interesa explícitamente el patrón espacial del factor de riesgo: **mapas de enfermedades**
- Porque el componente espacial recoge gran parte de la confusión no observada : **regresión espacial**
- Porque observamos aglomeraciones en el espacio: **detección de agrupaciones**
- Porque nos interesan los efectos de un foco contaminante en la salud de los habitantes residentes en sus alrededores: **identificación de focos**

ESTADÍSTICA Y EPIDEMIOLOGÍA ESPACIAL

1. Epidemiología y Epidemiología Espacial
2. Estadística Espacial
- 3. Mapas de Enfermedades**
4. Regresión ecológica
5. Agregación espacial (clustering)
6. Identificación de focos (cluster detection)



Epidemiología Espacial

3. MAPAS DE ENFERMEDADES

Los **mapas de enfermedades** proporcionan una primera visión de la distribución espacial de la enfermedad, del evento de salud o bien de los factores de riesgo de los mismos.

Son un resumen visual del riesgo geográfico

3. MAPAS DE ENFERMEDADES

Los **mapas de enfermedades** se utilizan para:

- **propósitos descriptivos:** para resumir la variación espacial y espacio-temporal del riesgo de la enfermedad.
- **generar hipótesis etiológicas:** los mapas de exposición permiten un examen informal (examen formal mediante regresión espacial)
- **vigilancia:** para resaltar áreas aparentemente de alto riesgo
- ayudar a la formulación de políticas y la asignación de recursos.

3. MAPAS DE ENFERMEDADES

La representación de la distribución espacial de la enfermedad en un mapa se puede llevar a cabo a diferentes escalas:

- a escala ***internacional***: comparaciones entre países (OMS)
- a escala ***nacional***: comparaciones entre comunidades autónomas, regiones, ABS, ...
- a escala ***local***: estudios en áreas pequeñas

3. MAPAS DE ENFERMEDADES

La representación de la distribución espacial de la enfermedad en un mapa se puede llevar a cabo a diferentes escalas:

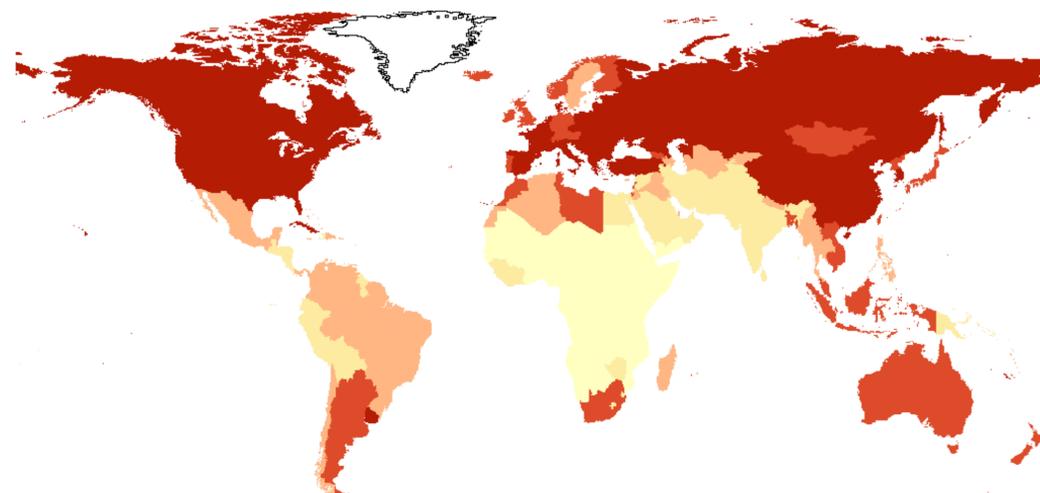
- a escala ***internacional***: comparaciones entre países (OMS)
 - grandes diferencias internacionales en las tasas de mortalidad por cáncer de pulmón, potencialmente explicadas por las diferencias en las prevalencias del tabaquismo
 - altas tasas de cancer de hígado en África y el sudeste asiático, relacionadas con la hepatitis B

3. MAPAS DE ENFERMEDADES

- a escala **internacional**: comparaciones entre países (OMS)

International scale: Lung cancer rates - worldwide, 2008

International Agency for Research on Cancer
Estimated age-standardised mortality rate per 100,000
Lung: male, all ages



3. MAPAS DE ENFERMEDADES

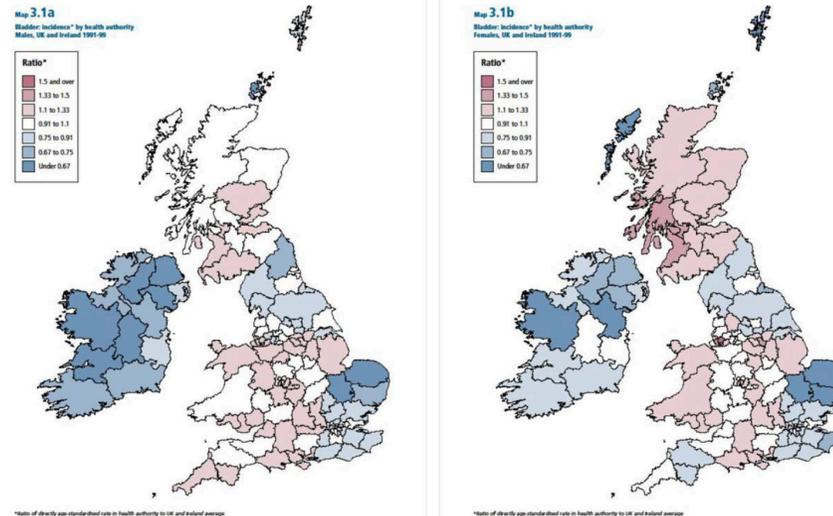
La representación de la distribución espacial de la enfermedad en un mapa se puede llevar a cabo a diferentes escalas:

- a escala **nacional**: comparaciones entre comunidades autónomas, regiones, ABS, ...
 - la mayoría de los atlas de enfermedades publicados se encuentran dentro de esta categoría

3. MAPAS DE ENFERMEDADES

- a escala **nacional**: comparaciones entre comunidades autónomas, regiones, ABS, ...

National scale: Bladder cancer incidence in the UK and Ireland - Cancer Atlas of the United Kingdom and Ireland, 1991 - 2000



3. MAPAS DE ENFERMEDADES

La representación de la distribución espacial de la enfermedad en un mapa se puede llevar a cabo a diferentes escalas:

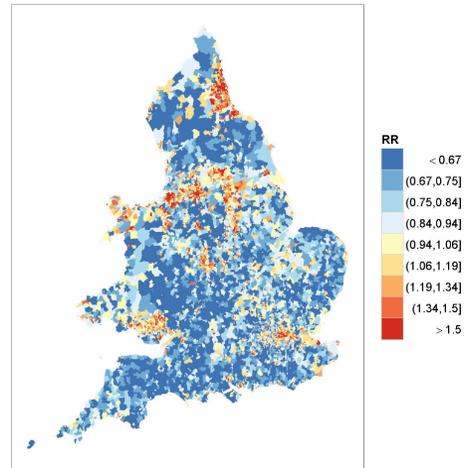
- a escala **local**: estudios en áreas pequeñas
 - escala subnacional, por ejemplo, municipios, barrios, secciones censales, ...
 - cada vez más común a medida que mejoran los datos y los métodos

3. MAPAS DE ENFERMEDADES

- a escala **local**: estudios en áreas pequeñas

Lung cancer incidence in males, 1985-2009, England and Wales

SIRs at ward level



Is the variability real or simply reflecting unequal E_i s?

Have the highlighted areas truly a raised relative risk?

	Min	Q1	Median	Q3	Max
O	0	26	47	84	456
E	3.25	32.14	53.60	82.47	390.49
SMR	0	0.70	0.89	1.13	2.63

3. MAPAS DE ENFERMEDADES

En la actualidad, estos mapas no tienen sentido si no se refieren a áreas pequeñas.



DISEÑO Y ANÁLISIS DE ESTUDIOS EPIDEMIOLÓGICOS EN ÁREAS PEQUEÑAS

3. MAPAS DE ENFERMEDADES

¿Por qué nos interesa representar gráficamente en mapas estas medidas de ocurrencia de la enfermedad en áreas pequeñas?

- Porque existe un interés en representar en un mapa las variaciones geográficas en los resultados de salud a pequeña escala
 - Destacar las fuentes de heterogeneidad y patrones espaciales
 - Sugerir determinantes de salud pública o pistas etiológicas
- Porque a pequeña escala (a nivel de municipio, barrio, sección censal, ...)
 - Estas medidas son menos susceptible al sesgo ecológico (agregación)
 - Existe una mayor capacidad de detectar efectos altamente localizados

3. MAPAS DE ENFERMEDADES

Pero, ¿qué indicadores de ocurrencia de la enfermedad deberíamos representar en estos mapas?

Algunas de las medidas de ocurrencia de la enfermedad más conocidas son:

- números absolutos (recuentos)
- prevalencia
- incidencia
- mortalidad

3. MAPAS DE ENFERMEDADES

➤ Morbilidad

- ***prevalencia***
 - ✓ registros
 - ✓ admisiones hospitalarias (problemas calidad datos por cuestiones administrativas)
- ***incidencia***
 - ✓ datos habitualmente disponibles solo para los diferentes tipos de cáncer (registros)
 - ✓ pueden ser más sensibles a los efectos de la exposición
 - ✓ menor lapso de tiempo entre la exposición y el evento en comparación con la mortalidad

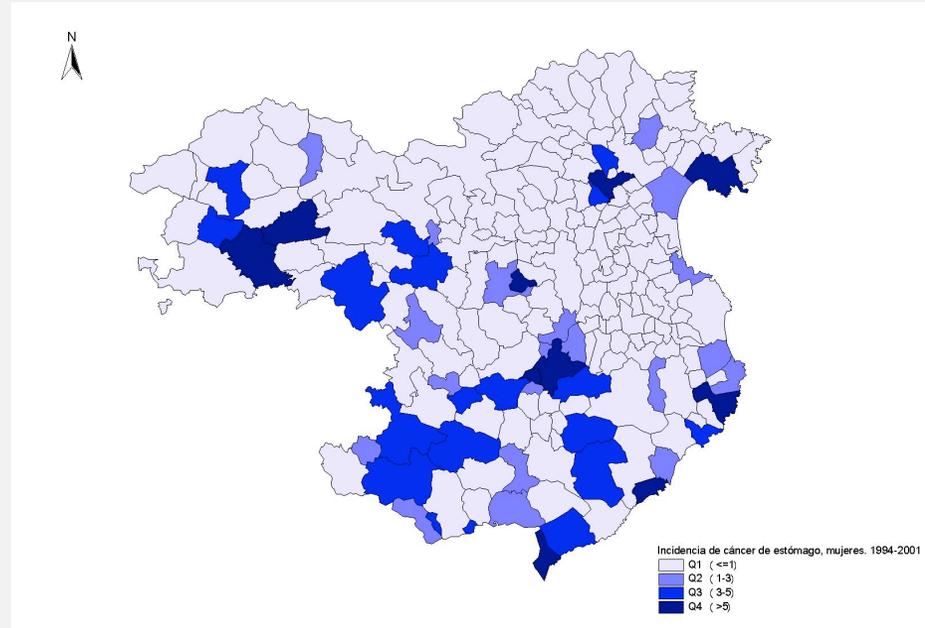
3. MAPAS DE ENFERMEDADES

➤ Mortalidad

- es la fuente de datos más fácilmente disponible para todas las enfermedades
- debería ser completa y relativamente precisa

3. MAPAS DE ENFERMEDADES

➤ Casos observados vs. tasas



Incidence of stomach cancer in women, 1994-2001. **Número de casos**

3. MAPAS DE ENFERMEDADES

➤ Tasas crudas vs. tasas estandarizadas

Supongamos dos municipios de 10.000 habitantes, A y B.

