

# EL ABUSO DE DATOS AGREGADOS EN LOS ESTUDIOS EPIDEMIOLÓGICOS DE ÁREAS PEQUEÑAS

R. Ocaña-Riola; C. Sánchez-Cantalejo; S. Toro-Cárdenas

## INTRODUCCIÓN Y OBJETIVOS

Habitualmente, los mapas de indicadores de salud describen la distribución geográfica de la morbi-mortalidad agrupando datos de varios años. En ocasiones, la amplitud del periodo de estudio ha llegado a agrupar más de 15 años de información.

Cuando se utilizan datos agregados, el peso de la incidencia o mortalidad de años pasados puede hacer que el exceso de enfermedad o de muertes observado en algunas áreas no se corresponda con la realidad actual. Además de la correlación espacial, puede existir una correlación temporal no contemplada al agrupar años.

El objetivo de este trabajo es mostrar que el uso de agregaciones de datos anuales puede sesgar las estimaciones del riesgo relativo en estudios epidemiológicos de áreas pequeñas.

## RESULTADOS

Las ecuaciones utilizadas mostraron que el riesgo relativo obtenido con datos agregados ( $\phi_m$ ) es igual al riesgo relativo estimado mediante datos desagregados ( $\mathcal{G}_m$ ) multiplicado por una constante ( $B_m$ ):

$$\phi_m = \mathcal{G}_m B_m \quad \text{con} \quad \mathcal{G}_m = e^{\beta_0 + u_m + v_m}, \quad B_m = \frac{\sum_{t=1}^T E_{mt} e^{c_i^{(p)} + d_i}}{E_m}$$

Si  $B_m$  es mayor que uno, entonces el modelo con datos agregados sobrestimaré el riesgo relativo del área. Si es menor que uno, entonces el modelo con datos agregados subestimaré el riesgo relativo. Únicamente si no existe una correlación temporal de los datos se tendrá un coeficiente igual a uno y por tanto el modelo con datos agregados en un único periodo será adecuado.

Ambos modelos fueron implementados en el software WinBUGS para aplicarlos al estudio de la mortalidad masculina por todas las causas en los municipios de la provincia de Granada durante el periodo 1985-1999.

El modelo espacio-temporal utilizado fue de orden  $p=2$ , considerando que cada error temporal estaba correlacionado con los dos anteriores o posteriores. Las estimaciones se realizaron utilizando inferencia bayesiana.

El coeficiente de sobrestimación del riesgo relativo,  $B_m$ , estuvo comprendido entre 1.09 y 1.11 en cada municipio (Tabla 1). Los resultados mostraron una sobreestimación media del riesgo por municipio del 10% cuando se utilizaron datos agregados en el tiempo que presentaban una correlación temporal.

## MÉTODOS

Se utilizaron las propiedades de la distribución de Poisson para relacionar el riesgo relativo estimado mediante datos desagregados en el tiempo, con el calculado a partir de datos agregados en un único periodo. La descripción de los modelos es la siguiente:

Información temporal	Supuesto	Modelo jerárquico bayesiano
Desagregada	$O_{mt} \sim \text{Poisson}(\mathcal{G}_m E_{mt})$	$\log(\mathcal{G}_m) = \beta_0 + (u_m + v_m) + (c_i^{(p)} + d_i)$
Agregada	$O_m \sim \text{Poisson}(\phi_m E_m)$	$\log(\phi_m) = \beta_0 + (u_m + v_m)$

$O_{mt}$  casos observados en el área  $m$  y tiempo  $t$   
 $E_{mt}$  casos esperados en el área  $m$  y tiempo  $t$   
 $\mathcal{G}_m$  riesgo relativo del área  $m$  en el tiempo  $t$

$O_m$  casos observados en el área  $m$  durante todo el periodo  
 $E_m$  casos esperados en el área  $m$  durante todo el periodo  
 $\phi_m$  riesgo relativo del área  $m$  en todo el periodo de estudio

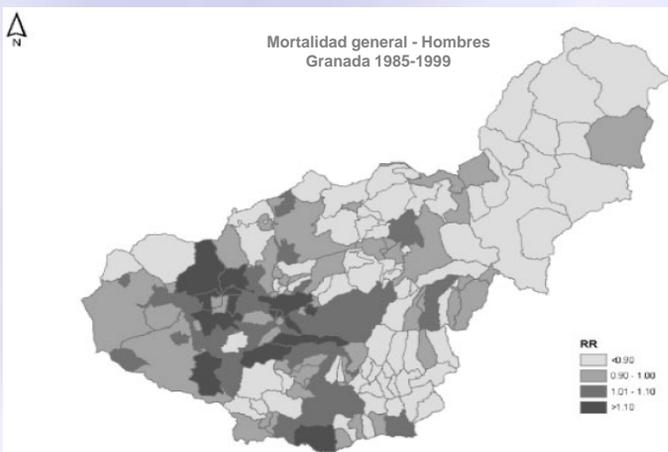
$u_m$  efecto aleatorio espacial estructurado  
 $v_m$  efecto aleatorio espacial no estructurado

$c_i^{(p)}$  efecto aleatorio temporal estructurado de orden  $p$   
 $d_i$  efecto aleatorio temporal no estructurado  
 $p$  errores temporales adyacentes con correlación

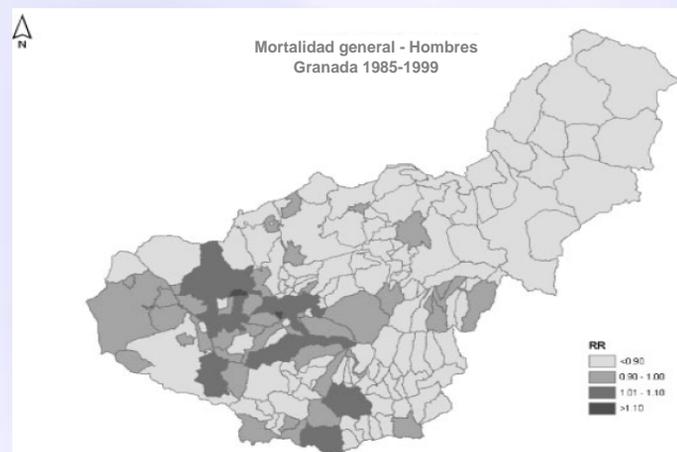
Tabla 1. Descripción del Riesgo Relativo suavizado (Granada, 1985-1999)

	Datos desagregados	Datos agregados en el tiempo	
	Riesgo relativo ( $\mathcal{G}_m$ )	Riesgo relativo ( $\phi_m$ )	Coficiente ( $B_m$ )
Mínimo	0.13	0.62	1.09
Máximo	1.13	1.22	1.11
Media	0.84	0.92	1.10
Desviación típica	0.15	0.13	0.004
Percentiles			
25	0.77	0.83	1.10
50	0.85	0.92	1.10
75	0.94	1.02	1.10

Riesgo relativo con datos agregados en un periodo de 15 años



Riesgo relativo con datos desagregados



## CONCLUSIONES

A pesar de ser una práctica habitual, la unión de varios años para formar un periodo de estudio más amplio puede llevar asociado un error importante en la estimación de los riesgos de incidencia o mortalidad de las áreas geográficas. Como consecuencia, la toma de decisiones y las políticas sanitarias aplicadas a partir de la interpretación de estos mapas pueden ser inadecuadas. Una descripción más detallada de este trabajo puede verse en *Statistics in Medicine* 2007; 26: 4489-4504.