VIGILANCIA EN SALUD PÚBLICA

INDICE

VIGILANCIA EN SALUD PÚBLICA....3
CAPTURA-RECAPTURA....4
Conceptos generales....4
Uso de la técnica de captura-recaptura en el campo de la salud....4
Limitaciones de la técnica de captura-recaptura....5
Pasos para aplicar la técnica de captura-recaptura....6
Bibliografía....10
DETECCIÓN DE CLUSTERS....12
Conceptos generales....12
El proceso de evaluación de un conglomerado....12
AGREGACIONES TEMPORALES....13
Casos individuales: método CHEN....14
Casos individuales: método SETS....16
Casos agrupados: Método Scan....18
Casos agrupados: Método Texas....22
Casos agrupados: Método Poisson....26
Casos agrupados: Método Cusum....28
AGREGACIONES ESPACIALES....30
Método Grimson....30
Método Ohno....33
AGREGACIONES ESPACIO-TEMPORALES....36
Método Knox....36
OTRAS AGREGACIONES: Método REMSA....39
Bibliografía....42
GRÁFICO DE LÍMITES HISTÓRICOS....44
Conceptos generales....44
Recomendaciones....44
Manejo del submódulo y solución del ejercicio....45
Bibliografía....48
GRÁFICO DE DISTRIBUCIÓN ESPACIAL....49
Conceptos generales....49
Recomendaciones....49
Manejo del submódulo y solución del ejercicio....50
Bibliografía....51
CORREDORES ENDÉMICOS....51
Conceptos generales....51
Método de construcción del corredor endémico....52
Recomendaciones....53
Manejo del submódulo y solución del ejercicio....53
Bibliografía....56
ONDAS EPIDÉMICAS....57
Conceptos generales....57
Análisis de la onda epidémica....58
Velocidad de difusión de una onda epidémica....59
Utilidades del método e interpretación de resultados....60
Tasa de propagación....60
Recomendaciones y advertencias....60

Manejo del submódulo y solución del ejercicio....61
Bibliografía....62
EFECTIVIDAD VACUNAL....64
Conceptos generales....64
Recomendaciones....65
Parámetros....66
Bibliografía....69

VIGILANCIA EN SALUD PÚBLICA

La *Vigilancia en Salud Pública* se puede definir como "la recogida, análisis e interpretación continua y sistemática de datos concretos para usarlos en la planificación, implementación y evaluación de la práctica en salud pública". Un sistema de vigilancia incluye la capacidad funcional para la recogida y el análisis de datos, así como la oportuna distribución de la información derivada de esos datos a personas que pueden llevar a cabo actividades de prevención y control efectivas.

El objetivo de este módulo de Epidat 3.1 es facilitar una serie de herramientas comúnmente utilizadas en las unidades de vigilancia. Incluye los siguientes procedimientos:

- Captura-Recaptura
- Detección de clusters
  - Agregaciones temporales
    - Datos individuales
    - Datos agrupados
  - Agregaciones espaciales
  - Agregaciones espacio-temporales
  - Otras agregaciones
- Gráficos
  - Gráfico de límites históricos
  - Gráfico de distribución espacial
  - Corredores endémicos
- Ondas epidémicas
- Efectividad vacunal

CAPTURA-RECAPTURA

Conceptos generales

El método de captura-recaptura fue originalmente desarrollado por los zoólogos con la finalidad de estimar el tamaño de poblaciones de animales salvajes[1]. Este método supone la captura de una muestra de la población de animales, su marcado, y su posterior liberación. Luego de un tiempo determinado, se selecciona otra muestra, lo que da lugar a que algunos especimenes sean "recapturados", y a través de la proporción de animales marcados y no marcados en la segunda muestra, y de los tamaños de ambas, se consigue estimar la magnitud de la población total de animales.

Esta técnica, adaptada para uso demográfico por Sekar y Deming[2], está siendo usada con frecuencia en el campo de la epidemiología y otros campos de la salud para estimar el tamaño de la población no registrada[3], y utilizar esas estimaciones para generar tasas de incidencia o prevalencia de una enfermedad o evento de salud específico. Por lo general, este tipo de información es crucial para la toma de decisiones de los directivos de la salud pública y los políticos.

Objetivo

La técnica de captura-recaptura tiene como objetivo determinar el número total de individuos de una población determinada, para lo cual se ha efectuado la recolección de datos de diversas fuentes de información que han "capturado" datos sobre los miembros de esa población.

Uso de la técnica de captura-recaptura en el campo de la salud

Para su uso en el campo de la epidemiología, el principio de este método es cruzar información de diferentes sistemas de registro, y estimar a partir de ellos el número de pacientes no identificados por todas esas fuentes.

Para realizar esas estimaciones, se contemplan los sujetos que aparecen en más de una lista, luego de lo cual se utilizan diversas técnicas estadísticas para calcular el número total de sujetos que deberían figurar en los registros. Estas técnicas están ya bien validadas, y están siendo usadas en estudios sobre la epidemiología de diabetes[4], epilepsia[5], cáncer[6-8], accidentes cerebrovasculares[9,10], enfermedades mentales[11] y consumo de drogas[12], entre otros. Para ello se están utilizando bases de datos con las listas de pacientes de médicos particulares, registros clínicos de hospitales especializados, estadísticas de egresos hospitalarios, etc.

Para garantizar la validez de los resultados obtenidos por este método han de cumplirse los siguientes criterios:

1. **La población es cerrada.** Es decir, que no existen cambios significativos en el tamaño de la población durante el periodo de estudio, por ejemplo, debido a migración o muerte de un importante número de personas.
2. **Las listas son independientes unas de otras**. Esto significa que la inclusión de un sujeto en un registro no influye en su inclusión en otro registro. Esto no siempre se da entre las poblaciones humanas; así, por ejemplo, un paciente diabético registrado en una clínica especializada estará, probablemente, en la lista de médicos de familia, ya que los pacientes diabéticos vistos en esa clínica son referidos por esos médicos.
3. **Las probabilidades de captura son homogéneas**, es decir, todos los miembros de la población tienen la misma probabilidad de ser capturados o registrados en cada una

de las fuentes de datos o listas. Dado que tal probabilidad puede estar influida por rasgos tales como la selección del lugar de estudio, los horarios de atención a los pacientes, especialmente si el periodo de estudio es corto, y la capacidad económica de los pacientes, hay que tener mucho cuidado en la selección y evaluación previa de las fuentes de datos.

4. **Todos los miembros de la población pueden ser identificados, en forma precisa, en cada una de las listas.** Para ello se debe tener al menos un identificador común (por ejemplo, el número nacional de identificación). Aunque en la practica, el nombre, apellido, edad o fecha de nacimiento, son usualmente suficientes para identificar a todas las personas en la mayoría de países, en algunos países en desarrollo pueden existir algunos problemas para la identificación de las personas, ya que en ocasiones los nombres pueden coincidir o ser similares, y la fecha de nacimiento y un número de identificación nacional no están disponibles.

Limitaciones de la técnica de captura-recaptura

Algunos de los problemas que puede encontrarse en la implementación de la técnica de captura-recaptura son:

- Dependencia entre las fuentes de información.
- Pobre calidad de los registros.
- Confidencialidad de los datos.
- Variaciones del identificador según fuentes de registro.
- Registros incompletos para algunas poblaciones específicas.

Un punto crucial para obtener datos fiables con esta técnica es la independencia entre las fuentes de los datos. Si hay dependencia, entonces la inclusión de un sujeto en un registro tiene un efecto causal directo sobre la inclusión de ese individuo en otro; cuando la identificación de casos por una fuente aumenta la probabilidad de que esos casos sean registrados por otra se habla de dependencia positiva y, si dicha probabilidad disminuye, la dependencia es negativa. La dependencia produce un sesgo en la estimación del número de casos: según sea positiva o negativa se produce infraestimación o sobreestimación, respectivamente.

Cuando los datos proceden de dos fuentes, no es posible evaluar la independencia entre ambas; en este caso, se considera que la posible dependencia se puede valorar a través del conocimiento que se tenga del entorno en que se aplica la estimación o de la simple afirmación de que los circuitos de recogida son diferentes.

En el contexto de múltiples listas, es decir, cuando se dispone de tres o más fuentes, se pueden aplicar modelos de regresión log-lineal que permiten incorporar y evaluar diferentes asunciones en cuanto a la dependencia existente entre las listas.

El modelo log-lineal permite estudiar la relación existente entre k variables cualitativas cruzadas en una tabla de contingencia de 2^k entradas; dicho modelo representa el logaritmo neperiano de las frecuencias observadas en las celdas como una combinación lineal de efectos principales y de interacciones.

En el caso particular del método de captura-recaptura, la tabla de contingencia es incompleta, es decir, tiene una celda desconocida que corresponde a los casos no notificados por ninguno de los registros, y que es necesario estimar para obtener la estimación del total de casos en la población.

Los términos de interacción del modelo varían según la relación de dependencia existente entre las fuentes. Con tres listas, por ejemplo, hay 8 modelos posibles: uno que asume

dependencia entre las tres, otro que las supone independientes y los restantes consideran las diferentes posibilidades de independencia entre pares de listas o de independencia de cada una con las otras dos. Para obtener la estimación del número de casos en la población, se ajustan los diferentes modelos log-lineales y se selecciona el más adecuado. Una medida de bondad de ajuste muy utilizada para la selección de modelos es el estadístico G^2 de razón de verosimilitudes, que debe ser lo más pequeño posible. Como medida de complejidad se usa el número de grados de libertad del modelo, que disminuye a medida que aumenta la complejidad. Un buen modelo debe tener un valor de G^2 pequeño y un valor alto en los grados de libertad; sin embargo, a medida que aumentan los grados de libertad, se tienen modelos menos complejos, pero su ajuste suele ser peor, por lo que también aumenta el valor de G^2. Por ejemplo, el modelo saturado, que incluye todos los términos de interacción y corresponde al caso de completa dependencia entre todas las fuentes, tiene un valor de $G^2=0$, lo que significa un ajuste perfecto, y también 0 grados de libertad, que se traduce en complejidad máxima. Un estadístico que contempla simultáneamente la complejidad y el ajuste y, por tanto, facilita la selección, es el criterio de información bayesiano (BIC) que debe tomar valores bajos[3,13]. Debe tenerse en cuenta que los criterios estadísticos rara vez conducen a la selección de un único modelo, sino que más bien permiten descartar modelos claramente inapropiados; así pues, el proceso de selección debe complementarse con la información y el conocimiento que el investigador tiene acerca de los registros utilizados.

Por otra parte, se requiere que los sistemas de monitoreo y de vigilancia en salud sean confiables, ya que los registros de mala calidad podrán subestimar o sobrestimar el tamaño real de la población en estudio. Se necesita, además, que todos los casos incluidos hayan sido confirmados, lo cual es a veces conflictivo, en particular en países o zonas con pobre infraestructura sanitaria. Además, el subregistro puede alterar significativamente los resultados obtenidos.