En A se observaron 125 casos de cáncer y en B, 182.

Tasas crudas: 12,5 (por 1.000 habitantes) en **A** y **18,2** (por 1.000 habitantes) en **B**.



¿Existe un mayor riesgo de morir de cáncer en el municipio B?

¿Es la composición por edad en los dos municipios la misma?

3. MAPAS DE ENFERMEDADES

➤ Tasas crudas vs. tasas estandarizadas

Antes de contestar, calcularemos las **tasas estandarizadas** mediante el método indirecto. Para ello nos fijaremos en la **composición por edad** en cada uno de los municipios.

Grupos edad	Población		Composición poblacional (%)		Número de casos		Tasa específica (por 10 ³)	
	A	B	A	B	A	B	A	B
0-4	1.500	2.500	15	25	63	90	42	36
5-14	2.500	3.500	25	35	50	84	20	24
15 o más	6.000	4.000	60	40	12	8	2	2
Total	10.000	10.000	100%		125	182	12.5	18.2

La composición por edad es diferente en A (edad más avanzada) y en B (más jóvenes).

3. MAPAS DE ENFERMEDADES

➤ **Tasas crudas vs. tasas estandarizadas**

Ahora estandarizaremos por edad mediante el método indirecto.

(1) Grupos edad	(2) Población estándar	(3) Tasas de incidencia
	A+B	
0-4	4.000	$(63+90)/4.000=0.03811$
5-14	6.000	$(50+84)/6.000=0.02233$
15 o más	10.000	$(12+8)/10.000=0.02000$
Total	20.000	

Grupos edad	Población		Tasas de incidencia		Casos esperado	
	A	B			A	B
0-4	1.500	2.500	0.03811	0.03811×1500 0.03811×2500	51,175	95,275
5-14	2.500	3.500	0.02233		55,825	78,725
15 o más	6.000	4.000	0.02000		12,000	8,000
Total	10.000	10.000	100%	Esperados totales	125,2	182,0
Observados totales	125	182				
Estandarizados	$125/125=1.000$	$182/182=1.000$				

3. MAPAS DE ENFERMEDADES

➤ Tasas crudas vs. tasas estandarizadas

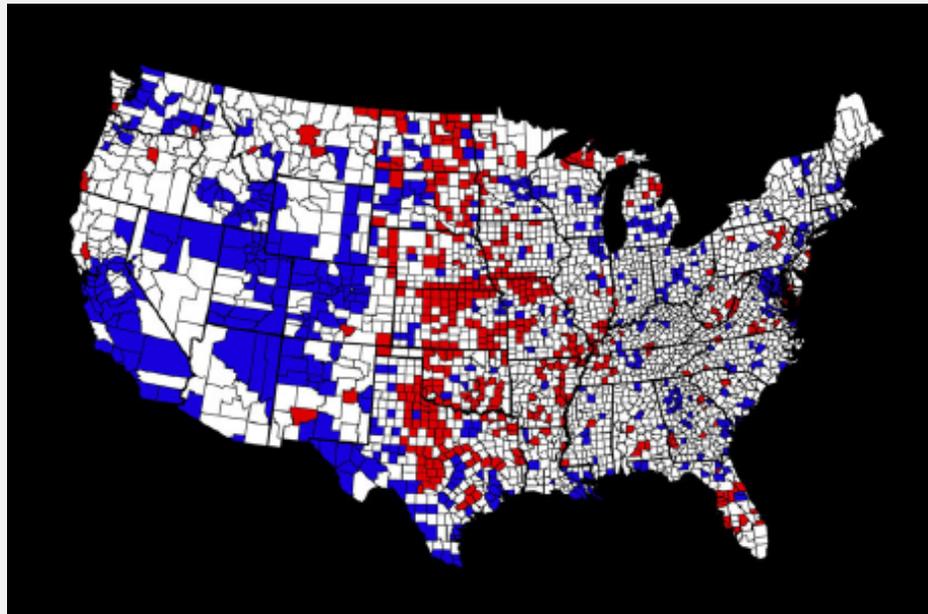


Figure 2
Unadjusted Mortality Rates 1993–1997 Red = High Mortality White = Normal Mortality Blue = Low Mortality

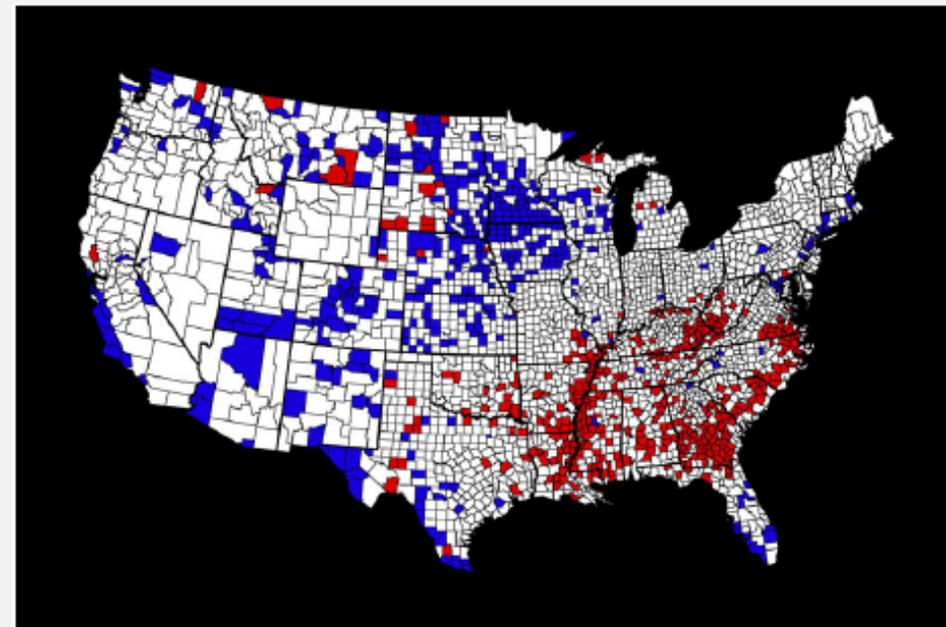


Figure 3
Age Adjusted Mortality Rates 1993–1997 Red = High Mortality White = Normal Mortality Blue = Low Mortality

3. MAPAS DE ENFERMEDADES

➤ Tasas crudas vs. tasas estandarizadas



Figura 3. Tasas crudas por muertes súbitas cardíacas. República mexicana, 2010.



Figura 4. Tasas estandarizadas por muertes súbitas cardíacas. República mexicana, 2010.

3. MAPAS DE ENFERMEDADES

Así pues, **se deben estandarizar o ajustar las tasas.**

Hemos visto un tipo de estandarización (estandarización indirecta), pero existen **dos métodos de estandarización:**

- **Estandarización directa:** los casos esperados se calculan a partir de la estructura de edad y sexo de la población de referencia.
- **Estandarización indirecta:** los casos esperados se calculan a partir de las tasas específicas de enfermedad de la población de referencia.

3. MAPAS DE ENFERMEDADES

- **Datos:** habitualmente disponemos de datos para una región de interés/área de referencia, a nivel geográfico y para un periodo de tiempo determinado. Por ejemplo, Cataluña, nivel de sección censal, periodo 2011-2020
 - O_i : número de casos observados en el área i
 - E_i : número de casos esperados en el área i , basados en el tamaño de la población, ajustados por edad, sexo, otros estratos, ...
 - n_i : población a riesgo en el área i
- **Parámetro de interés:** Riesgo relativo λ_i en cada área comparado con el área de referencia escogida

3. MAPAS DE ENFERMEDADES

- Calcularemos el número de casos esperados si la población tiene la misma tasa de mortalidad/incidencia estrato específica que en el área de referencia
- Ajustaremos (estratificaremos) por: edad, sexo, ...
- Estandarización indirecta:

$$E_i = \sum_j n_{ij} r_j$$

donde r_{ij} es la tasa de enfermedad para el estrato j en la población de referencia y n_{ij} es la población a riesgo en el área i , estrato j (if internal comparison: $\sum_{i=1}^N O_i = \sum_{i=1}^N E_i$)

3. MAPAS DE ENFERMEDADES

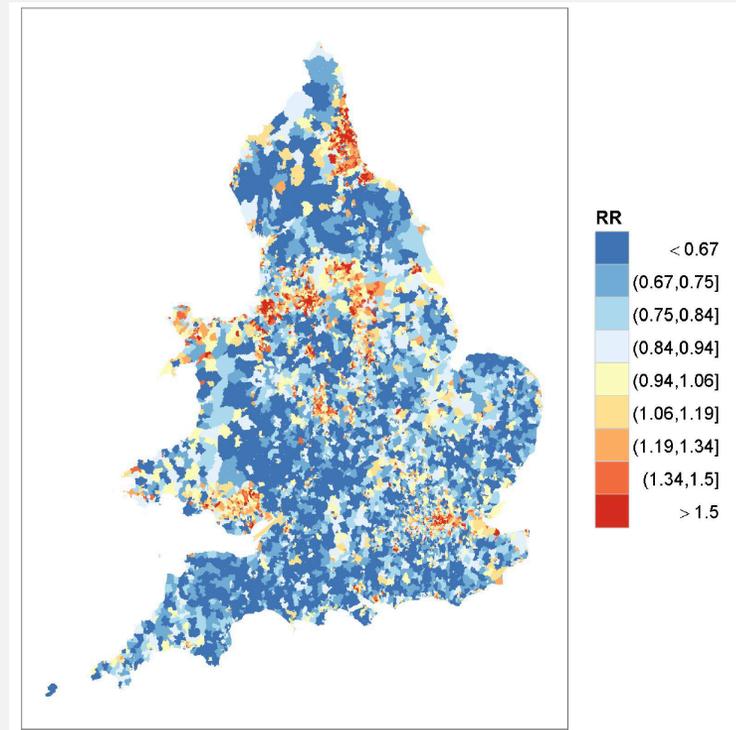
- Otro ejemplo: Incidencia de cáncer de pulmón en hombres, para todas las edades, usando las tasas en Inglaterra y Gales como referencia, para el período 1985-2009

Strata Age group	Reference area=EW			Ward A		
	Population n_j	Observed O_j	Age-specific rate per 100,000 males $r_j = \frac{O_j}{n_j}$	Population n_{ij}	Observed O_{ij}	Expected $E_{ij} = \frac{n_{ij} * r_j}{100000}$
0-4	41,400,692	15	0.04	11,438	0	0.00
5-9	41,143,722	6	0.01	9,697	0	0.00
10-14	41,469,696	9	0.02	9,026	0	0.00
15-19	43,087,823	39	0.09	8,650	0	0.01
20-24	45,441,353	79	0.17	12,409	0	0.02
25-29	46,873,725	172	0.37	16,963	0	0.06
30-34	46,927,658	518	1.10	17,303	0	0.19
35-39	46,936,367	1,465	3.12	13,847	0	0.43
40-44	45,304,711	4,136	9.13	11,843	1	1.08
45-49	41,657,557	9,835	23.61	9,457	5	2.23
50-54	38,451,416	20,929	54.43	8,561	3	4.66
55-59	35,842,426	40,427	112.79	7,613	8	8.59
60-64	32,480,032	68,230	210.07	6,968	5	14.64
65-69	28,231,499	95,794	339.32	6,290	15	21.34
70-74	23,315,240	110,371	473.39	5,098	27	24.13
75-79	17,297,264	102,038	589.91	4,049	22	23.89
80-84	10,498,214	68,273	650.33	2,616	20	17.01
85+	6,289,452	38,748	616.08	1,312	12	8.08
TOTAL	632,648,846	561,084		163,140	118	126.38

$$SIR_A = \frac{118}{126.38} = 0.93 \rightarrow$$
 Tenemos menos casos incidentes de cancer de pulmón en hombres en el área de salud A que los que se esperaría (un 7% menos) en esta área después de ajustar por grupos de edad

3. MAPAS DE ENFERMEDADES

- Razón de incidencia estandarizada (RIE=O/E) de cáncer de pulmón en hombres en Inglaterra y Gales (1985-2009) a nivel de barrios



3. MAPAS DE ENFERMEDADES

- La práctica común consiste en **representar** en un mapa las **RME (o RIE)**:
 - **en el caso de enfermedades raras y/o áreas pequeñas, las RME (o RIE) son muy imprecisas** porque la varianza es proporcional al cuadrado del denominador (es decir, es proporcional al número de casos esperados al cuadrado)
 - la varianza asociada a las áreas que tengan casos esperados pequeños será muy alta
- Las **RME (o RIE) se estiman independientemente** en cada área
- Las RME (o RIE) no hacen uso de las estimaciones del riesgo en otras áreas del mapa, aunque es probable que sean similares. Es decir, **no tienen en cuenta la muy probable dependencia espacial**

3. MAPAS DE ENFERMEDADES

- Las tasas estandarizadas (**RME** o **RIE**) son muy inestables

Observado = 100	Observado = 2
Esperado = 80	Esperado = 1,6
RIE = 1,25	RIE = 1,25
Observado = 101	Observado = 3
Esperado = 80	Esperado = 1,6
RIE = 1,26 (1% más)	RIE = 1,88 (50% más)

3. MAPAS DE ENFERMEDADES

- **Parámetro de interés:** Riesgo relativo λ_i en cada área comparado con el área de referencia escogida.
- En lugar de calcular el parámetro de interés (como hemos hecho anteriormente), lo que se suele hacer es **estimar** este mediante una regresión.
- **Para estimar este parámetro**, lo más habitual es utilizar una **regresión de Poisson** donde la variable respuesta es el numerador del riesgo relativo, es decir, los casos observados (variable discreta o de recuento) y los casos esperados se introducen como offset.

3. MAPAS DE ENFERMEDADES

- **Modelo estadístico** estándar cuando tenemos **enfermedades raras y/o áreas pequeñas**:

$$O_i \sim \text{Poisson}(\lambda_i E_i)$$

donde E_i es el número de casos esperados en el área i

- **Parámetro de interés**: el riesgo relativo λ_i que se estima habitualmente mediante razones de mortalidad estandarizadas (RME) y/o razones de incidencia estandarizadas (RIE):

$$\hat{\lambda}_i = SMR_i \circ SIR_i = \frac{O_i}{E_i} \quad \text{y} \quad \text{Var}(\hat{\lambda}_i) = \frac{\lambda_i}{E_i} \rightarrow \widehat{\text{Var}}(\hat{\lambda}_i) = \frac{O_i}{E_i^2}$$

Recordemos: $X \sim \text{Poisson}(\mu) \leftrightarrow E(X) = \text{Var}(X) = \mu$

3. MAPAS DE ENFERMEDADES

- En el caso de **enfermedades más comunes** suele utilizarse el **modelo Binomial** :

$$O_i \sim \text{Binomial}(p_i, N_i)$$

donde

N_i = población a riesgo

p_i = probabilidad de enfermedad

- $\text{logit}(p_i) = \alpha + V_i$
- **Parámetro de interés:** odds ratio = $OR_i = \exp(\alpha + V_i)$

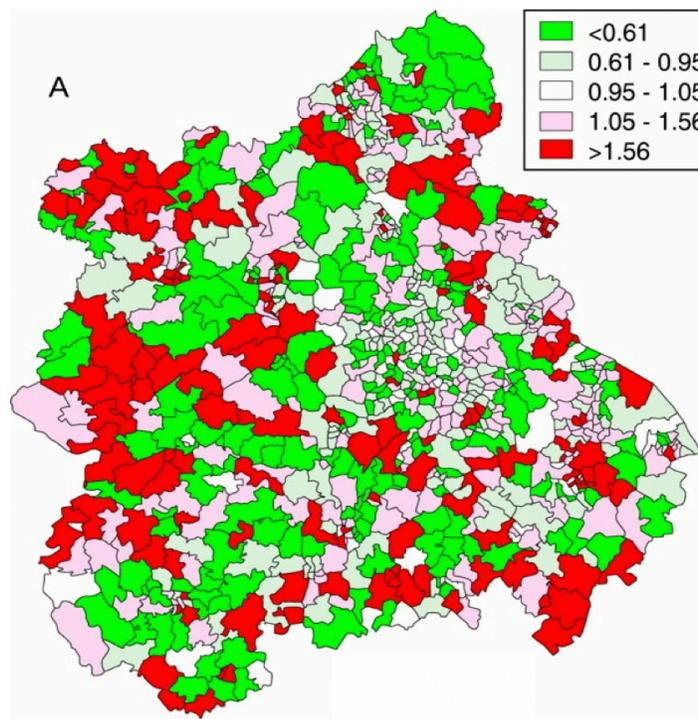
3. MAPAS DE ENFERMEDADES

- A menudo, estamos interesados en proporcionar una **medida resumen** general de la **variabilidad entre áreas** para:
 - comparar la variabilidad de diferentes resultados
 - cuantificar cuanta variación puede venir explicada por las covariables
- Razón percentil: cociente de resultados (riesgos relativos) en áreas clasificadas en el percentil q^{th} y en el percentil $(100-q)^{\text{th}}$ = $QR_{90} = \frac{\lambda_{95\%}}{\lambda_{5\%}}$
- Distribución a posteriori de QR_{90} es fácil de calcular a partir del resultado de MCMC

3. MAPAS DE ENFERMEDADES

Map of SMR of adult leukaemia in West Midlands Region, England 1974-86

(Olsen, Martuzzi and Elliott, *BMJ* 1996;313:863-866).

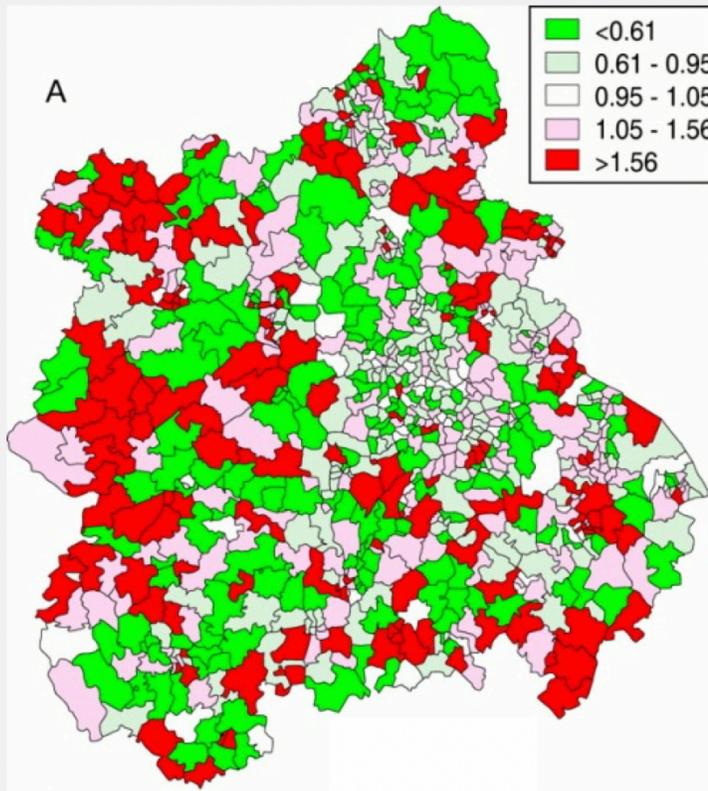


Is the variability real or simply reflecting unequal expected counts ?

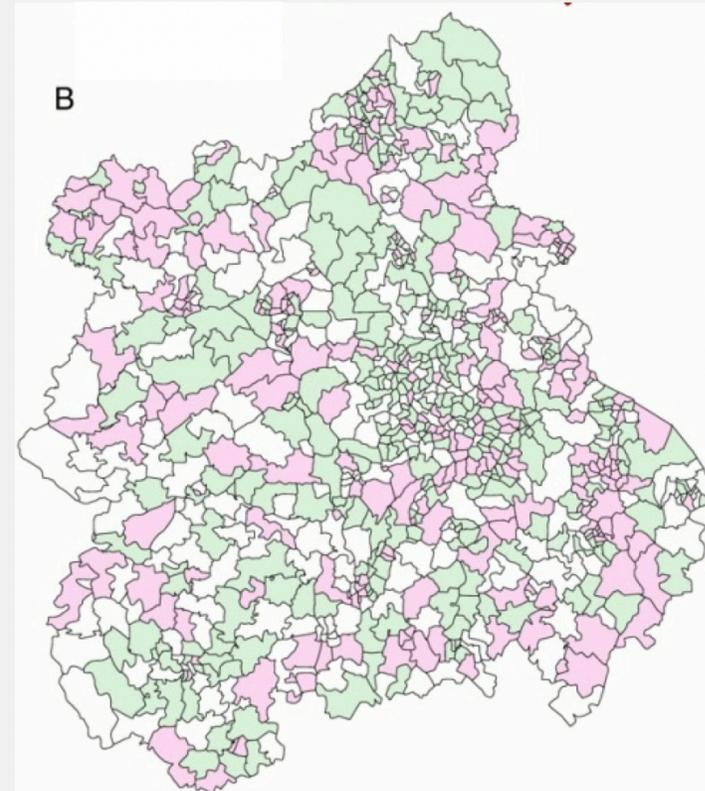
Have the red highlighted areas truly got a raised relative risk?

3. MAPAS DE ENFERMEDADES

(A) unsmoothed SMR



(B) smoothed by Bayesian methods

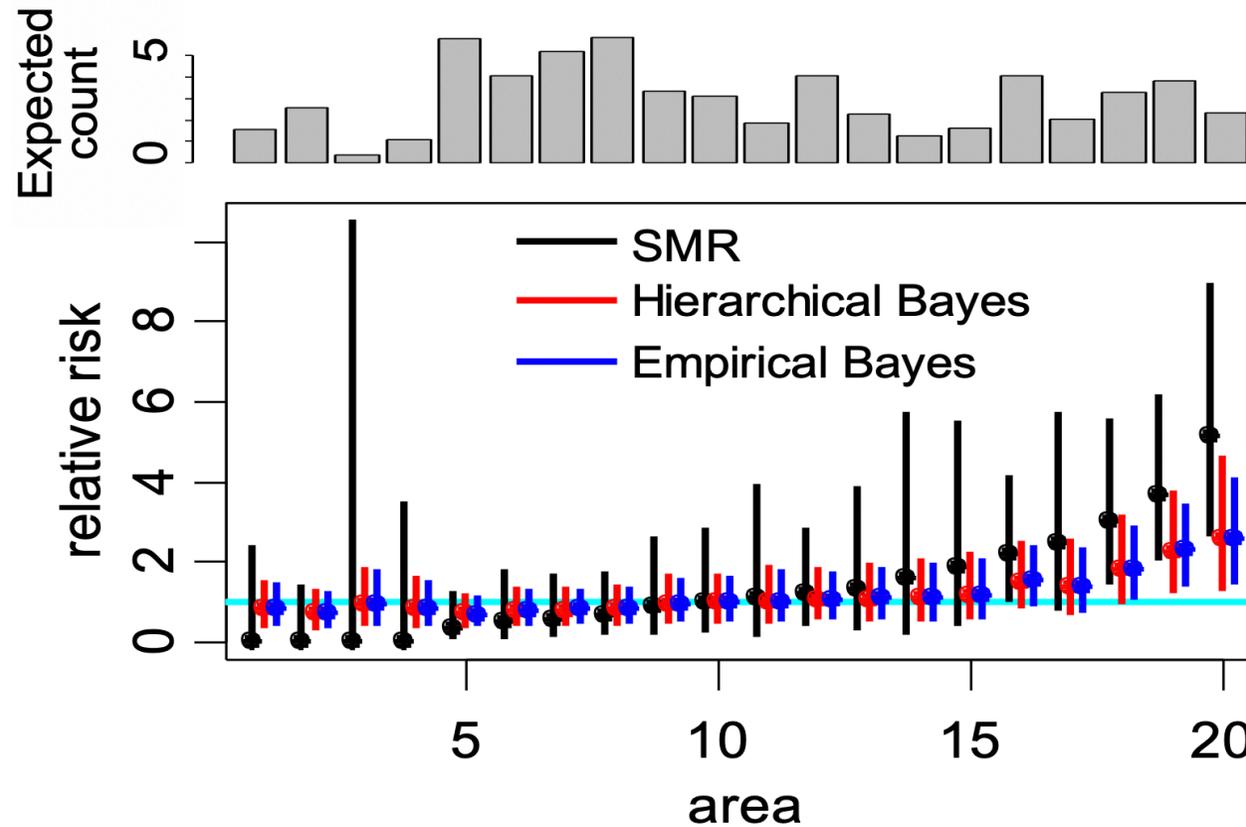


Map of SMR of adult leukaemia in West Midlands Region, England 1974-1986

(Olsen, Martuzzi and Elliott, *BMJ* 1996;313:863-866)