Si bien la validez de esta metodología ha sido probada y reconocida, un estudio que tomaba como base un valor referencial (*'gold standar'*) encontró que los procedimientos de captura-recaptura proveían inferencias equivocadas en cuanto al número de casos no registrados, y que para lograr un valor cercano al del valor referencial se requería de modelos más complejos de los utilizados regularmente[14].

Pasos para aplicar la técnica de captura-recaptura

1. Identificar las diferentes listas o registros de pacientes. Dependiendo del problema a evaluar, así como de los objetivos específicos del estudio, las fuentes de los datos pueden variar; entre las más comunes se hallan los registros hospitalarios, clínicas especializadas, médicos particulares, escuelas, encuestas, registros de defunciones, registros laborales, etc.
2. Determinar si los registros seleccionados cumplen con los criterios para su utilización. Se deberá evaluar si las listas son independientes, permiten la identificación de todos los pacientes y si éstos pueden ser emparejados en las distintas listas.
3. Crear un identificador único para cada caso incluido. Cada persona incluida en cada una de las listas deberá tener un identificador único que permita ubicarlo en cada una de las listas en forma precisa y sin posibilidad de errores. Puede usarse el número de identidad, el nombre de la persona, etc., dependiendo de las variables que se hayan registrado en cada una de las listas.
4. Comparar las diferentes listas para identificar los casos que se encuentran registrados en varias listas. Teniendo la forma de identificar a cada una de las personas en las diferentes listas, se debe hacer el recuento de cuántos de ellos se encuentran repetidos

en las diferentes listas y cuántos sólo se encuentran registrados en una única lista. Los valores obtenidos pueden ser colocados en un diagrama de Venn.

Epidat 3.1 permite aplicar el método de captura-recaptura con dos o tres registros, y en ambos casos, la entrada de datos es manual. En el primer caso, el programa asume que las dos fuentes son independientes y estima el número total de casos de la población, con su intervalo de confianza. Además, estima la exhaustividad de cada una de las listas por separado y conjuntamente. Si se cuenta con tres fuentes de datos, Epidat 3.1 estima, mediante un modelo log-lineal, el número total de casos con su intervalo de confianza y la exhaustividad de los registros, bajo diferentes asunciones sobre la independencia de los mismos. Para cada uno de los modelos ajustados se presentan el estadístico G^2 de razón de verosimilitudes, el número de grados de libertad del modelo y el criterio de información bayesiano (BIC).

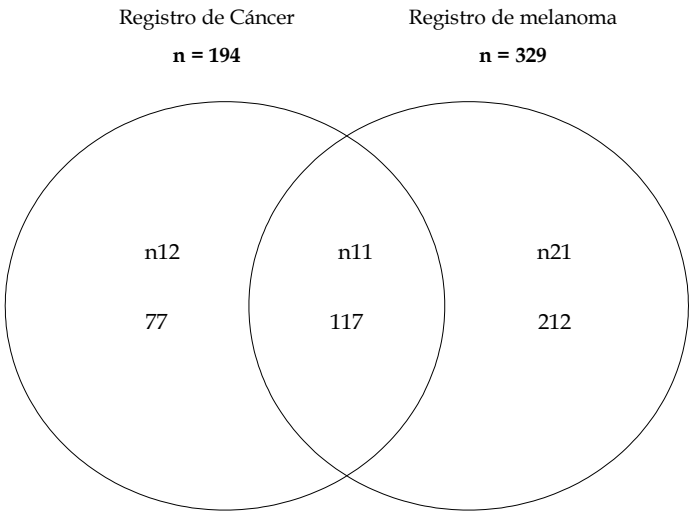
Ejemplos

Ejemplo 1: Dos listas

En el año 1998 se realizó en Gales un estudio con la finalidad de determinar la prevalencia de melanoma[8], para lo cual se utilizaron el Registro de cáncer de Gales y el Registro de melanoma. Se encontró un total de 406 casos, de los cuales 194 habían sido incluidos en el Registro de cáncer, 329 en el Registro de melanoma y 117 en ambos (Tabla 1). El objetivo fue determinar el número de personas con melanoma. La población en riesgo se estimó en 4,4 millones de personas.

Tabla 1. Casos de melanoma registrados en dos fuentes de información. Gales, 1998.

		Registro de melanoma (B)	
		Notificados	**No notificados**
Registro de cáncer (A)	**Notificados**	n_{11} = 117	n_{12} = 77
	No notificados	n_{21} = 212	$\mathbf{n_{22}}$



Resultados con Epidat 3.1

```
Captura-recaptura

Número de registros:     Dos
Nivel de confianza:  95,0%
```

```
                          Registro B
Registro A          Notificados   No notificados            Total
--------------- ---------------  ---------------  ---------------
Notificados                 117               77              194
No notificados              212              138              350
--------------- ---------------  ---------------  ---------------
Total                       329              215              544

Casos estimados       IC (95%)
--------------- ---------------------
            544       495   593

Exhaustividad                          %
------------------------------ ---------
Registro A                        35,64
Registro B                        60,44
Registros A y B                   74,59
```

El número total de personas afectadas de melanoma fue estimado en 544, con un intervalo de confianza de 495 a 593. Por tanto, la tasa de prevalencia de este tipo de cáncer estaría entre 11,3 y 13,5 por 100.000 habitantes, con una confianza del 95%. En cuanto a la exhaustividad de estos registros, puede observarse que el de melanoma detecta un 60% de los casos que se estiman, en tanto que el registro de cáncer sólo detecta el 36%; aún contando con los dos registros se pierden un 25% de los casos de melanoma.

Ejemplo 2: Tres listas

Entre abril y mayo de 1997 se realizó un estudio[5] para estimar la prevalencia de epilepsia en dos poblados de la región de Benin de Zinvié, los cuales cuentan con una población de 3.134 habitantes.

Se utilizaron las siguientes fuentes de datos:

- Registro A: fuentes no médicas (practicantes de medicina tradicional, profesores, líderes comunales y religiosos).
- Registro B: registros médicos en los centros de salud
- Registro C: estudio transversal casa por casa.

El total de casos registrados fue 66; 26 de ellos fueron identificados por fuentes no médicas, 4 por fuentes médicas y 50 en el estudio transversal casa por casa. En la Tabla 2 puede verse la distribución de estos casos al cruzar las tres fuentes.

Tabla 2. Casos de epilepsia registrados en tres fuentes de información.

Registro C: Notificados

		Registro B	
		Notificados	No notificados
Registro A	Notificados	0	12
	No notificados	1	37

Registro C: No notificados

		Registro B	
		Notificados	No notificados
Registro A	Notificados	1	13
	No notificados	2	

El análisis puede comenzar con una estimación del número de casos para cada par de listas. Si la estimación con uno de los modelos da un valor mucho más pequeño que los demás, puede asumirse que hay dependencia positiva entre las dos fuentes utilizadas.

Al hacer los cálculos con Epidat 3.1 se obtiene una estimación de 66 casos con los registros A y B, 104 con A y C, y 126 con B y C:

	Registro B		
Registro A	Notificados	No notificados	Total
Notificados	1	25	26
No notificados	3	37	40
Total	4	62	66

Casos estimados	IC (95%)	
66	10	123

	Registro C		
Registro A	Notificados	No notificados	Total
Notificados	12	14	26
No notificados	38	40	78
Total	50	54	104

Casos estimados	IC (95%)	
104	71	139

	Registro C		
Registro B	Notificados	No notificados	Total
Notificados	1	3	4
No notificados	49	73	122
Total	50	76	126

Casos estimados	IC (95%)	
126	17	236

Los resultados que Epidat 3.1 produce utilizando las tres fuentes son:

Captura-recaptura

Número de registros: Tres
Nivel de confianza: 95,0%

	Casos notificados
Registro A	26
Registro B	4
Registro C	50
Total	66

Hipótesis	X	N	IC (95,0%)	G^2	gl	BIC

A, B y C independientes	45	111	73	148	1,46	3	-5,59
A y B independientes de C	**45**	**111**	**73**	**150**	**1,46**	**2**	**-3,25**
A y C independientes de B	62	128	5	251	1,35	2	-3,36
B y C independientes de A	40	106	71	141	0,68	2	-4,02
A independiente de B	26	92	28	156	0,56	1	-1,80
A independiente de C	40	106	70	142	0,68	1	-1,67
B independiente de C	74	140	-39	320	1,27	1	-1,08
A, B y C dependientes	0	66	-	-	0,00	0	0,00

X: Estimación de los casos no notificados por ningún registro
N: Estimación del total de casos
G^2: Estadístico de razón de verosimilitudes
BIC: Criterio de información Bayesiano

Exhaustividad de los registros (%)

Hipótesis	A	B	C	A,B	A,C	B,C	A,B,C
A, B y C independientes	23,4	3,6	45,0	26,1	57,6	47,7	59,5
A y B independientes de C	**23,4**	**3,6**	**45,0**	**26,1**	**57,7**	**47,7**	**59,5**
A y C independientes de B	20,3	3,1	39,1	22,7	50,0	41,4	51,6
B y C independientes de A	24,5	3,8	47,2	27,4	60,4	50,0	62,3
A independiente de B	28,3	4,3	54,3	31,5	69,6	57,6	71,7
A independiente de C	24,5	3,8	47,2	27,4	60,4	50,0	62,3
B independiente de C	18,6	2,9	35,7	20,7	45,7	37,9	47,1
A, B y C dependientes	39,4	6,1	75,8	43,9	97,0	80,3	100,0

El primer modelo puede descartarse ya que no es matemáticamente plausible. Asimismo, los valores del BIC permiten descartar los cuatro últimos; para los restantes modelos se obtienen valores próximos, por lo que no está clara la elección entre ellos. El análisis preliminar de cada par de fuentes muestra evidencia de que los registros A y B pueden ser dependientes, ya que parecen infraestimar el número total de casos; con el modelo log-lineal que asume independencia entre A y B se llega a la misma conclusión. Por tanto, el modelo más adecuado parece ser el que asume que A y B son independientes de C, lo que coincide con el conocimiento que se tiene de estas fuentes. Teniendo en cuenta esto, el número estimado de epilépticos es de 111, con un intervalo de confianza de 73 a 150 personas, por lo que la tasa de prevalencia de epilepsia, en los dos poblados en estudio, varía entre 23,3 y 47,9 por 1.000 habitantes, con una confianza del 95%. La exhaustividad de los registros médicos es muy baja, solo detectan un 3,6% de los casos y aportan muy poco a las fuentes A y C, que pasan de un 57,7% de exhaustividad a un 59,5% al añadir el registro B.

Bibliografía

1. Schnabel ZE. The estimation of the total fish population of a lake. *Am Math Monthly* 1938; 45: 348-3.

2. Sekar CC, Deming WE. On a method of estimating birth and death rates and the extent of registration. *Am Stat Ass J* 1949; 44: 101-15.

3. Hook EB, Regal RR. Capture-recapture methods in epidemiology: methods and limitations. *Epidemiol Rev* 1995; 17: 243-64.

4. Ismail AA, Beeching NJ, Gill GV, Bellis MA. Capture-recapture-adjusted prevalence rates of type 2 diabetes are related to social deprivation. *Q J Med* 1999; 92: 707-10.