3. MAPAS DE ENFERMEDADES

Comparison of estimation methods



3. MAPAS DE ENFERMEDADES

- Resumiendo, los **problemas** que nos encontramos al estimar razones estandarizadas (tanto de mortalidad como de incidencia) son la **inestabilidad de los estimadores** y los **efectos de no tener en cuenta la muy probable dependencia espacial**.
- Estos problemas pueden solucionarse en parte mediante el **suavizado espacial de las razones**.
- Existen varios métodos para suavizar estas razones.

3. MAPAS DE ENFERMEDADES

Métodos para suavizar mapas

- Algoritmos de suavizado local, ad hoc, por ejemplo, medias móviles espaciales, algoritmo de headbanging
- Análisis de superficies de tendencia, por ejemplo, kriging, suavizado mediante polinomios/splines
- Modelos de efectos aleatorios

3. MAPAS DE ENFERMEDADES

Métodos para suavizar mapas

- **Algoritmos de suavizado local**, ad hoc, por ejemplo, medias móviles espaciales, algoritmo de headbanging
 - Ventajas: rápidos y sencillos de implementar
 - Inconvenientes: pueden ser muy sensibles a la elección ad hoc de ponderaciones, etc. y no existen estimaciones para la incertidumbre (errores estándar)

3. MAPAS DE ENFERMEDADES

Métodos para suavizar mapas

- **Análisis de superficies de tendencia**, por ejemplo, kriging, suavizado mediante polinomios/splines
 - Ventajas: la estimación de los “parámetros de suavizado” está basada en el equilibrio entre el ajuste y la suavización y los errores estándar están generalmente disponibles
 - Inconvenientes: puede ser sensibles a la elección del parámetro de penalización para el equilibrio

3. MAPAS DE ENFERMEDADES

Métodos para suavizar mapas

- **Modelos de efectos aleatorios:** las alternativas más populares a los métodos vistos anteriormente consisten en los modelos de efectos aleatorios
- Dos tipos de modelos aleatorios:
 - **Empirical Bayes:** incluyen como efecto aleatorio únicamente la heterogeneidad espacial
 - **Modelos jerárquicos Bayesianos (modelos mixtos):** incluyen dos efectos aleatorios, la heterogeneidad espacial y la dependencia espacial

3. MAPAS DE ENFERMEDADES

Métodos para suavizar mapas

- *Modelos de efectos aleatorios*
- Antes del gran desarrollo informático que se produjo a partir del año 2000, se utilizaban básicamente los modelos empíricos Bayes, ya que los otros eran computacionalmente complejos. **Actualmente**, estos han quedado obsoletos y se utilizan los **modelos jerárquicos Bayesianos**

3. MAPAS DE ENFERMEDADES

➤ *Modelos de efectos aleatorios*

Con el objetivo de solucionar estos problemas, actualmente se utilizarán estimadores suavizados bayesianos de las RME y/o RIE obtenidos mediante :

- ***Modelo de Poisson-lognormal-espacial*** (modelo jerárquico Bayesiano con dos efectos aleatorios)

3. MAPAS DE ENFERMEDADES

- **Modelos de efectos aleatorios: Modelo de Poisson-lognormal-espacial**
 - ***El efecto aleatorio utilizado para recoger la heterogeneidad espacial*** es el mismo en todos los modelos (iid normal).
 - ***El efecto aleatorio para recoger la dependencia espacial*** se puede aproximar mediante un CAR (Conditional Autoregressive) (modelo denominado ***Besag-York-Mollié***, utilizado cuando se tienen datos de área) o mediante una Mátern (modelo denominado ***log-Cox-Gaussian***, utilizado para todo tipo de datos espaciales).

3. MAPAS DE ENFERMEDADES

➤ Modelos de efectos aleatorios: Modelo de Poisson-lognormal-espacial

Poisson-logNormal model

$$O_i \sim \text{Poisson}(\lambda_i E_i)$$

$$\log \lambda_i = \alpha + V_i$$

$$V_i \sim \text{Normal}(0, \sigma_v^2)$$

Priors (vague, non informative):

- between-area variance σ_v^2 :
 $\sigma_v^2 \sim \text{Inverse Gamma}(0.5, 0.0005) \Leftrightarrow \tau_v \sim \text{Gamma}(0.5, 0.0005)$
 $\sigma_v \sim \text{Truncated Normal}(0, 100)_{[0, \text{Inf}]}$
- mean log relative risk: $\alpha \sim N(0, 100)$

donde:

- O_i, E_i es el número de casos observados y esperados en el área i
- $\lambda_i = \exp(\alpha + V_i)$: RR en el área i comparado con el riesgo esperado en base a la edad y sexo de la población
- parámetros V_i : efectos aleatorios area-specific
- residuos $RR = \exp(V_i)$

3. MAPAS DE ENFERMEDADES

Modelos de efectos aleatorios: Modelo de Poisson-lognormal-espacial

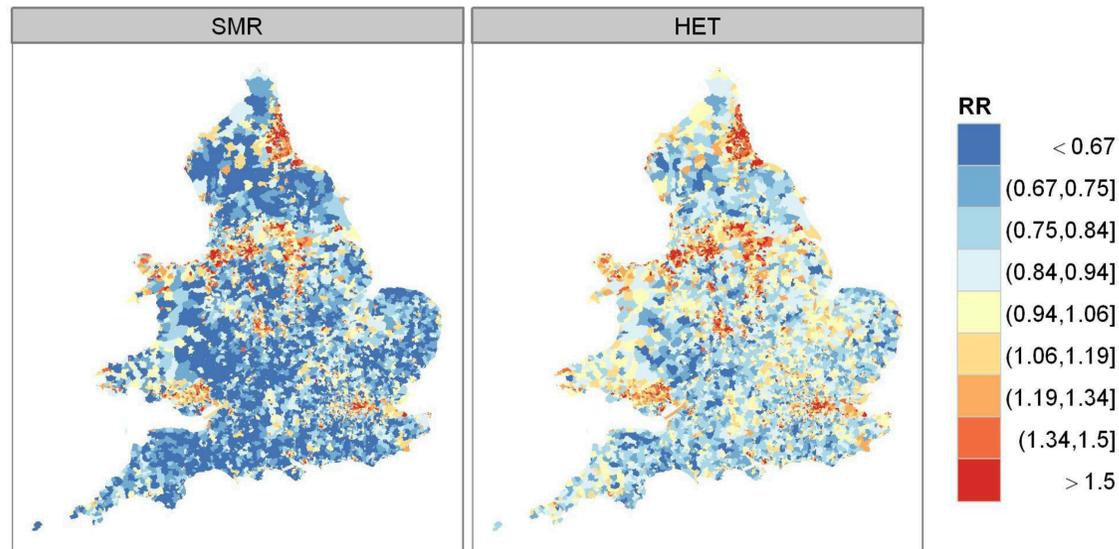
- Parámetros V_i : efectos aleatorios específicos para cada área para controlar la heterogeneidad (sobredispersión)
- Resumen de la variabilidad de los efectos aleatorios usando el cociente entre los cuantiles de su distribución empírica: es decir, se trata de clasificar los efectos aleatorios y calcular el cociente entre dos unidades en extremos opuestos
 - $\lambda_{5\%}$ y $\lambda_{95\%}$: RR para el área clasificada en el percentil 5% y 95%, respectivamente
 - $QR_{90} = \frac{\lambda_{95\%}}{\lambda_{5\%}}$: cociente de RR entre el 5% superior e inferior de las áreas

3. MAPAS DE ENFERMEDADES

➤ Modelos de efectos aleatorios: Modelo de Poisson-lognormal-espacial

Lung cancer incidence in males, 1985-2009, England and Wales (I)

RR estimates using 2 methods



SMRs and non-spatially smoothed RRs

3. MAPAS DE ENFERMEDADES

Software

- Para realizar el suavizado mediante Empirical Bayes, se puede utilizar la Rapid Inquiry Facility (RIF) de la SAHSU del Imperial College.
- La estimación de modelos jerárquicos bayesianos se realizaba mediante métodos de simulación computacionalmente complejos (MCMC)
 - Implementados en el software gratuito WinBUGS y GeoBUGS
- Desde 2008, el software gratuito INLA (Rue et al, 2008) implementa una aproximación rápida y más eficiente a los MCMC para estimar estos modelos.