5. Debrock C, Preux PM, Houinato D, et al. Estimation of the prevalence of epilepsy in the Benin region of Zinvié using the capture-recapture methods. *Int J Epidemiol* 2000; 29: 330-5.

6. Robles SC, Marrett LD, Clarke EA, Rish HA. An application of capture-recapture methods to the estimation of completeness of cancer registration. *J Clin Epidemiol* 1988; 41: 495-501.

7. Schouten LJ, Straatman H, Kiemeney LA, Gimbrere CH, Verbeek AL. The capture-recapture method for estimation of cancer registry completeness: a useful tool?. *Int J Epidemiol* 1994; 23: 1111-6.

8. Paterson IC, Beer H, Adams Jones D. Under-registration of melanoma in Wales in 1998: use of the capture-recapture method to estimate the "true" incidence. *Melanoma Res* 2001; 11: 141-5.

9. Taub NA, Lemic-Stojcevic N, Wolfe CD. Capture-recapture methods for precise measurement of the incidence and prevalence of stroke. *J Neurol Neuroserg Psychiatr* 1996; 60: 696-7.

10. Carolei A, Marine C, Di Napoli M, et al. High stroke incidence in the prospective community-based L'Aquila registry (1994-1998). First year's results. *Stroke* 1997; 28: 2500-6.

11. Fisher N, Turner SW, Pugh R, Taylor C. Estimating numbers of homeless and homeless mentally ill people in north east Westminster by using capture-recapture analysis. *Br Med J* 1994; 308: 27-30.

12. Frischer M, Bloor M, Finlay A, et al. A new method of estimating prevalence of injecting drug use in an urban population: results from a Scottish city. *Int J Epidemiol* 1991; 20: 997-1000.

13. Gallay A, Nardone A, Vaillant V, Desenclos JC. La méthode capture-recapture appliquée à l'épidémiologie: principles, limites et applications. *Rev Epidemiol Sante Publique* 2002; 50: 219-32.

14. Cormack RM, Chang YF, Smith GS. Estimating death from industrial injury by capture-recapture: a cautionary tale. *Int J Epidemiol* 2000; 29: 1053-9.

DETECCIÓN DE CLUSTERS

Conceptos generales

Una de las tareas mas importantes de la Vigilancia Epidemiológica es el análisis del comportamiento y tendencias de las enfermedades y/o daños sujetos a vigilancia y, dentro de ello, el análisis de los patrones epidémicos observados, que permitan la estimación numérica de parámetros de importancia médica, e incluso el tamaño de la población afectada.

Sin embargo, con la mejora de las condiciones de saneamiento, la evolución de los sistemas y servicios de salud, el descubrimiento de nuevos y más eficaces agentes terapéuticos y vacunas, así como de los métodos de control de enfermedades, la capacidad de expansión de las epidemias de enfermedades transmisibles de mayor prevalencia ha disminuido con el tiempo. En contraste, asistimos al desarrollo de enfermedades degenerativas, a los daños derivados de la contaminación y agentes tóxicos presentes en el ambiente, y a la emergencia de nuevas enfermedades transmisibles, condiciones de las que, por lo general, no se puede disponer de registros continuos de casos debido a su reciente aparición, su presentación esporádica y, en ocasiones, su localización en áreas geográficas muy pequeñas.

Una de las necesidades que plantea el comportamiento de estas entidades es el perfeccionamiento de los sistemas de vigilancia, de manera que puedan identificar cuando una agregación de casos de enfermos observada en un área geográfica determinada, en un período de tiempo limitado, o teniendo en cuenta ambos escenarios a la vez, es superior a lo esperado, y si ello representa un brote. Las técnicas estadísticas pueden ayudar a los epidemiólogos a reconocer si la agrupación de casos fuera de lo usual es estadísticamente significativa, y puede corresponder al inicio de una epidemia.

La aplicación de estos métodos a la epidemiología se debe a la capacidad de detección de casos especialmente próximos, por encima de lo esperado, que se denominan *clusters* o conglomerados. De esta forma, es posible la investigación de eventos raros, ligados a contaminación ambiental o a transmisibles[1]. Ejemplos clásicos son los estudios de Leucemia próximos a centrales nucleares, síndrome de muerte súbita infantil[2] y enfermedad de Hodgkin[3-6]. Desde el inicio de los años 1960 el Centro de Control de Enfermedades de los Estados Unidos (CDC) investigó, según Caldwell[7], más de 108 ocurrencias de conglomerados en cáncer, aplicando diversos métodos.

Históricamente, las investigaciones de conglomerados llevaron a diversos descubrimientos que mejoraron la comprensión tanto de dolencias transmisibles como de las crónicas degenerativas. Como ejemplo reciente de investigación de conglomerados de enfermedades infecciosas se encuentran la de los brotes de Pneumocistitis carinii[8] y sarcoma de Kaposi[9] en hombres jóvenes; entre las dolencias no infecciosas puede citarse la ocurrencia de la enfermedad de Minamata relacionada con la contaminación por mercurio en Japón[10].

El proceso de evaluación de un conglomerado

Tres criterios deben tenerse en cuenta para evaluar los conglomerados: la consistencia, la plausibilidad biológica y el efecto dosis-respuesta.

La *consistencia*, en términos de las investigaciones de conglomerados, puede referirse a la observación de un patrón histórico en los casos notificados, o su compatibilidad con datos de la literatura; sin embargo, la consistencia se refiere generalmente a la homogeneidad de los casos. Los atributos personales, como edad, raza, sexo y ocupación, pueden reforzar un resultado estadístico sugestivo; además, ciertas características específicas de la enfermedad (como la estirpe celular en los tumores, las vías de exposición, la localización anatómica, ...) pueden ser importantes evidencias de consistencia. Dado que los eventos raros de salud se

estudian mejor en conglomerados, la uniformidad dentro de la agregación aporta validez a las evidencias e interpretación estadística.

La *plausibilidad biológica* es siempre un aspecto importante en la interpretación epidemiológica de los datos. La presencia de un riesgo ambiental reconocido, o un factor de exposición importante, proporcionan plausibilidad biológica para una investigación; sin embargo, ésta no puede ser una base para descartar un conglomerado, aunque sí un criterio para reforzar y/o interpretar las pruebas estadísticas.

El *efecto dosis-respuesta* es uno de los criterios más fuertes para evaluar los datos epidemiológicos y en muchas ocasiones forma la base para detectar relaciones de causalidad.

Podemos definir el análisis de conglomerados como un conjunto de procedimientos que buscan agrupar y discriminar grupos de individuos, regiones u objetos. Estas agregaciones se detectan mediante criterios basados en distancias, es decir, medidas matemáticas de similitud geográfica, temporal o basada en cualquier característica del objeto. Sin embargo, en la evaluación de un presunto conglomerado no debe dependerse exclusivamente de los datos estadísticos, sino que debe evaluarse también la consistencia, la plausibilidad biológica y el efecto dosis-respuesta.

Epidat incluye, en el módulo de Vigilancia en Salud Pública, un submódulo de detección de clusters basado en el programa de análisis epidemiológico CLUSTER 3.1, desarrollado por el CDC (Centres for Disease Control)[11]. Dicho submódulo ofrece una serie de técnicas útiles para la detección de agregaciones temporales, espaciales y espacio-temporales:

Para la detección de agregaciones temporales:

Casos individuales:

- Método Chen
- Método Sets

Casos agrupados:

- Método Scan
- Método Texas
- Método Poisson
- Método Cusum

Para la detección de agregaciones espaciales:

- Método Grimson
- Método Ohno

Para la detección de agregaciones espacio-temporales:

- Método Knox

Para la detección de otras agregaciones:

- Método Remsa

AGREGACIONES TEMPORALES

En este tipo de agregaciones el investigador busca decidir si el número o proporción de casos de una enfermedad determinada, que aparece en intervalos de tiempo consecutivos, suceden con una frecuencia diferente a la esperada si se tratara de una distribución aleatoria. Las técnicas de agregaciones temporales pretenden dar una respuesta acertada, especialmente cuando los datos disponibles se refieren a series cortas de tiempo, que no pueden ser analizadas usando los métodos convencionales; en Epidat se incluyen 6 de estos métodos.

Chen y Sets, que se aplican a casos individuales, y Scan, Texas, Poisson y Cusum para series temporales de casos.

Casos individuales: método CHEN

El test de Chen[12], que permite detectar agregaciones de casos en el tiempo, intenta responder a la pregunta ¿están ocurriendo todos los casos consecutivos de una enfermedad, registrados en una cierta zona geográfica, en un intervalo crítico de tiempo?

Para contrastar la hipótesis nula de que no existe agregación temporal en los casos observados, los periodos de tiempo que transcurren entre casos consecutivos se comparan con una longitud crítica, que se determina a partir de la tasa de incidencia de la enfermedad y el tamaño de la población en riesgo. Si cada uno de los intervalos observados entre los casos es más corto que la longitud critica, se puede concluir que existe evidencia de agregación temporal.

Para la aplicación de esta técnica, se necesita la siguiente información:

- La serie de casos con sus respectivas fechas de ocurrencia.
- T: la tasa esperada de incidencia anual de la enfermedad (puede ser una tasa histórica o del pasado reciente que ha prevalecido hasta antes de la sospecha de epidemia).
- P: el tamaño de la población a riesgo.
- FP: la tasa de falsos positivos, es decir, la probabilidad de rechazar la hipótesis nula siendo cierta, por motivo de la aparición de casos falsos de la enfermedad en el período estudiado. Es el nivel de significación de la prueba.
- Fecha inicial: debe especificarse como fecha de inicio una fecha anterior al primer caso considerado; puede ser, por ejemplo, la fecha de ocurrencia de un caso no perteneciente al brote en estudio.

Con estos datos se calcula la longitud temporal crítica o intervalo de referencia, que viene dado por:

$$LC = -\log\left(1 - FP\right)^{1/N} LE$$

donde N es el número total de casos y LE es la longitud esperada del intervalo entre casos consecutivos (en meses):

$$LE = \frac{12}{\textit{número esperado de casos anuales}} = \frac{12}{P \times T}$$

Si cada uno de los intervalos de tiempo entre casos consecutivos es menor que LC, entonces se rechaza la hipótesis de que no existe agregación temporal en los casos observados y se declara, por tanto, la evidencia de epidemia con una significación del FP%.

Limitaciones del método Chen

- Este método está diseñado para detectar sólo agregaciones temporales de casos de una enfermedad.
- Para la aplicación del método Chen es necesario conocer el tamaño de la población a riesgo y la tasa de incidencia "basal" en dicha población, y ésta puede ser difícil de determinar en comunidades pequeñas.
- Las tasas de incidencia de enfermedades de muy baja frecuencia, así como en áreas muy pequeñas, pueden ser muy inestables.

- Es necesario "restringir" los casos a aquellos que se considera que forman una agregación, es decir, en los que hay sospecha de brote.