3. MAPAS DE ENFERMEDADES

Modelos de efectos aleatorios: Modelo de Poisson-lognormal-espacial

- Los λ_i suelen estar correlacionados espacialmente porque reflejan, en parte, factores de riesgo que varían espacialmente
 - Incorporación de la dependencia espacial en la distribución de los λ_i
 - Modelo condicional autoregresivo (CAR)

$$\log (R_i) \sim \text{Normal} (m_i, v_i)$$

$$m_i = \sum_k R_k / n_i = \text{average risk in neighbouring areas}$$

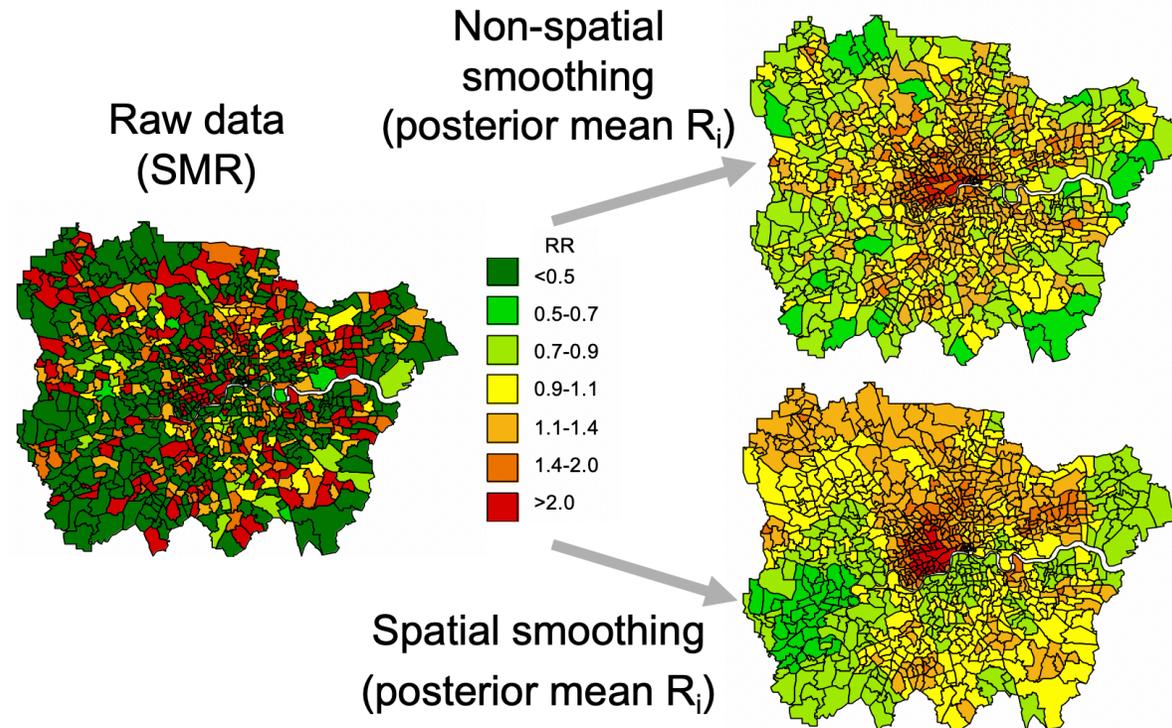
$$v_i = v / n_i \rightarrow \text{variance inversely proportional to number of neighbours}$$

Besag, York, Mollie (1991) *Annals of the Institute of Statistics and Mathematics*, 43: 1-59

3. MAPAS DE ENFERMEDADES

Modelos de efectos aleatorios: Modelo de Poisson-lognormal-espacial

Childhood leukaemia incidence in London, 1986-1998

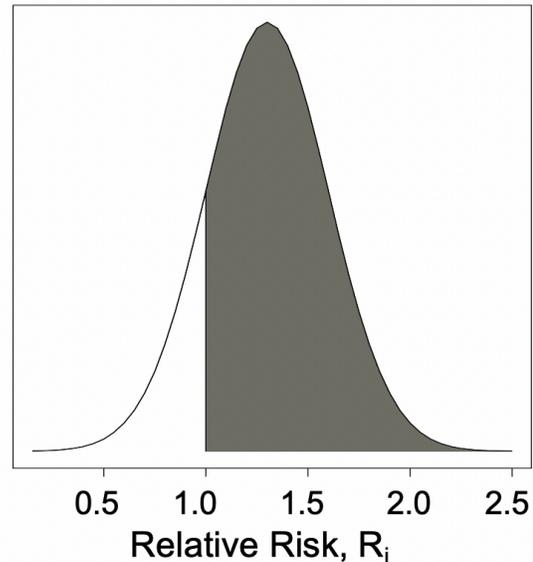


3. MAPAS DE ENFERMEDADES

Modelos de efectos aleatorios: Modelo de Poisson-lognormal-espacial

Mapping uncertainty

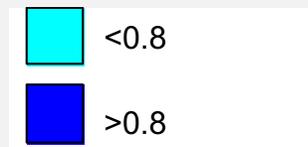
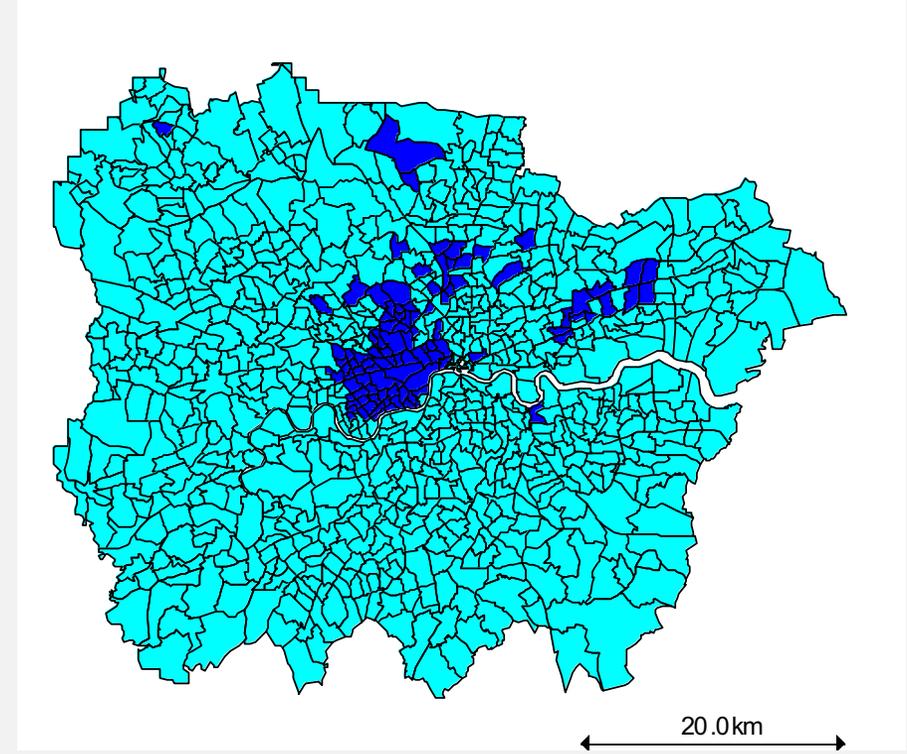
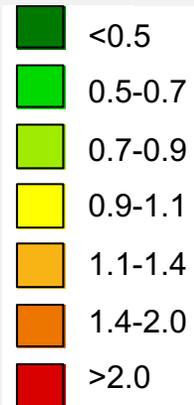
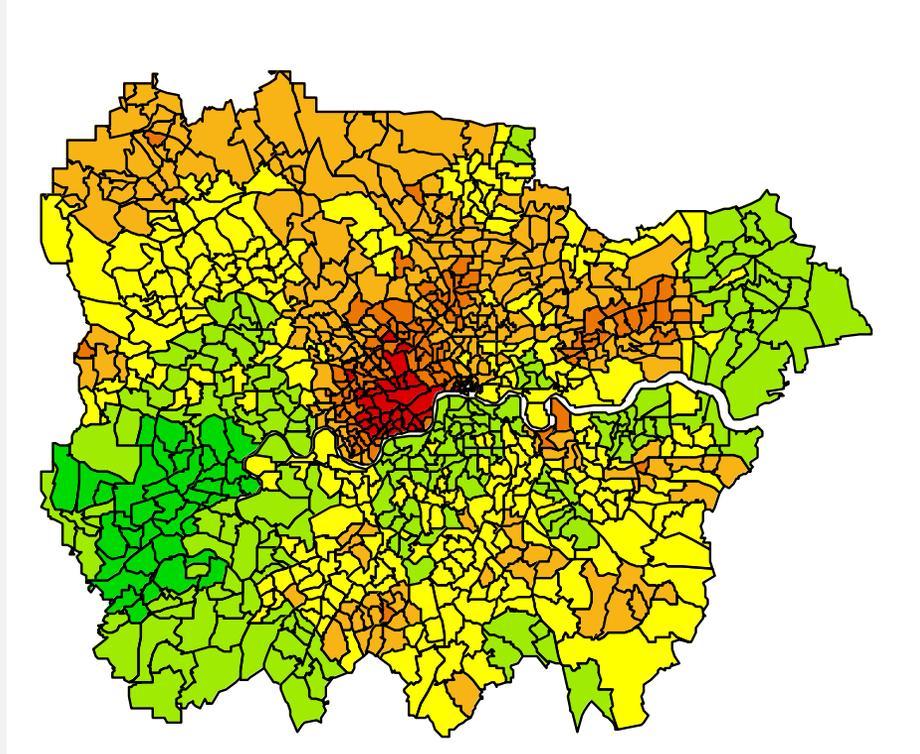
- Mapping the mean posterior value of R_i does not make full use of the posterior distribution



- Map **posterior SD**
- Map **Probability ($R_i > 1$)**

Note – this is not the same as a classical p-value

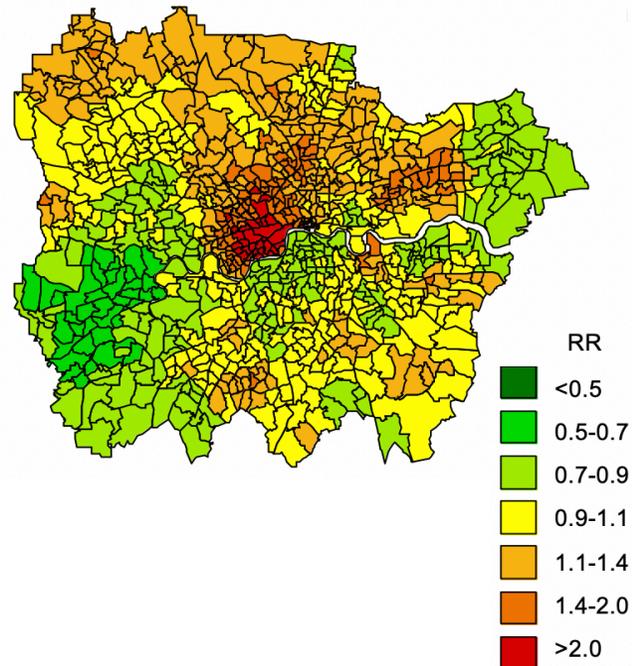
Childhood leukaemia incidence in London, 1986-1998