Ejercicio

En una comunidad de 20.000 habitantes se produjeron durante el año 1999 11 casos de una enfermedad con incidencia esperada de 12 casos por 100.000, y el último caso del año 1998 ocurrió el 10 de noviembre, ¿hay evidencia de agregación temporal? El archivo CHEN-SETS.xls que se distribuye con Epidat contiene la serie de casos con sus fechas de ocurrencia, según se muestra en la Tabla 1.

Tabla 1. Casos observados con sus respectivas fechas de ocurrencia.

CASO	FECHA
1	11/01/99
2	21/01/99
3	14/02/99
4	06/04/99
5	22/04/99
6	30/04/99
7	29/05/99
8	01/06/99
9	11/06/99
10	12/07/99
11	15/08/99

Manejo del submódulo y solución del ejercicio.-

Los datos pueden introducirse desde el teclado (entrada manual) o importarse en formato Dbase, Access o Excel (entrada automática). Si la entrada es manual, basta con informar del número de casos en la población de estudio y el programa ya genera una tabla con tantas filas como número de casos en la que el usuario sólo tiene que rellenar las fechas de ocurrencia correspondientes a cada uno de los casos. Si la entrada es automática, Epidat necesita que la tabla de datos tenga una estructura determinada, con una variable que identifique la fecha de ocurrencia de cada uno de los casos (véase la Tabla 1).

Resultados con Epidat 3.1

```
Vigilancia: Detección de clusters, agregaciones temporales

Archivo de trabajo: C:\Archivos de programa \Epidat 3.1 \Ejemplos
\Vigilancia \CHEN-SETS.xls

Campo que contiene:
  Fecha: FECHA

Fecha inicial                    : 10/11/1998
Tamaño de la población a riesgo  : 20000
Tasa esperada de incidencia anual: 12
Tasa de falsos positivos (%)  : 5
Método                         : Chen


  Caso       Fecha  Meses entre casos consecutivos
------  ----------  ------------------------------
```

```
     1  11/01/1999      2,0370
     2  21/01/1999      0,3285
     3  14/02/1999      0,7885
     4  06/04/1999      1,6756
     5  22/04/1999      0,5257
     6  30/04/1999      0,2628
     7  29/05/1999      0,9528
     8  01/06/1999      0,0986
     9  11/06/1999      0,3285
    10  12/07/1999      1,0185
    11  15/08/1999      1,1171

Longitud en meses del periodo de estudio: 9,13
Longitud temporal crítica entre casos:  7,1689 Meses
Hay evidencia de epidemia con significación del 5%
```

Puesto que cada uno de los intervalos temporales (meses) entre casos consecutivos es menor que la longitud crítica LC = 7,17 meses, hay evidencia de que la incidencia de la enfermedad considerada ha experimentado un aumento en la comunidad de estudio.

Casos individuales: método SETS

La técnica Sets fue diseñada para estudiar y controlar la incidencia de malformaciones congénitas[13-15] y de otras enfermedades raras, como ciertos tipos de cáncer[16]. El método se basa en el concepto de *Set*. Un *Set* es el intervalo de tiempo entre dos casos consecutivos de una enfermedad o, en general, entre dos eventos consecutivos. Si cada *Set* de una secuencia de casos es menor que cierto valor de referencia, entonces se declara una alarma para indicar la existencia de una agregación de casos no aleatoria y significativa en el área de estudio.

Para aplicar el método Sets se necesitan los siguientes datos:

- La serie de casos con sus respectivas fechas de ocurrencia.
- T: la tasa esperada de incidencia anual de la enfermedad (puede ser una tasa histórica o del pasado reciente que ha prevalecido hasta antes de la sospecha de epidemia).
- P: el tamaño de la población a riesgo.
- A: el número de años entre falsos positivos (al aumentar A se reduce la probabilidad de rechazar la hipótesis nula siendo cierta por motivo de la aparición de casos falsos de la enfermedad en el período estudiado).
- Fecha inicial: debe especificarse como fecha de inicio una fecha anterior al primer caso considerado; puede ser, por ejemplo, la fecha de ocurrencia de un caso no perteneciente al brote en estudio.

El análisis, que se realiza para cada secuencia de n casos, se basa en comparar los Sets de la secuencia con el intervalo o valor de referencia, que se determina como un múltiplo del tiempo esperado entre casos consecutivos (inverso del número esperado de casos anuales):

$$R = \frac{k}{\textit{número esperado de casos anuales}} = \frac{k}{P \times T}$$

Los valores n y k son dos parámetros asociados al método y que definen, respectivamente, el tamaño de la secuencia de casos (número de casos consecutivos a utilizar en cada análisis) y la constante para calcular el valor mínimo del *Set* o intervalo de referencia. Estos valores fueron calculados por Chen utilizando una ecuación aproximada primero[16] y aplicando un método exacto posteriormente[18], y tabulados en función del número esperado de casos entre dos falsos positivos:

$$M = P \times T \times A$$

A partir de estos datos, el procedimiento consiste en lo siguiente:

- El primer conjunto o secuencia se conforma con los *n* primeros casos y se analiza si cada *Set* es menor que el valor crítico R; de ser así, se declara una alarma y se pasa a analizar el siguiente conjunto de *n* elementos, que no tendrá casos comunes con el anterior.
- Por el contrario, si en la secuencia inicial de *n* casos al menos un *Set* es mayor que R, se concluye que no hay evidencia de agregación temporal en esa secuencia y se pasa a la siguiente, que ahora estará formada por los casos desde el segundo hasta el *n*+1.

Limitaciones del método Sets

A pesar de que el método Sets es una de las técnicas más empleadas en la detección de agregaciones temporales, y puede incluso modificarse para vigilar varias comunidades a la vez, también presenta algunas limitaciones:

- Para la aplicación del método es necesario conocer el tamaño de la población a riesgo y la tasa de incidencia "basal" en dicha población, y ésta puede ser difícil de determinar en comunidades pequeñas.
- Los resultados obtenidos y la sensibilidad del método dependen, en gran medida, del número de años entre falsos positivos (A) y de la tasa de incidencia anual que, en el caso de enfermedades de muy baja frecuencia, así como ENáreas muy pequeñas, puede ser muy inestable.
- El método asume que la población a riesgo es estable.

Recomendaciones

Debe tenerse en cuenta para este análisis que el parámetro A se selecciona con cierto grado de subjetividad, por lo que son necesarias algunas recomendaciones. Así, cuando se analizan períodos de tiempo largos y varias enfermedades, es preferible tomar un valor de A elevado; por el contrario, si se pretende estudiar una sola enfermedad debe utilizarse un valor menor, pues contribuye a elevar la potencia para la detección de agregaciones.

Ejercicio

En el ejercicio del método Chen (Tabla 1), consideremos que el número de años entre dos falsas alarmas es A=35, ¿qué se obtiene al aplicar el método Sets?

Manejo del submódulo y solución del ejercicio.-

Los datos pueden introducirse desde el teclado (entrada manual) o importarse en formato Dbase, Access o Excel (entrada automática). Si la entrada es manual, basta con informar del número de casos en la población de estudio y el programa ya genera una tabla con tantas filas como número de casos en la que el usuario debe rellenar las fechas de ocurrencia correspondientes a cada caso. Si la entrada es automática, Epidat necesita que éstos tengan una estructura determinada, con una variable que identifique la fecha de ocurrencia de cada uno de los casos (véase la Tabla 1).

Resultados con Epidat 3.1

```
Vigilancia: Detección de clusters, agregaciones temporales

Archivo de trabajo: C:\Archivos de programa \Epidat 3.1 \Ejemplos
\Vigilancia \CHEN-SETS.xls

Campo que contiene:
  Fecha: FECHA

Fecha inicial                    : 10/11/98
Población a riesgo               : 20000
Tasa esperada de incidencia anual: 12
N° de años entre falsos positivos: 35
Método                           : Sets

  Caso       Fecha  Meses entre casos consecutivos
------  ----------  ------------------------------
     1  11/01/1999     2,0370
     2  21/01/1999     0,3285
     3  14/02/1999     0,7885
     4  06/04/1999     1,6756
     5  22/04/1999     0,5257
     6  30/04/1999     0,2628
     7  29/05/1999     0,9528
     8  01/06/1999     0,0986
     9  11/06/1999     0,3285
    10  12/07/1999     1,0185
    11  15/08/1999     1,1171

Longitud en meses del periodo de estudio: 9,13
Longitud temporal crítica entre casos:  3,9050 Meses
Tamaño del cluster: 6
Hay evidencia de 1 cluster

Cluster 1:

  Caso       Fecha  Años entre casos consecutivos
------  ----------  ------------------------------
     1  11/01/1999     2,0370
     2  21/01/1999     0,3285
     3  14/02/1999     0,7885
     4  06/04/1999     1,6756
     5  22/04/1999     0,5257
     6  30/04/1999     0,2628
```

Como entre los 6 primeros casos todos los *sets* son menores que el valor crítico (3,9 meses), hay evidencia de agregación temporal en esa secuencia de casos. A continuación, se analizaría el siguiente conjunto de 6 elementos, que empezaría en el caso número 7, pero al no disponer de otra secuencia completa de 6 casos a partir de éste (sólo hay 11 casos), entonces se para el proceso.

Casos agrupados: Método Scan

El test Scan, desarrollado por Nauss[17-18], es una prueba estadística para detectar agrupaciones temporales en una serie temporal de casos de un cierto evento o enfermedad. La hipótesis nula establece que los casos ocurren al azar en el tiempo, frente a la alternativa de que los casos se agrupan en ciertos períodos. Es un método útil para detectar "picos" en la incidencia de una enfermedad y, al no tener una componente espacial, es adecuado para estudios de enfermedades en los que la población se distribuye de forma desigual en el área de estudio.

El estadístico Scan, en el que se basa la prueba, es el número máximo de casos observados en una ventana móvil de longitud fija, que se mueve de forma continua a lo largo del tiempo. Si los casos se agrupan en el tiempo, entonces el número máximo de casos en la ventana temporal será grande. El ancho de la ventana se basa en la duración esperada de una epidemia.

Los datos necesarios para aplicar la prueba son:

- La serie temporal de casos observados (en días, semanas, cuatrisemanas, meses o años): X_i, i=1, ..., T.
- w: la amplitud de la ventana, expresada en las mismas unidades de tiempo.

Si $X_{t,t+w}$ denota el número de casos en el período (t, t+w], el estadístico Scan se expresa como:

$$S_w = \max_{0 \le t \le T-w} Y_{t,t+w}$$

Se rechaza la hipótesis nula, al nivel de significación α, si la probabilidad de que el estadístico sea mayor que el valor observado en la serie es mayor o igual que α:

$$Valor\ p = \Pr(S_w \ge n) \ge \alpha$$

Este valor p depende de 3 datos:

- El número total de casos observados (N).
- La amplitud relativa de la ventana respecto al período total de estudio (r=w/T).
- El número máximo de casos observados en todas las ventanas temporales de amplitud w (n).

Se denota como $\Pr(n, N, r)$. Matemáticamente, es la probabilidad de que si N puntos están distribuidos aleatoriamente a lo largo de una línea de longitud 1, exista un intervalo de longitud r que contenga al menos n de ellos. Como la distribución del estadístico Scan bajo la hipótesis nula es desconocida, el valor p debe calcularse con aproximaciones o por simulación; Epidat utiliza una aproximación basada en la distribución binomial[19].