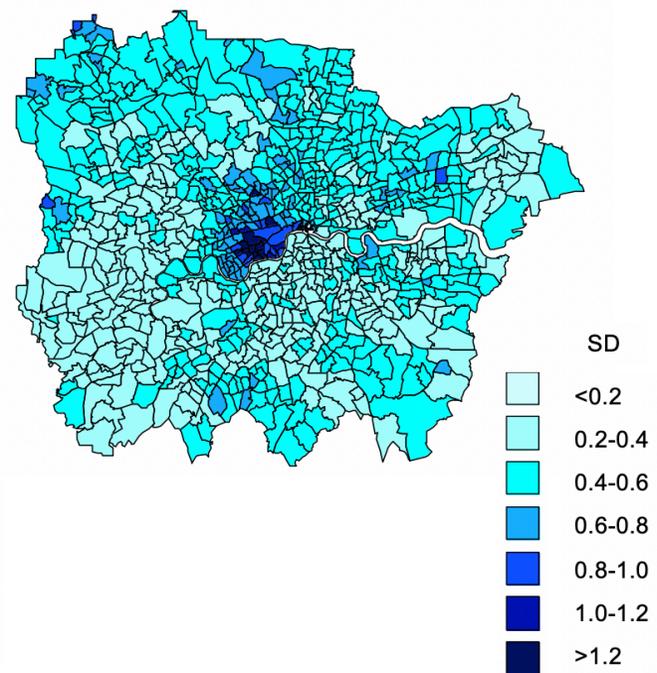
3. MAPAS DE ENFERMEDADES

Posterior SD of relative risk estimates

Posterior mean relative risk



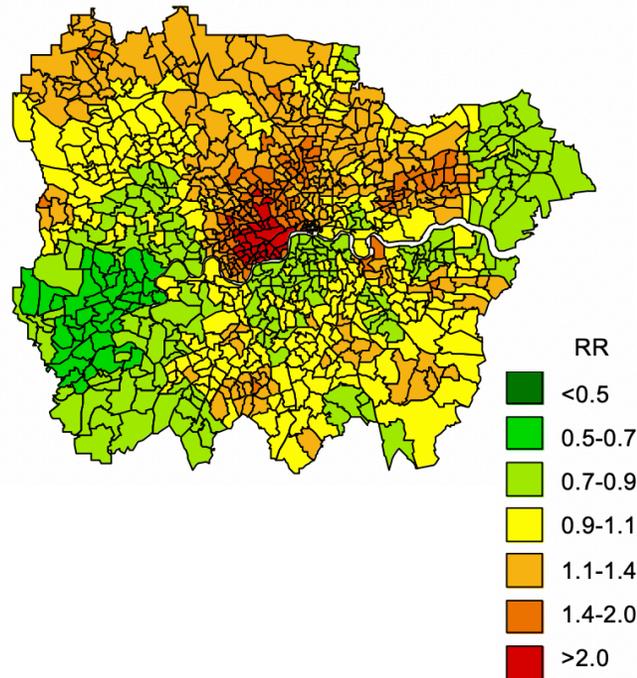
Posterior sd of relative risk



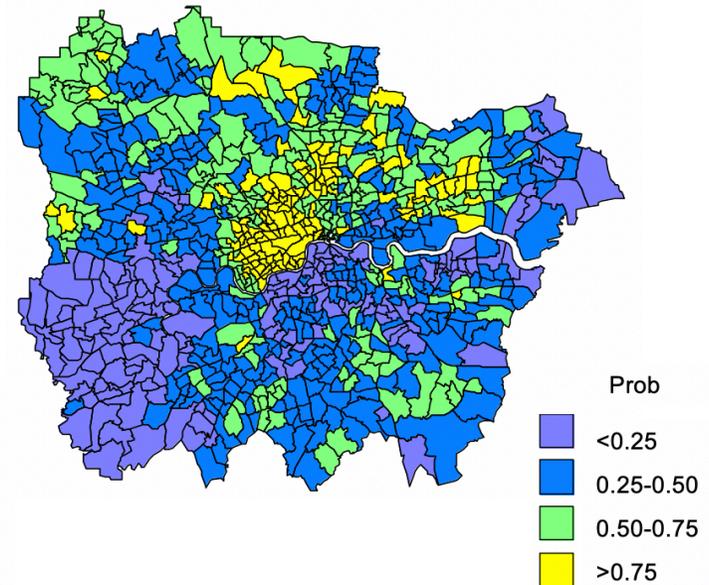
3. MAPAS DE ENFERMEDADES

Posterior probability that relative risk > 1

Posterior mean relative risk



Posterior probability that relative risk > 1



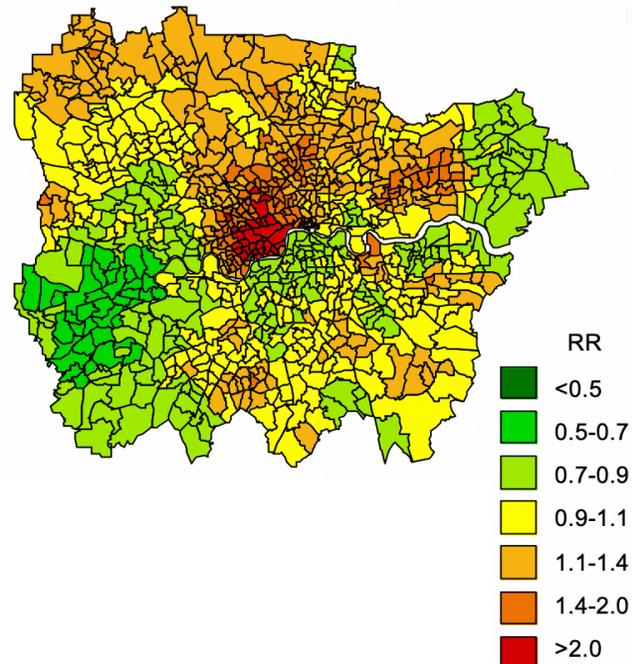
3. MAPAS DE ENFERMEDADES

- Richardson et al. (2004): estudio de simulación que investiga el uso de probabilidades a posteriori **en estudios enfocados en la representación de enfermedades en mapas**
 - Se clasificará una área como de elevado riesgo si $[Prob(R_i > 1)] > 0.8$
 - Se clasificará una área como de elevada especificidad si *falsa detección* $< 10\%$
 - Sensibilidad 60%-95% para E_i de 5-20 y verdadero R_i de 1.5-3.0

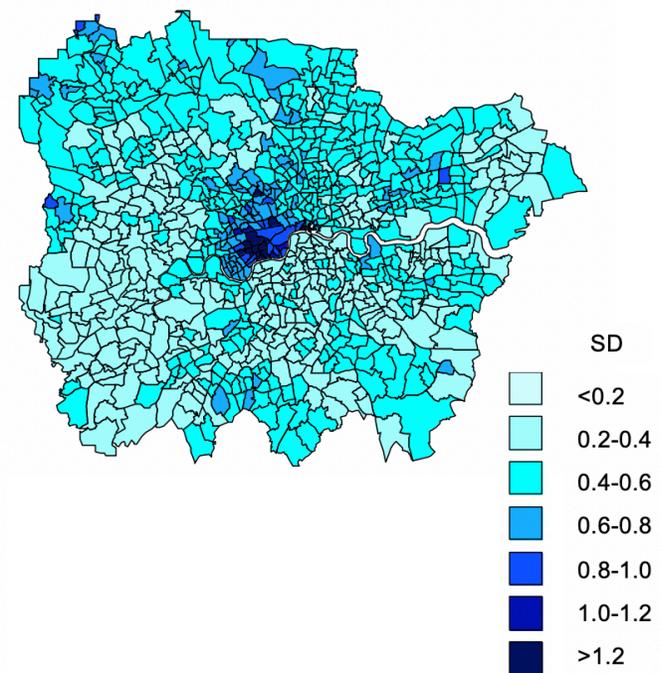
3. MAPAS DE ENFERMEDADES

Posterior SD of relative risk estimates

Posterior mean relative risk



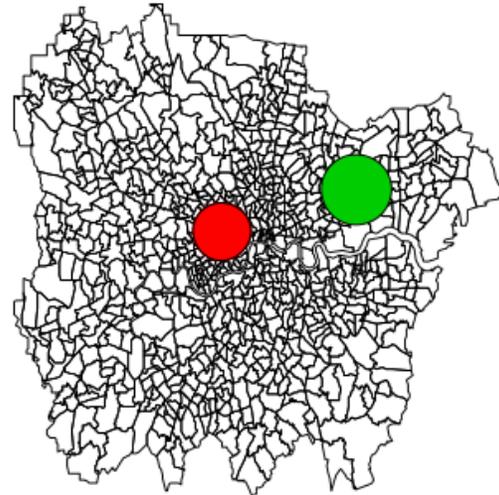
Posterior sd of relative risk



3. MAPAS DE ENFERMEDADES

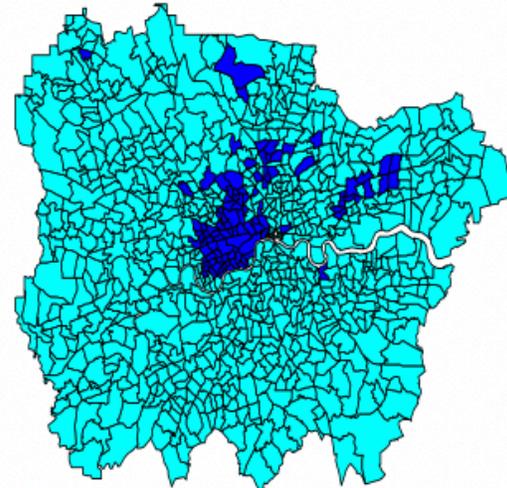
Comparison of SaTScan and Bayesian classification rule

SaTScan (Kuldorff,
www.satscan.org):
Location of most likely cluster



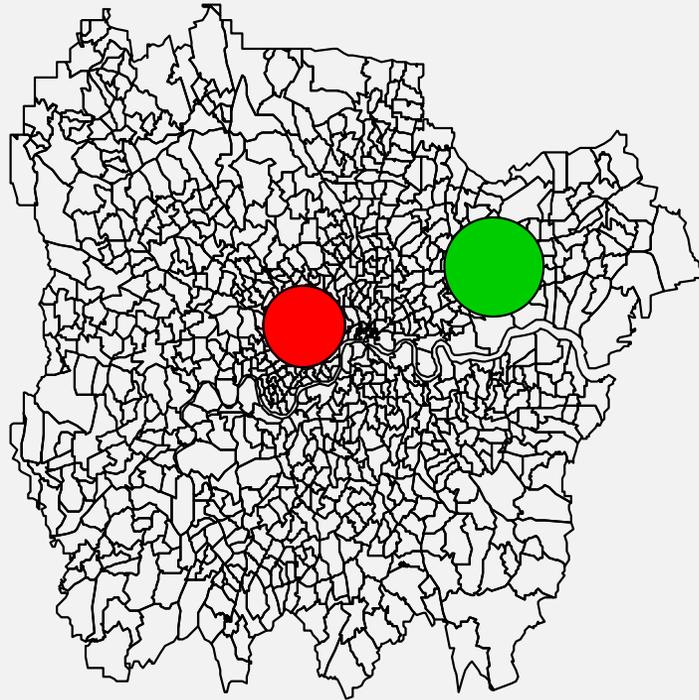
- Most likely cluster; $p < 0.001$
- 2nd most likely cluster; $p = 0.2$

Bayesian:
Probability of excess risk

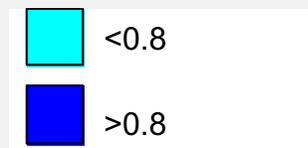
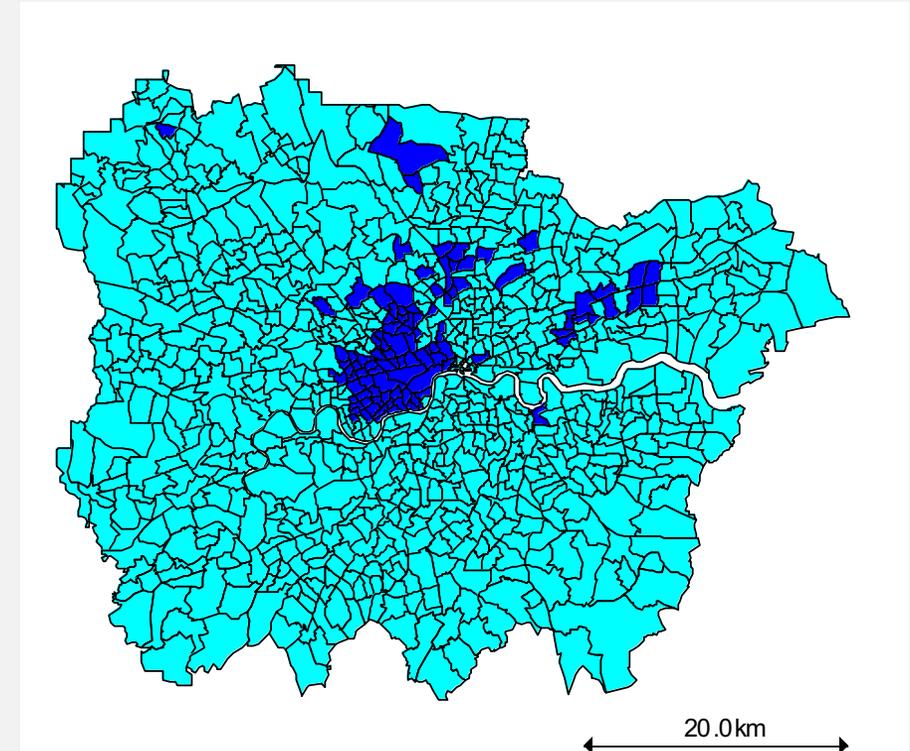