Limitaciones del método Scan

- El método Scan asume que la población a riesgo permanece constante a lo largo de todo el período de estudio, al igual que la probabilidad de detección de casos, que también se supone constante. Por tanto, el método no es válido en situaciones en las que haya algún cambio en la población o en los métodos diagnósticos de la enfermedad.
- No puede distinguir entre un cluster debido a un cambio en el patrón de los factores de riesgo conocidos de la enfermedad, y un cluster debido a otras causas.
- Hay que tener especial cuidado en la selección de la ventana temporal, pues dicha elección determina los resultados obtenidos. Cuando no se conoce la duración esperada de la supuesta epidemia, conviene aplicar varias veces el test utilizando diferentes amplitudes de ventana.
- El procedimiento no es útil cuando el número de casos detectados es extremadamente bajo (<10).

Ejercicio

En un período de 28 años, entre 1975 y 2002, se notifican 255 casos de cierto tipo de cáncer en una región. La serie anual, que puede verse en la Tabla 2, se encuentra en el archivo SCAN.xls de los ejemplos de Epidat. En el período de estudio, ¿hay evidencia de alguna agregación de casos en el lapso de un año? ¿y en 3 años?

Tabla 2. Serie anual de casos observados de 1975 a 2002.

AÑO	CASOS
1975	15
1976	12
1977	13
1978	11
1979	15
1980	12
1981	9
1982	12
1983	8
1984	4

AÑO	CASOS
1985	17
1986	13
1987	15
1988	12
1989	9
1990	12
1991	6
1992	7
1993	8
1994	4

AÑO	CASOS
1995	6
1996	5
1997	3
1998	7
1999	8
2000	2
2001	4
2002	6

Manejo del submódulo y solución del ejercicio.-

Los datos pueden introducirse desde el teclado (entrada manual) o importarse en formato Dbase, Excel o Access (entrada automática). En ambos casos, el usuario debe seleccionar en primer lugar la unidad de medida del tiempo entre días, semanas, cuatrisemanas, meses y años. A continuación, si la entrada es manual, se introducen el período inicial (por ejemplo, el mes y año cuando la serie de casos es mensual) y el número total de períodos (en ese caso, número de meses); entonces, el programa genera una tabla con tantas filas como períodos y sólo hay que introducir el número de casos en cada fila.

Si la entrada es automática, Epidat 3.1 necesita que la tabla de datos tenga una estructura determinada, con una variable que contenga la serie temporal de casos observados, y una o dos variables que identifiquen el período temporal de ocurrencia, según la unidad de medida seleccionada (días, semanas, cuatrisemanas, meses o años). Por ejemplo, si el período elegido es días, sólo se necesita una variable temporal que contenga la fecha de ocurrencia de los casos. Si la serie de casos es semanal, mensual o cuatrisemanal, la tabla debe contener una variable para identificar el año y otra para el número de semana, mes o cuatrisemana, respectivamente. Por último, en el caso de series anuales, sólo es necesario el año de ocurrencia (como ejemplo, véase la Tabla 3).

Para analizar si hay agregaciones anuales de casos debe tomarse una ventana de amplitud 1. Para ese valor, el número máximo de casos en una ventana de amplitud 1 es el máximo de los casos anuales, que corresponde a los 17 casos del año 1985 (Tabla 2).

Tabla 3. Formato de tabla preparada para importar datos desde Epidat 3.1 para la aplicación del método Scan.

AÑO	CASOS
1975	15

1976	12
1977	13
1978	11
...	...
2001	4
2002	6

Resultados con Epidat 3.1 para una ventana de amplitud 1 año:

```
Vigilancia: Detección de clusters, agregaciones temporales


Archivo de trabajo: C:\Archivos de programa \Epidat 3.1 \Ejemplos
\Vigilancia \SCAN.xls

Campo que identifica:
  Casos observados: CASOS
  Años: AÑO

Método         : Scan
Tipo de período: Años
Ventana temporal en Años: 1

N° máximo de casos en la ventana temporal: 17
Probabilidad {Casos esperados >= 17} =  0,7699

 Intervalo 1
       Año      Casos
---------- ----------
      1985         17
---------- ----------
     Total         17
```

En este caso, como la probabilidad de observar un número de casos igual o mayor que n = 17 es tan elevada (0,7699), no hay evidencias para rechazar la hipótesis nula con un nivel de significación del 5% y, por tanto, se admite que no existe ninguna agrupación anual de casos.

Si tomáramos una ventana de longitud 3 años, los resultados serían diferentes. En primer lugar, habría que calcular el número de casos en todas las posibles ventanas de 3 años:

Ventana de 3 años			Casos
1975	1976	1977	40
1976	1977	1978	36
1977	1978	1979	39
1978	1979	1980	38
1979	1980	1981	36
1980	1981	1982	33
...	...	...	...
2000	2001	2002	12

Resultados con Epidat 3.1 para una ventana de amplitud 3 años:

```
Vigilancia: Detección de clusters, agregaciones temporales

Archivo de trabajo: C:\Archivos de programa \Epidat 3.1 \Ejemplos
\Vigilancia \SCAN.xls

Campo que identifica:
  Casos observados: CASOS
```

```
  Años: AÑO

Método         : Scan
Tipo de período: Años
Ventana temporal en Años: 3

N° máximo de casos en la ventana temporal: 45
Probabilidad {Casos esperados >= 45} =  0,0457

 Intervalo 1
       Año      Casos
---------- ----------
      1985         17
      1986         13
      1987         15
---------- ----------
     Total         45
```

Ahora, en cada intervalo de 3 años, comenzando en 1975, se busca la mayor agrupación de casos que, en efecto, se encuentra en la ventana formada por los años 1985, 1986 y 1987. La probabilidad de observar 45 casos o más en tres años es un valor pequeño (p=0,0457 < 0,05), por lo que hay evidencia estadística de que la agrupación observada no es producto de la mera casualidad.

Hay que tener en cuenta que muchas veces se desconoce la duración esperada de la enfermedad, de modo que en la selección de la amplitud de la ventana puede haber un grado de subjetividad elevado; por esta razón, es aconsejable probar con varios valores.

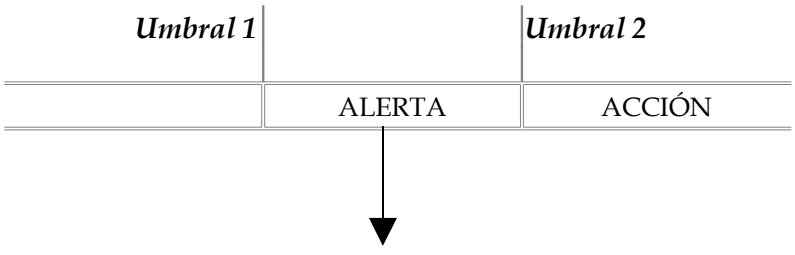
Casos agrupados: Método Texas

El método Texas es un procedimiento estadístico que permite la detección de agregaciones temporales en la incidencia de una enfermedad de baja frecuencia y ha sido desarrollado en el contexto de una comunidad potencialmente expuesta a una fuente contaminante[20]. Por ejemplo, ha sido utilizado para el estudio de efectos adversos sobre la salud provocados por emisiones tóxicas de industrias locales[21].

Esta metodología se basa en una regla de decisión, en dos etapas, que utiliza la razón estandarizada de mortalidad (RME o SMR en inglés) para identificar un aumento significativo en la mortalidad o morbilidad de una región con respecto a lo esperado. El método establece dos valores umbrales (*umbral1* < *umbral2*), que definen una *fase de alerta* cuando los datos indican un posible problema de salud, o una *fase de actuación* cuando hay evidencias de un problema de salud grave o prolongado y es necesario intervenir.

La regla de decisión para definir los estados de alerta o actuación es la siguiente: si se supera el *umbral2* se declara el estado de actuación, es decir, se identifica una agregación temporal inusual; si ese nivel no es excedido pero sí el *umbral1*, entonces se declara el estado de alerta. Además, dos alertas consecutivas definen también un estado de actuación (Figura 1).

Figura 1.- Esquema de la regla de decisión en el método TEXAS.



2 alertas consecutivas

ACCIÓN

Los datos necesarios para aplicar el método son:

- La serie temporal de casos observados (en semanas, cuatrisemanas, meses o años): O_i.
- La serie temporal de casos esperados (en semanas, cuatrisemanas, meses o años): E_i.
- P_1: la probabilidad de alerta.
- P_2: La probabilidad de acción ($P_2 < P_1$).

El cálculo de los valores umbrales y la regla de decisión se basan en la distribución de Poisson si el número de casos esperado es menor o igual que 30, o en una aproximación a la distribución normal en caso contrario.

Si $E_i > 30$, el procedimiento tiene los siguientes pasos:

1. Se calculan los valores umbrales de cada período:

$$Umbral1 = 1 + \frac{\phi^{-1}(1 - P_1)}{\sqrt{E_i}} \qquad Umbral2 = 1 + \frac{\phi^{-1}(1 - P_2)}{\sqrt{E_i}}$$

   donde ϕ es la función de distribución de la Normal estándar.

2. Se calcula el SMR de cada período como el cociente entre el número de casos observado y el esperado.
3. Se aplica la siguiente regla:

- ALERTA: $Umbral1 < SMR_i < Umbral2$
- ACTUACIÓN:
  - $SMR_i \geq Umbral2$
  - $Umbral1 < SMR_i < Umbral2$ y $Umbral1 < SMR_{i-1} < Umbral2$ (2 alertas consecutivas)

Por el contrario, si $E_i \leq 30$, los pasos son:

1. Se calculan los valores umbrales:

$$Umbral1 = P_1 \qquad Umbral2 = P_2$$

2. Se calcula, para cada período, la probabilidad de observar al menos los O_i casos observados a partir de una distribución de Poisson de media E_i:

$$P_i = \Pr[Poisson(E_i) \geq O_i]$$

3. Se aplica la siguiente regla:

- ALERTA: $Umbral1 < P_i < Umbral2$
- ACTUACIÓN:
  - $P_i \geq$ Umbral2
  - $Umbral1 < P_i < Umbral2$ y $Umbral1 < P_{i-1} < Umbral2$ (2 alertas consecutivas)

Limitaciones del método Texas

- Para la estimación del número de casos esperado (generalmente se ajusta por edad, sexo, raza, etc) se requiere conocer las características demográficas de la población. También es necesario conocer las tasas específicas de mortalidad de una población de referencia y ésta última determina los resultados del método.
- La razón estandarizada de mortalidad es apropiada para determinar los estados de alerta o actuación si la población que se controla es pequeña.
- La determinación de las probabilidades de alerta y actuación, P_1 y P_2 respectivamente, es arbitraria.

Ejercicio

En una población cercana a una fuente contaminante se establece un sistema de vigilancia de abortos espontáneos. Los casos observados mensualmente durante el año 1999 se presentan en la Tabla 4, y se estima que el número esperado de abortos espontáneos al mes es de 10 casos. El ejemplo se ha tomado del manual del CLUSTER[13], y los datos se encuentran en el archivo TEXAS.xls distribuido con Epidat 3.1. ¿En qué meses se producen alertas si se fijan las probabilidades de alerta y actuación en el 5% y 1%, respectivamente?

Manejo del submódulo y solución del ejercicio.-

Los datos pueden introducirse desde el teclado (entrada manual) o importarse en formato Dbase, Excel o Access (entrada automática). En ambos casos, el usuario debe seleccionar en primer lugar la unidad de medida del tiempo entre semanas, cuatrisemanas, meses y años. A continuación, si la entrada es manual, se introducen el período inicial (por ejemplo, el mes y año cuando la serie de casos es mensual) y el número total de períodos (en ese caso, número de meses); entonces, el programa genera una tabla con tantas filas como períodos y sólo hay que introducir el número de casos observados y esperados en cada período.

Si la entrada es automática, Epidat 3.1 necesita que la tabla de datos tenga una estructura determinada, con dos variables que contengan, respectivamente, las series de casos observados y esperados, y una o dos variables que identifiquen el período temporal de ocurrencia, según la unidad de medida seleccionada (semanas, cuatrisemanas, meses o años). Por ejemplo, si la serie de casos es semanal, mensual o cuatrisemanal, la tabla debe contener una variable para identificar el año y otra para el número de semana, mes o cuatrisemana, respectivamente (como ejemplo, véase la Tabla 4); en el caso de series anuales, sólo es necesario el año de ocurrencia.

En cualquiera de los dos casos deben introducirse, además, las probabilidades de alerta y actuación, en porcentaje, que en el ejemplo son 5% y 1%, respectivamente.

Tabla 4. Casos de abortos espontáneos observados y esperados mensualmente durante el año 1999.

AÑO	MES	OBSERVADOS	ESPERADOS
1999	1	8	10
1999	2	10	10
1999	3	9	10
1999	4	11	10
1999	5	14	10
1999	6	16	10
1999	7	15	10
1999	8	13	10
1999	9	17	10
1999	10	14	10
1999	11	23	10
1999	12	13	10

Resultados con Epidat 3.1

```
Vigilancia: Detección de clusters, agregaciones temporales

Archivo de trabajo: C:\Archivos de programa \Epidat 3.1 \Ejemplos
\Vigilancia \TEXAS.xls

Campo que identifica:
   Casos observados: OBSERVADOS
   Casos esperados: ESPERADOS
  Meses: MES
  Años: AÑO

Método                        : Texas
Tipo de período               : Meses
Probabilidad de alerta (en %): 5
Probabilidad de acción (en %): 1

       Mes        SMR      Estado
---------- ----------  ----------
    1-1999     0,8000      ---
    2-1999     1,0000      ---
    3-1999     0,9000      ---
    4-1999     1,1000      ---
    5-1999     1,4000      ---
    6-1999     1,6000      Alerta
    7-1999     1,5000      ---
    8-1999     1,3000      ---
    9-1999     1,7000      Alerta
   10-1999     1,4000      ---
   11-1999     2,3000      Acción
   12-1999     1,3000      ---
```

A la vista de los resultados anteriores, se declara el estado de *alerta* en los meses de junio y septiembre, y el estado de *acción* en el mes de noviembre.

Casos agrupados: Método Poisson

El método de Poisson, como su nombre indica, está basado en la distribución de Poisson y permite detectar agregaciones de casos en el tiempo. Es un procedimiento muy sencillo que ha sido aplicado, por ejemplo, en la vigilancia de malformaciones congénitas en nacidos vivos[22].

Los datos necesarios para la aplicación de esta técnica son los siguientes:

- La serie temporal de casos observados (en semanas, cuatrisemanas, meses o años): O_i.
- La serie temporal de casos esperados (en semanas, cuatrisemanas, meses o años): E_i.

Suponiendo que el número de casos observados en el período i-ésimo (O_i) sigue una distribución de Poisson con media igual al número esperado de casos (E_i), este método basa su regla de decisión en la probabilidad de observar O_i casos o más:

$$P_i = \Pr[Poisson(E_i) \geq O_i]$$

Si esta probabilidad es pequeña (menor que el nivel de significación de la prueba) se rechaza la hipótesis nula de que no existe agregación de casos en el período considerado.

Limitaciones del método Poisson

- No siempre se puede asumir que las enfermedades siguen una distribución de Poisson. Si, por ejemplo, el área de estudio es pequeña, o la enfermedad es poco frecuente, puede producirse lo que se denomina sobredispersión cuando la varianza es mayor que la media (en una distribución de Poisson son iguales). En esos casos, distribuciones como la binomial negativa pueden ser más adecuadas.

Ejercicio

Durante el año 1987 se diagnostican 105 casos de cierto tipo de cáncer en una región. La serie mensual de casos observados y esperados, que puede verse en la Tabla 5, se encuentra en el archivo POISSON.xls de los ejemplos de Epidat. En el período de estudio, ¿hay evidencia de agregación de casos en algún mes?

Tabla 5. Casos de cáncer observados y esperados mensualmente durante el año 1987.

AÑO	MES	OBSERVADOS	ESPERADOS
1987	1	11	10
1987	2	8	10
1987	3	9	10
1987	4	14	10
1987	5	19	10
1987	6	11	10
1987	7	9	8
1987	8	8	8
1987	9	7	8
1987	10	4	8
1987	11	0	8
1987	12	5	8

Manejo del submódulo y solución del ejercicio.-

Los datos pueden introducirse desde el teclado (entrada manual) o importarse en formato Dbase, Excel o Access (entrada automática). En ambos casos, el usuario debe seleccionar en primer lugar la unidad de medida del tiempo entre semanas, cuatrisemanas, meses y años. A continuación, si la entrada es manual, se introducen el período inicial (por ejemplo, el mes y año cuando la serie de casos es mensual) y el número total de períodos (en ese caso, número de meses); entonces, el programa genera una tabla con tantas filas como períodos y sólo hay que introducir el número de casos observados y esperados en cada período.

Si la entrada es automática, Epidat 3.1 necesita que la tabla de datos tenga una estructura determinada, con dos variables que contengan, respectivamente, las series de casos observados y esperados, y una o dos variables que identifiquen el período temporal de ocurrencia, según la unidad de medida seleccionada (semanas, cuatrisemanas, meses o años). Por ejemplo, si la serie de casos es semanal, mensual o cuatrisemanal, la tabla debe contener una variable para identificar el año y otra para el número de semana, mes o cuatrisemana, respectivamente (como ejemplo, véase la Tabla 5); en el caso de series anuales, sólo es necesario el año de ocurrencia.

Resultados con Epidat 3.1

```
Vigilancia: Detección de clusters, agregaciones temporales

Archivo de trabajo: C:\Archivos de programa \Epidat 3.1 \Ejemplos
\Vigilancia \POISSON.xls

Campo que identifica:
  Casos observados: OBSERVADOS
  Casos esperados: ESPERADOS
  Meses: MES
  Años: AÑO

Método         : Poisson
Tipo de período: Meses

       Mes Observados   Esperados     Valor p
---------- ----------  ----------  ----------
    1-1987         11     10,0000      0,4170
    2-1987          8     10,0000      0,7798
    3-1987          9     10,0000      0,6672
    4-1987         14     10,0000      0,1355
    5-1987         19     10,0000      0,0072
    6-1987         11     10,0000      0,4170
    7-1987          9      8,0000      0,4075
    8-1987          8      8,0000      0,5470
    9-1987          7      8,0000      0,6866
   10-1987          4      8,0000      0,9576
   11-1987          0      8,0000      1,0000
   12-1987          5      8,0000      0,9004


Valor p: probabilidad de que n° de casos esperado sea >= al observado.
```

Epidat, con la información de los casos observados y de los casos esperados, calcula, para cada mes, la probabilidad de observar esos casos o más bajo una distribución de Poisson. Si esa probabilidad resulta ser muy pequeña, se concluye que hay evidencia de una agregación temporal que no es producto del azar. Nótese como en el mes de mayo se da un número inusitado de casos con una baja probabilidad.

Casos agrupados: Método Cusum

Esta técnica fue diseñada por Page[23] para el control de calidad de algunos procesos industriales y se utiliza en epidemiología para detectar incrementos en la incidencia de una enfermedad, por ejemplo para vigilar malformaciones congénitas[15,22].

Los datos necesarios para la aplicación de esta técnica son los siguientes:

- La serie temporal de casos observados (en días, semanas, cuatrisemanas o meses): O_i.
- E: el número esperado de casos por unidad de tiempo.
- El CUSUM inicial.

A partir del número esperado de casos, E, se obtienen dos valores K y h, conocidos como valor de referencia y de alarma, respectivamente, de modo que maximicen la capacidad de detección del método y minimicen la probabilidad de producir falsas alarmas. Estos valores están tabulados[24] en función de E cuando E < 9. Si E ≥ 9 se toman K=1 y h=2.

La técnica basa su decisión en la comparación del valor CUSUM, calculado para cada intervalo de tiempo, con el valor h. Si CUSUM > h, se detecta un cambio significativo en la frecuencia de la enfermedad y se declara una alerta.

Para obtener los valores CUSUM se procede del modo siguiente:

1. Para cada período k se calcula:

$$S_k = Max\{0; O_i - K\} \text{ si E < 9}$$

$$S_k = Max\left\{0; \frac{O_i - E}{\sqrt{E}} - 1\right\} \text{ si E ≥ 9}$$

2. El CUSUM del período i es la suma acumulada de valores S_k desde el período inicial, si aún no se ha declarado ninguna alerta o, en caso contrario, desde el período siguiente al de la última alerta, y hasta el período i. Para calcular el primer valor acumulado $CUSUM_1$ se añade a S_1 el valor del cusum inicial $CUSUM_0$.

Limitaciones del método Cusum

- La sensibilidad del método depende de la incidencia de la enfermedad en la población.
- Cuando el número esperado de casos es mayor o igual que 9 el método funciona peor.

Ejercicio

En una población se han observado 26 casos de anencefalia en los 7 primeros meses del año 1990 y una revisión de los registros del año anterior indica que el número esperado de casos para esta región es de 2 casos mensuales. Los datos, que se han tomado del manual del CLUSTER[13] se presentan en la Tabla 6 y se encuentran, además, en el archivo CUSUM.xls distribuido con Epidat 3.1. Partiendo de un CUSUM inicial igual a 0, ¿hay evidencia de un cluster de anencefalia en algún mes del período estudiado ?

Manejo del submódulo y solución del ejercicio.-

Los datos pueden introducirse desde el teclado (entrada manual) o importarse en formato Dbase, Excel o Access (entrada automática). En ambos casos, el usuario debe seleccionar en primer lugar la unidad de medida del tiempo entre días, semanas, cuatrisemanas o meses. A

continuación, si la entrada es manual, se introducen el período inicial (por ejemplo, el mes y año cuando la serie de casos es mensual) y el número total de períodos (en ese caso, número de meses); entonces, el programa genera una tabla con tantas filas como períodos y sólo hay que introducir el número de casos observados en cada período.