- < 0.8
- ≥ 0.8

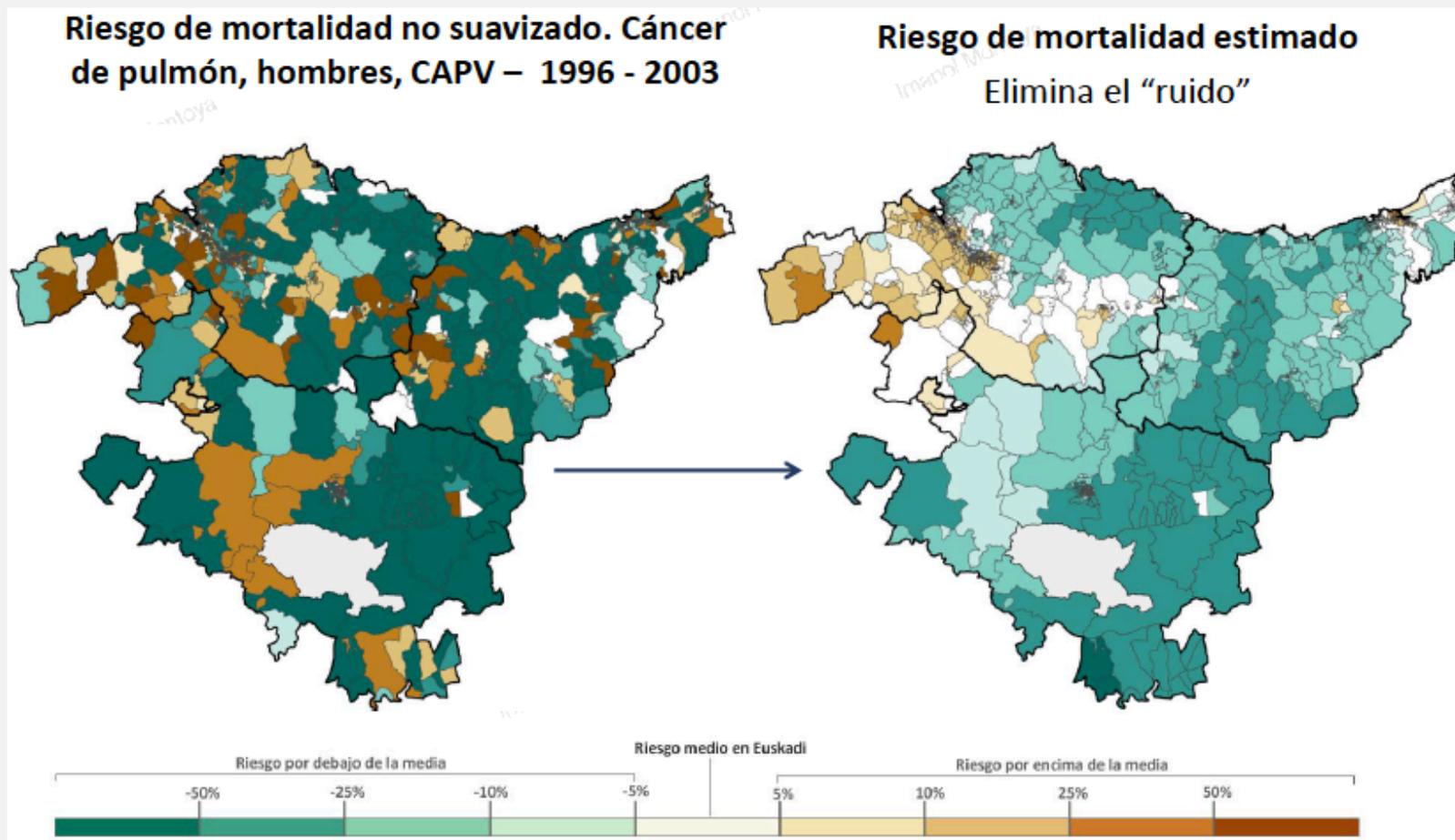
Childhood leukaemia incidence in London, 1986-1998



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



3. MAPAS DE ENFERMEDADES



3. MAPAS DE ENFERMEDADES. CONCLUSIONES

Carrying out a disease mapping study - protocol

Study health outcome, region of interest, period

- ▶ ICD codes, incidence, mortality
- ▶ Adjustments (strata): age, sex, socio-economic status
- ▶ Comparison area
- ▶ Map of the region of interest (shapefiles)

Data descriptive summary

- ▶ Population and observed cases per stratum and area
- ▶ Calculation of the expected numbers of cases, adjusted for the covariates, in each area
- ▶ Calculation of the SMRs

Non-spatially smoothed RRs using Poisson-logNormal model

- ▶ Write the model
- ▶ Format the data
- ▶ Run WinBUGS/OpenBUGS
- ▶ Sensitive analyses (priors on variance parameters)

Presentation of the results

- ▶ Quantitative summary (variance, quantile, ratio)
- ▶ Visual summary (histograms, maps)

ESTADÍSTICA Y EPIDEMIOLOGÍA ESPACIAL

1. Epidemiología y Epidemiología Espacial
2. Estadística Espacial
3. Mapas de Enfermedades
- 4. Regresión ecológica**
5. Agregación espacial (clustering)
6. Identificación de focos (cluster detection)



4. REGRESIÓN ECOLÓGICA

El objetivo de la **regresión ecológica** es examinar cómo las variaciones geográficas en las variables respuesta en salud se relacionan con las variaciones geográficas en la exposición de interés (por ejemplo, contaminación).

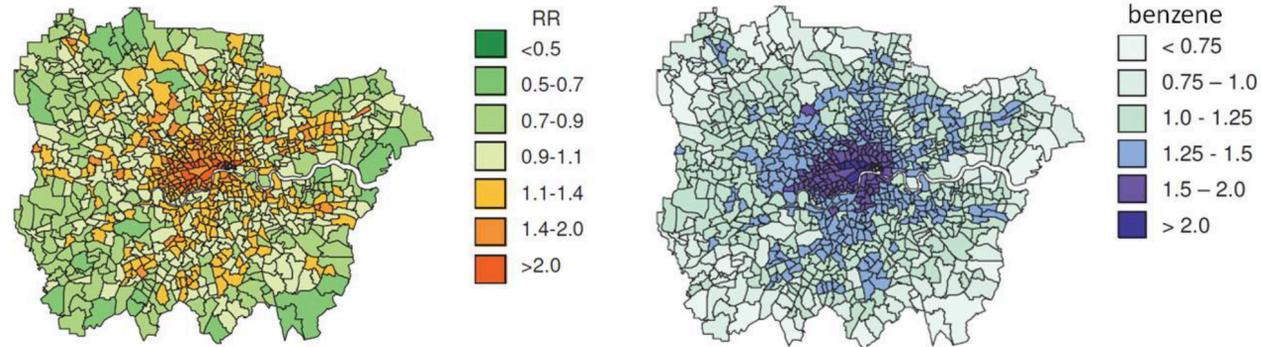
4. REGRESIÓN ECOLÓGICA

- Asociación entre el riesgo y las exposiciones a nivel de área
- Consiste en introducir variables explicativas (que recojan las variaciones geográficas en la exposición de interés) en los modelos jerárquicos utilizados para suavizar las razones de mortalidad y/o incidencia (BYM o log-Cox-Gaussian).
- En estos modelos, podemos contralar otros posibles confusores, introduciéndolos como variables explicativas adicionales.

4. REGRESIÓN ECOLÓGICA

- Leucemia en niños y benceno

Best et al, 2001, JRSSA



Small positive association found between average annual benzene emissions and risk of childhood leukaemia

ESTADÍSTICA Y EPIDEMIOLOGÍA ESPACIAL

1. Epidemiología y Epidemiología Espacial
2. Estadística Espacial
3. Mapas de Enfermedades
4. Regresión ecológica
5. **Agregación espacial (clustering)**
6. Identificación de focos (cluster detection)



5. AGREGACIÓN ESPACIAL (CLUSTERING)

- Hemos visto que los **mapas de enfermedades** proporcionan una ***primera vision de la distribución espacial de la enfermedad/evento de salud.***
- Un ***segundo paso*** en el estudio de la distribución espacial de la enfermedad es el ***análisis de agregación espacial.***

5. AGREGACIÓN ESPACIAL (CLUSTERING)

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

Por lo tanto, el estudio de aglomeración de casos (cluster detection) consiste en la identificación de agrupaciones “inusuales” de casos, permite identificar hotspots

5. AGREGACIÓN ESPACIAL (CLUSTERING)

¿Qué es un **cluster**?

- Según Knox (1989), *'Un cluster 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 cluster 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.

5. AGREGACIÓN ESPACIAL (CLUSTERING)

Métodos para el análisis de agregaciones

- Evaluación (detección) de la presencia de agrupaciones (clusters)
- Localización de agrupaciones (clusters)
- Agregaciones entorno a una fuente

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

5. AGREGACIÓN ESPACIAL (CLUSTERING)

Proceso de análisis de agregaciones espaciales (datos en rejilla)

- Análisis de heterogeneidad de riesgo
- Análisis de autocorrelación espacial
- Test de clustering general
- Localización de la agregación/es
- Análisis de asociación espacial

5. AGREGACIÓN ESPACIAL (CLUSTERING)

Análisis de heterogeneidad de 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 (mayores riesgos relacionados con una mayor exposición a dicho factor)
 - Otras fuentes

5. AGREGACIÓN ESPACIAL (CLUSTERING)

Análisis de heterogeneidad de riesgo

- En cada área de estudio, casos observados y esperados
- Test global de homogeneidad

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - \theta E_i)^2}{\theta E_i} \quad \theta = SMR = \frac{\sum_i O_i}{\sum_i E_i}$$

- Test de homogeneidad Potthoff and Whittinghill (distribución gamma)
 - Test de sobredispersión

$$\begin{aligned} H_0 : \theta_1 = \dots = \theta_n = \lambda \\ H_1 : \theta_i \sim Ga(\lambda^2/\sigma^2, \lambda/\sigma^2) \end{aligned} \quad PW = E_+ \sum \frac{O_i(O_i - 1)}{E_i}$$

- Otros

5. AGREGACIÓN ESPACIAL (CLUSTERING)

Análisis de heterogeneidad de riesgo

Tests for heterogeneity (tests of overdispersion)

To determine whether there is evidence of global heterogeneity of the RRs

$$H_0 : \lambda_1 = \lambda_2 = \dots = \lambda_N \text{ versus } H_1 : \lambda_i \neq \lambda_k \text{ for some } i \neq k$$

When evidence against H_0 is found, suggests that underlying disease risk varies across the study region \rightarrow interpreted as evidence of disease clustering i.e. pattern of location of disease cases is different to the pattern of population

- Likelihood ratio test

$$T = 2 \sum_{i=1}^N O_i [\log \text{SMR}_i - \log(O^+/E^+)] \underset{H_0}{\sim} \chi^2 \text{ with } N - 1 \text{ df}$$

- Score statistic of Potthoff and Whittinghill

$$T = E^+ \sum_i \frac{O_i(O_i - 1)}{E_i} \underset{H_0}{\sim} \text{Normal}(O^+(O^+ - 1), 2(N - 1)O^+(O^+ - 1))$$

Notations: $O^+ = \sum_i^N O_i$, $E^+ = \sum_i^N E_i$, with $N = \text{nb of areas}$

5. AGREGACIÓN ESPACIAL (CLUSTERING)

Autocorrelación espacial

- Dado un conjunto de entidades y un atributo asociado, evalúa si el patrón expresado está agrupado, disperso o es aleatorio

5. AGREGACIÓN ESPACIAL (CLUSTERING)

Autocorrelación espacial

- Test I de Moran
 - Comparar un índice de autocorrelación observado con uno esperado si no hubiera autocorrelación
 - Índice I de Moran

5. AGREGACIÓN ESPACIAL (CLUSTERING)

Autocorrelación espacial

- Test I de Moran
 - Índice I de Moran
 - Positivo = los valores altos se agrupan cerca de otros valores altos, los valores bajos se agrupan cerca de otros valores bajos
 - Negativo = los valores altos rechazan otros valores altos y tienden a estar cerca de valores bajos
 - 0 = los valores positivos equilibran los valores negativos

5. AGREGACIÓN ESPACIAL (CLUSTERING)

Autocorrelación espacial

- Test I de Moran
 - La hipótesis nula establece que el atributo que se analiza está distribuido en forma aleatoria entre las entidades del área de estudio, es decir, los procesos espaciales que promueven el patrón de valores observado constituyen una opción aleatoria
 - Se calcula el valor medio y la varianza para el atributo que se evalúa. A continuación, resta el valor medio en cada valor de la entidad, lo que crea una desviación del valor medio. Se evalúan las desviaciones de las entidades vecinas.