Si la entrada es automática, Epidat 3.1 necesita que la tabla de datos tenga una estructura determinada, con una variable que contenga la serie temporal de casos observados, y una o dos variables que identifiquen el período temporal de ocurrencia, según la unidad de medida seleccionada (días, semanas, cuatrisemanas o meses). Por ejemplo, si el período elegido es días, sólo se necesita una variable temporal que contenga la fecha de ocurrencia de los casos. Si la serie de casos es semanal, mensual o cuatrisemanal, la tabla debe contener una variable para identificar el año y otra para el número de semana, mes o cuatrisemana, respectivamente (como ejemplo, véase la Tabla 6).

Tabla 6. Casos mensuales de anencefalia entre enero y julio de 1990.

AÑO	MES	OBSERVADOS
1990	1	1
1990	2	3
1990	3	2
1990	4	4
1990	5	5
1990	6	5
1990	7	6

En cualquiera de los dos casos deben introducirse, además, el número esperado de casos por período de tiempo y el CUSUM inicial.

Resultados con Epidat 3.1

```
Vigilancia: Detección de clusters, agregaciones temporales

Archivo de trabajo: C:\Archivos de programa \Epidat 3.1 \Ejemplos
\Vigilancia \CUSUM.xls

Campo que identifica:
  Casos observados: OBSERVADOS
  Días: FECHA
  Meses: MES
  Años: AÑO

Método         : Cusum
Tipo de período: Meses
Cusum inicial  : 0
N° esperado de casos por Meses: 2

          Mes Observados       Cusum
------------ ---------- ----------
       1-1990          1       0,00
       2-1990          3       0,00
       3-1990          2       0,00
       4-1990          4       1,00
       5-1990          5       3,00
       6-1990          5       5,00
       7-1990          6       8,00  Alarma
```

Los resultados indican que el nivel de alerta se superó en el mes de julio, indicando una posible agrupación de casos de anencefalia en ese mes.

AGREGACIONES ESPACIALES

Los métodos para detectar agregaciones espaciales utilizan como criterio las "distancias", de modo que objetos próximos constituyen el mismo grupo y distancias grandes separan grupos.

El investigador dispone de dos tipos diferentes de datos: tasas correspondientes a áreas de variado tamaño (barrios, por ejemplo) o coordenadas de localización geográfica, las cuales se obtienen, por ejemplo, a partir de la dirección de residencia de los casos. En dependencia del tipo de dato disponible se explora cierto patrón espacial. Cuando se dispone de tasas se trata de determinar si las áreas con altas tasas forman clusters espaciales; si se dispone de coordenadas se busca determinar cuándo se agrupan localizaciones específicas de casos. Los dos métodos incluidos en Epidat para la detección de agregaciones espaciales, los de Grimson y Ohno, utilizan tasas de áreas geográficas.

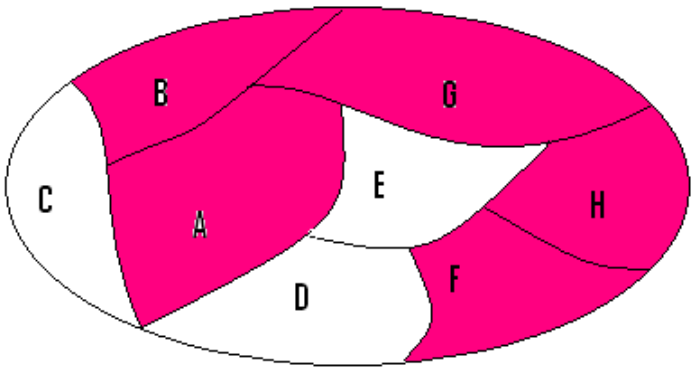
Método Grimson

Esta técnica[24] ha sido diseñada para la detección de agregaciones espaciales de alguna enfermedad, dentro de una región subdividida en áreas más pequeñas, cuando las áreas de mayor riesgo tienden a agruparse espacialmente.

La prueba compara el número de pares de áreas adyacentes de alto riesgo con un número esperado, o valor crítico, bajo el supuesto de que dichas áreas de alto riesgo estuvieran distribuidas aleatoriamente en la región de estudio.

Supongamos, por ejemplo, que se conocen las tasas de incidencia de una enfermedad en cada uno de los municipios de cierta provincia, con la distribución reflejada en la Figura 2.

Figura 2.- División en municipios de la provincia de estudio.



Las áreas marcadas son las consideradas de alto riesgo de la enfermedad, cuestión que se decide definiendo un punto de corte en las tasas de incidencia de la enfermedad, y que se supone que es la tasa de incidencia más pequeña que se considera alta. Por tanto, los municipios A, B, F, G y H son aquellos con tasas de incidencia iguales o mayores que el valor tomado como punto de corte. Se observa, además, que son adyacentes A con B, A con G, B con G, G con H y H con F; es decir, existen cinco pares de áreas adyacentes de alto riesgo en la provincia. La cuestión estriba, entonces, en decidir si este hecho se da por azar o se debe a que realmente existe una agregación espacial de estas áreas.

En resumen, la hipótesis nula de esta prueba establece que las áreas de alto riesgo están distribuidas aleatoriamente dentro de la región de estudio, y esta hipótesis se rechaza si el número observado de pares de áreas adyacentes de alto riesgo es mayor de lo esperado.

Los datos que se necesitan para la aplicación del método Grimson son:

- La identificación de cada una de las áreas (suelen utilizarse letras mayúsculas).
- Los índices que identifican, para cada área, las áreas que son adyacentes a ella, es decir, con las que comparte fronteras.
- Las tasas de incidencia de cada área.
- El punto de corte en las tasas de incidencia que define las áreas de alto riesgo.

Con estos datos se calculan:

- N: el número total de áreas.
- n: el número de áreas de alto riesgo.
- P: el número de pares de áreas adyacentes de alto riesgo.
- n_i: el número de áreas adyacentes al área i (i=1, ..., N) y la media y varianza de estos N valores:

$$\mu = \frac{\sum_{i=1}^{N} n_i}{N}; \sigma^2 = \frac{\sum_{i=1}^{N} (n_i - \mu)^2}{N}$$

Cuando n es pequeño comparativamente con N, se asume que P sigue una distribución de Poisson, cuestión que puede verificarse comprobando si la media y la varianza de P son aproximadamente iguales, E(P)≈V(P). Dichos valores pueden calcularse de forma exacta a partir de μ y σ, según el método propuesto por Grimson et al[25].

En esta situación, la probabilidad de observar P o más áreas adyacentes de alto riesgo, es decir, el valor p de la prueba, se obtiene a partir de la distribución de Poisson:

$$Valor\ p = \Pr[Poisson(E(P)) \geq P]$$

A medida que n aumenta, se hace mayor la diferencia entre los valores E(P) y V(P) y, en ese caso, puede utilizarse la distribución normal para calcular el valor p:

$$Valor\ p = \Pr\left[N(0,1) \geq \frac{P - E(P)}{\sqrt{V(P)}}\right]$$

Limitaciones del método Grimson

- Esta técnica otorga el mismo peso a todas las áreas que comparten fronteras con un área determinada, sin tener en cuenta la longitud de estos bordes. Esto podría introducir un sesgo en el análisis, por el hecho de que las áreas con fronteras más extensas tienen mayor probabilidad de conformar agregaciones que otras áreas con bordes más cortos. Este efecto podría neutralizarse, en alguna medida, tomando áreas lo más homogéneas posibles dentro de la región de estudio.
- La sensibilidad del método depende, en gran medida, de la elección del punto de corte para distinguir las áreas de alto riesgo.

Ejercicio

En una región con 11 áreas sanitarias se observaron, durante el año 2000, las tasas de mortalidad por cáncer de colon que se presentan en la Tabla 7; los datos se encuentran también en el archivo GRIMSON.xls de los ejemplos de Epidat 3.1. Si una tasa mayor o igual a 14 por 100.000 define un área de alto riesgo ¿existe evidencia de alguna agregación espacial entre esas áreas?

Manejo del submódulo y solución del ejercicio.-

Los datos pueden introducirse desde el teclado (entrada manual) o importarse en formato Dbase, Excel o Access (entrada automática). Si la entrada es manual, hay que indicar el número de áreas y el programa ya genera una tabla con tantas filas como áreas. En esa tabla, el usuario debe rellenar tres columnas correspondientes a la identificación de las áreas (puede ser un número o una letra), los índices de sus áreas adyacentes y las tasas de incidencia correspondientes. En el caso de la entrada automática, los datos se cargan desde un fichero en formato Dbase, Access o Excel, con una configuración como la mostrada en la Tabla 7.

Tabla 7. Áreas de la región junto a sus áreas adyacentes y tasas de incidencia.

AREA	INDICES*	TASAS
A	B,G	8,0
B	A,C,D,F,G	9,5
C	B,D,K	15,7
D	B,C,E,F	21,1
E	D,F,J,K	8,9
F	B,D,E,G,I,J,K	4,4
G	A,B,F,H,I	7,7
H	G,I	12,1
I	F,G,H,J,K	8,0
J	E,F,I	2,1
K	I,F,E,C	18,9

* Áreas adyacentes al área correspondiente

En cualquiera de los dos casos debe introducirse, además, el punto de corte en las tasas de incidencia que define un área de alto riesgo. En el ejemplo, el punto de corte se ha tomado a 14, por lo que se tienen tres áreas de alto riesgo (C, D y K) cuyas tasas son superiores a dicho valor. Así mismo, se tendrán sólo dos pares de áreas adyacentes de alto riesgo (CK, CD), ya que entre D y K no hay contigüidad.

Resultados con Epidat 3.1

```
Vigilancia: Detección de clusters, agregaciones espaciales

Archivo de trabajo: C:\Archivos de programa \Epidat 3.1 \Ejemplos
\Vigilancia \Grimson.xls

Campo que contiene:
  Área: AREA
  Índices de áreas adyacentes: INDICES
  Tasas: TASAS
Número de áreas     : 11

Método Grimson
```

```
Pares de áreas adyacentes de alto riesgo
----------------------------------------
Observadas:       2
Esperadas :  1,2000
Probabilidad { >= 2} =  0,1683
```

Los resultados indican que la probabilidad de observar 2 ó más pares de áreas adyacentes de alto riesgo es superior al valor convencional de 0,05, por tanto, no hay evidencia para rechazar la hipótesis nula a un nivel de significación del 5%. Se concluye, entonces, que las agregaciones observadas se han producido por mero azar.

Método Ohno

Este método ha sido desarrollado[26] para identificar patrones geográficos de mortalidad o morbilidad observados visualmente en el mapa de una región dividida en áreas más pequeñas (municipios, comarcas, o áreas sanitarias, por ejemplo).

Si las áreas adyacentes tienden a presentar niveles de la enfermedad más similares de lo que se esperaría por azar, entonces existe evidencia de agregación espacial.