5. AGREGACIÓN ESPACIAL (CLUSTERING)

Autocorrelación espacial

➤ Test I de Moran

Moran's autocorrelation test

$$I = \frac{\sum_i \sum_j w_{ij} (Z_i - \bar{Z})(Z_j - \bar{Z})}{\left(\sum_i \sum_j w_{ij}\right) \sum_{k=1}^N (Z_k - \bar{Z})^2 / N} \text{ with } Z = \text{SMR}$$

- If there is no spatial dependence, I will be close to 0
- If there is clustering then areas close together (as defined by w_{ij}) will tend to have similar responses (i.e. SMR), so I will be positive
- If $I < 0$, clustering of dissimilar values
- "Significance" will be done using Monte Carlo approach

5. AGREGACIÓN ESPACIAL (CLUSTERING)

Autocorrelación espacial

➤ Test I de Moran

El valor P <i>no</i> es estadísticamente significativo.	No puede rechazar la hipótesis nula. Es posible que la distribución espacial de los valores de entidades sea el resultado de procesos espaciales aleatorios. El patrón espacial observado de los valores de entidades podría ser cualquiera de las tantas versiones posibles de aleatoriedad espacial completa (CSR).
El valor P es estadísticamente significativo y la puntuación z es positiva.	Puede rechazar la hipótesis nula. La distribución espacial de los valores altos y los valores bajos en el dataset está más agrupada espacialmente de lo que se esperaría si los procesos espaciales subyacentes fueran aleatorios.
El valor P es estadísticamente significativo y la puntuación z es negativa.	Puede rechazar la hipótesis nula. La distribución espacial de los valores altos y los valores bajos en el dataset está más dispersa espacialmente de lo que se esperaría si los procesos espaciales subyacentes fueran aleatorios. Un patrón espacial disperso suele reflejar algún tipo de proceso competitivo: una entidad con un valor alto rechaza a otras entidades con valores altos; del mismo modo, una entidad con un valor bajo rechaza a otras entidades con valores bajos.

5. AGREGACIÓN ESPACIAL (CLUSTERING)

Test de clustering general

- Test de Tango
 - Test global de clustering comparando observados y esperados en cada region
 - Considera el decaimiento de la fuerza de la interacción entre vecinos con la distancia

5. AGREGACIÓN ESPACIAL (CLUSTERING)

Test de clustering general

➤ Test de Tango

Tango's test

$$C = (\mathbf{r} - \mathbf{p})^t \mathbf{A} (\mathbf{r} - \mathbf{p})$$

where

- $A_{ij} = \exp(-d_{ij}/\phi)$ measure of closeness, with ϕ chosen a priori to reflect the expected extend of spatial dependence
- $\mathbf{r} = [O_1/O^+, \dots, O_N/O^+]$ Observed proportions of cases in each area
- $\mathbf{p} = [E_1/E^+, \dots, E_N/E^+]$ Expected proportions of cases in each area

Monte Carlo tests can be used to obtain p-values

5. AGREGACIÓN ESPACIAL (CLUSTERING)

Test de clustering general

➤ Test

Autocorrelation statistics for assessment of clustering

- A number of approaches have been suggested for measuring spatial autocorrelation - these are global measures
- Form of the measure of global clustering

$$T = c \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij} \times \text{similarity}_{ij}}{\sum_{i=1}^N \sum_{j=1}^N w_{ij}}$$

where

- ▶ c constant
 - ▶ N number of areas
 - ▶ w_{ij} weight reflecting the proximity between areas i and j
 - ▶ similarity_{ij} measure of similarity between data values in areas i and j
- For the most common choices, the mean and the variance of T under the null hypothesis of no clustering are not available, and asymptotic normality is not reliable → Monte Carlo approach

5. AGREGACIÓN ESPACIAL (CLUSTERING)

Test de clustering general

➤ Test

Measures of proximity

- Various ways of measuring the 'closeness' of 2 areas
 - ▶ $w_{ij} = 1$ if areas i and j are adjacent (common boundary), 0 otherwise
 - ▶ $w_{ij} = 1$ if the centroids of areas i and j are within a certain distance of each other
 - ▶ $w_{ij} = d_{ij}^{-1}$ where d_{ij}^{-1} is the inverse of the distance between the centroids of areas i and j
- The choice of weights depends on the type of spatial dependence we want to detect.

ESTADÍSTICA Y EPIDEMIOLOGÍA ESPACIAL

1. Epidemiología y Epidemiología Espacial
2. Estadística Espacial
3. Mapas de Enfermedades
4. Regresión ecológica
5. Agregación espacial (clustering)
6. **Identificación de focos (cluster detection)**



Epidemiología Espacial

5. IDENTIFICACIÓN DE FOCOS (CLUSTER DETECTION)

Localización de la agregación/es

- Scan Statistics: familia de métodos
- Ejemplos:
 - Openshaw's Geographical Analysis Machine
 - Kulldorff
 - Test de Stone

5. IDENTIFICACIÓN DE FOCOS (CLUSTER DETECTION)

Localización de la agregación/es

- 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

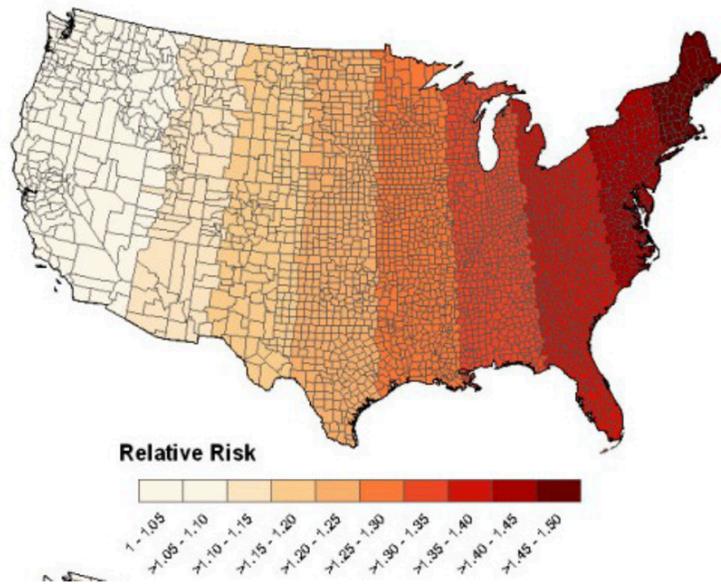
5. IDENTIFICACIÓN DE FOCOS (CLUSTER DETECTION)

Localización de la agregación/es

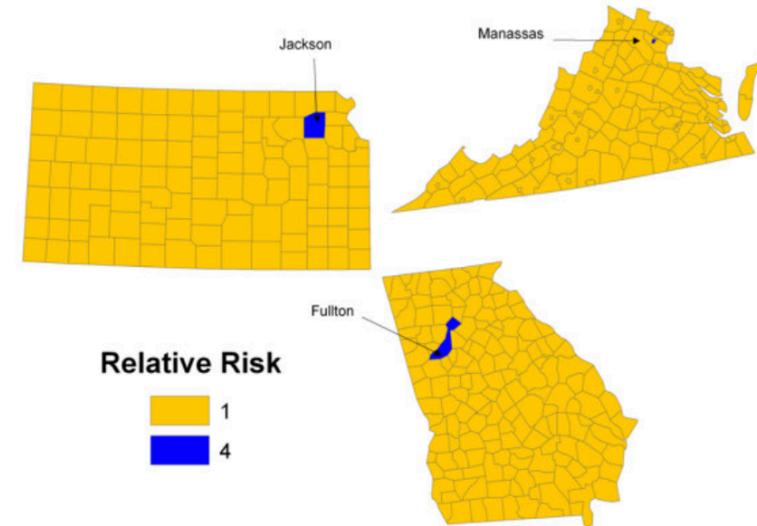
- Scan Statistics: familia de métodos
- Kulldorff
 - Hipótesis nula: no hay agregación = los dos riesgos son iguales
 - Hipótesis alternativa: el riesgo es mayor dentro de la ventana
 - Likelihood ratio test
 - El cluster más probable es aquel donde se encuentre el mayor likelihood ratio con la mejor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

5. IDENTIFICACIÓN DE FOCOS (CLUSTER DETECTION)

Localización de la agregación/es



Global clustering
From Jackson et al, 2009



Localised clusters

5. IDENTIFICACIÓN DE FOCOS (CLUSTER DETECTION)

Detección de patrones de heterogeneidad

- Sobredispersión: recuentos independientes O_i con $\text{Var}(O_i) > E(O_i)$
- Dependencia espacial, clustering: dependencia entre O_i y O_k que está relacionada con la posición geográfica de las áreas i y k

5. IDENTIFICACIÓN DE FOCOS (CLUSTER DETECTION)

Detección de patrones de heterogeneidad

- Las agrupaciones pueden darse debido a
 - Un agente infeccioso
 - Susceptibilidad genética
 - Factores de riesgo que varían espacialmente. Ajustaremos por factores de riesgo relevantes
- Tests de dependencia espacial basados en una medida de la "cercanía"

5. IDENTIFICACIÓN DE FOCOS (CLUSTER DETECTION)

Identificación de focos

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

5. IDENTIFICACIÓN DE FOCOS (CLUSTER DETECTION)

Identificación de focos

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



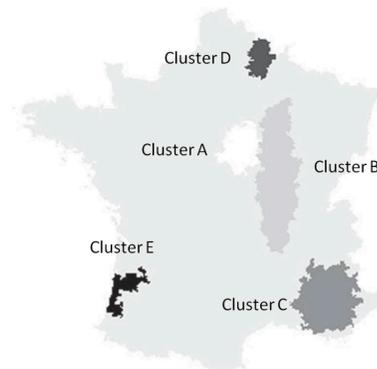
5. IDENTIFICACIÓN DE FOCOS (CLUSTER DETECTION)

Identificación de focos

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



- cluster A: AL for 90–94, $P=0.08$
- cluster B: AML for 95–00, $P=0.08$
- cluster C: ALL for 95–00, $P=0.06$
- cluster D: ALL 1–4 years for 95–00, $P=0.09$
- cluster E: ALL 1–4 years for 01–06 $P=0.08$

5. IDENTIFICACIÓN DE FOCOS (CLUSTER DETECTION)

Agregación espacial e Identificación de focos

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)

ESTADÍSTICA Y EPIDEMIOLOGÍA ESPACIAL

1. Epidemiología y Epidemiología Espacial
2. Estadística Espacial
3. Mapas de Enfermedades
4. Regresión ecológica
5. Agregación espacial (clustering)
6. **Identificación de focos (cluster detection)**



Epidemiología Espacial

THANK YOU

For Your Precious Time and Attention...