Para la aplicación de la técnica de Ohno, se necesita la siguiente información:

- La identificación de cada una de las N áreas en que se divide la región de estudio (suelen utilizarse letras mayúsculas).
- Los índices que identifican, para cada área, las áreas que son adyacentes a ella, es decir, con las que comparte fronteras.
- Las tasas de incidencia de cada área.
- El número de categorías (k) en las que se quieren agrupar las áreas en función de sus tasas de incidencia, y los k-1 puntos de corte de las tasas de incidencia que definen esas categorías. La decisión de qué valores tomar compete al investigador o grupo de investigadores, basándose en la experiencia y conocimiento de la enfermedad objeto de estudio, y debe abarcar una gama de riesgos, desde los menores hasta los más elevados observados en la región de estudio.

El primer paso del método consiste en clasificar las áreas según el nivel de riesgo de enfermedad que presenta cada una, es decir, atendiendo a los puntos de corte definidos para las tasas de incidencia. Así, a partir de los k-1 puntos de corte V_i (i=1,...,k-1) y las tasas de incidencia de la enfermedad para cada área T_i (i= 1,...,N), las k categorías de riesgo quedan establecidas de la siguiente manera:

Tabla 8. Clasificación de las áreas en categorías.

Categoría	Punto de corte	Definición	Número de áreas
1	V_1	$T_i<V_1$	N_1
2	V_1, V_2	$V_1\leq T_i<V_2$	N_2
...	...	...	...
k-1	V_{k-2}, V_{k-1}	$V_{k-2}\leq T_i<V_{k-1}$	N_{k-1}
k	V_{k-1}	$T_i\geq V_{k-1}$	N_k

Luego se calcula el número de pares de áreas, dentro de cada categoría, que son adyacentes, es decir, el número de pares de áreas que cumplen a la vez dos condiciones: están en el mismo nivel de riesgo (son *concordantes*) y tienen contigüidad espacial (son *adyacentes*).

El método de Ohno consiste en comparar, mediante un estadístico Ji-cuadrado, el número observado de pares de áreas *adyacentes y concordantes* en la categoría i-ésima (AC_i, i=1, ..., k) con el número esperado (EAC_i, i=1, ..., k) bajo la hipótesis de que la distribución es uniforme en toda la región de estudio:

$$\chi^2 = \left(\frac{AC_i - EAC_i}{\sqrt{EAC_i}} \right)^2 , \text{ i=1, ..., k}$$

donde:

$$EAC_i = \frac{A}{\frac{N(N-1)}{2}} \frac{N_i(N_i - 1)}{2} , \text{ i=1, ..., k}$$

y A el número total de pares de áreas adyacentes

En esta última fórmula el primer factor representa la proporción de pares de áreas en toda la región de estudio que son adyacentes, y el segundo es el número máximo de pares de áreas concordantes (de las cuales no necesariamente todas son adyacentes) que pueden formarse en la categoría i-ésima.

También se realiza una prueba global para comparar el número observado de áreas adyacentes y concordantes de toda la región (AC) con su número esperado (EAC), que de nuevo se calcula bajo la hipótesis de que dichas áreas se distribuyen uniformemente en la región:

$$\chi^2 = \left(\frac{AC - EAC}{\sqrt{EAC}} \right)^2$$

donde:

$$AC = \sum_{i=1}^{k} AC_i \text{ y } EAC = \sum_{i=1}^{k} EAC_i$$

El objetivo de esta prueba global es la detección de agregaciones espaciales en toda la región, sin considerar una categoría de riesgo específica.

Limitaciones del método Ohno

- Al igual que el método de Grimson, esta técnica otorga el mismo peso a todas las áreas que comparten fronteras con un área determinada, sin tener en cuenta la magnitud de esas fronteras.
- La aproximación Ji-cuadrado del estadístico es válida si el número de pares de áreas adyacentes es mucho más pequeño que el número total de pares posibles: N(N-1)/2.

Ejercicio

En una región con 10 áreas de salud se observaron, durante el año 2000, las tasas de mortalidad por cáncer de colon que se presentan en la Tabla 9; los datos se encuentran también en el archivo OHNO.xls de los ejemplos de Epidat 3.1. El equipo de investigadores ha decidido utilizar cuatro categorías de nivel de riesgo de la enfermedad, para lo cual definen tres puntos de corte en las tasas: 2, 6 y 10, ¿existe evidencia de alguna agregación espacial entre esas áreas?

Manejo del submódulo y solución del ejercicio.-

Los datos pueden introducirse desde el teclado (entrada manual) o importarse en formato Dbase, Excel o Access (entrada automática). Si la entrada es manual, hay que indicar el número de áreas y el programa ya genera una tabla con tantas filas como áreas. En esa tabla, el usuario debe rellenar tres columnas correspondientes a la identificación de las áreas (puede ser un número o una letra), los índices de sus áreas adyacentes y las tasas de incidencia correspondientes. En el caso de la entrada automática, los datos se cargan desde un fichero en formato Dbase, Access o Excel, con una configuración como la mostrada en la Tabla 9.

Tabla 9. Áreas de la población junto a sus áreas adyacentes y tasas de incidencia.

AREA	INDICES	TASAS
A	B,G,H	7
B	A,G,F,D,C	8,5
C	B,D	11,9
D	C,B,F,E	15,7
E	D,F,J	1,5
F	B,G,I,J,E,D	4,4
G	B,A,H,I,F	7,7
H	A,G,I	1,1
I	H,G,F,J	8,3
J	I,F,E	3,3

En cualquiera de los dos casos debe introducirse, además, el número de categorías de riesgo y los puntos de corte en las tasas de incidencia para definir esas categorías.

Resultados con Epidat 3.1

```
Vigilancia: Detección de clusters, agregaciones espaciales

Archivo de trabajo: C:\Archivos de programa \Epidat 3.1 \Ejemplos
\Vigilancia \OHNO.xls

Campo que contiene:
  Área: AREA
  Índices de áreas adyacentes: INDICES
  Tasas: TASAS

Número de áreas     : 10
Nº de categorías    : 4
Método Ohno

                 Áreas adyacentes concordantes
Categoría   Áreas   Observadas    Esperadas Ji-cuadrado  Valor p
---------   -----  ------------ ------------ -----------  -------
        1       2             0       0,4222      0,4222   0,5158
        2       2             1       0,4222      0,7906   0,3739
        3       4             4       2,5333      0,8491   0,3568
        4       2             1       0,4222      0,7906   0,3739
---------   -----  ------------ ------------ -----------  -------
    Total      10             6       3,8000      1,2737   0,2591
```

En los resultados, Epidat presenta las categorías de riesgo con el número de áreas incluidas en cada una de ellas. También incluye el número observado y esperado de pares de áreas adyacentes y concordantes para cada nivel de riesgo y, por último, los valores χ^2 y los

valores p correspondientes a las pruebas parciales y global. En el ejemplo, no se observa evidencia de agregación espacial en ninguna categoría ni globalmente.

AGREGACIONES ESPACIO-TEMPORALES

El estudio de aglomeraciones en dominios espacio-temporales va más allá de la detección de clusters en el espacio, por un lado, y en el tiempo, por otro; de hecho, ambos pueden existir por separado sin que haya interacción. La interacción supone que los casos cercanos en el espacio son, además, cercanos en el tiempo, de modo que la localización de un caso depende de la localización del caso que lo precede. Este tipo de patrones es muy útil cuando se desea investigar enfermedades transmisibles. También se pone de manifiesto la interacción cuando la causa de la enfermedad es la exposición a un agente geográficamente localizado (una sustancia tóxica, ciertos tipos de radiaciones, etc.). Epidat incluye el método de Knox para la detección de agregaciones espacio-temporales.

Método Knox

Este método fue diseñado por Knox[28] para detectar agregaciones espacio-temporales, que ocurren cuando los casos observados de una enfermedad en determinada región guardan cercanía, tanto en espacio como en tiempo. El método de Knox tiene la ventaja de que no es necesario conocer el tamaño ni las características de la población en estudio.

Para aplicar el método de Knox es necesario disponer de los siguientes datos:

- Las coordenadas espaciales X e Y correspondientes a cada uno de los casos de enfermedad ocurridos durante el período de estudio. Estas coordenadas se determinan en un sistema cartesiano, tomando como origen un punto arbitrario dentro de la región estudiada, y están expresadas en una unidad de longitud.
- Las fechas de ocurrencia de los casos (dd/mm/aaaa).

Además, debe definirse de antemano qué se entenderá por proximidad espacial y temporal. Para ello, el usuario tiene que fijar dos valores críticos *e* y *t*, los cuales constituyen, respectivamente, la distancia máxima aceptada entre dos casos para que puedan ser considerados cercanos espacialmente (distancia espacial crítica) y el tiempo máximo entre la ocurrencia de dos casos para considerarse próximos en tiempo (distancia temporal crítica). Los criterios para definir la cercanía en el espacio y en el tiempo dependerán de las características epidemiológicas de la enfermedad.

El procedimiento consiste en calcular las distancias espaciales y temporales entre todos los pares de casos y, a partir de los valores críticos establecidos, definir dos variables dicotómicas que expresan si un par de casos están o no próximos en el espacio y en el tiempo.

Para calcular la distancia espacial entre dos casos (x_i, y_i) y (x_j, y_j) se utiliza la distancia euclídea (la recta más corta que une dos puntos):

$$e_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

Por su parte, la distancia temporal es la diferencia, en valor absoluto, entre las fechas de ocurrencia de los dos casos:

$$t_{ij} = |f_i - f_j|$$

A partir de estos valores se clasifican los N(N-1)/2 pares de casos en una tabla de doble entrada (2x2), según las cuatro posibilidades que existen, de la siguiente manera:

Tabla 10. Disposición de los pares según su cercanía espacial y temporal.

		Cercanos en tiempo		
Cercanos espacialmente		Sí	No	Total
	Sí	A (pares cercanos en tiempo y espacio)	B (pares cercanos sólo en espacio)	A+B
	No	C (pares cercanos sólo en tiempo)	D (pares sin cercanía de ningún tipo)	C+D
	Total	A+C	B+D	n =A+C+B+D

El criterio de proximidad en cada dimensión se define teniendo en cuenta la distancia espacial crítica *e* y la distancia temporal crítica *t*:

- Pares cercanos en tiempo y espacio: $e_{ij} \leq e$ y $t_{ij} \leq t$
- Pares cercanos sólo en tiempo: $t_{ij} \leq t$ y $e_{ij} > e$
- Pares cercanos sólo en espacio: $e_{ij} \leq e$ y $t_{ij} > t$
- Pares sin cercanía de ningún tipo: $e_{ij} > e$ y $t_{ij} > t$

En este método, la hipótesis nula establece que no existe agregación espacio-temporal entre los pares de casos, y para contrastarla se compara el número observado de pares cercanos en espacio y tiempo con el número esperado bajo la hipótesis nula. Bajo ciertas condiciones[29], puede asumirse que A sigue una distribución de Poisson con media $\lambda=(A+C)(A+B)/n$, y con esta asunción se calcula el valor p de la prueba:

$$Valor\ p = \Pr[Poisson(\lambda) \geq A]$$

Limitaciones del método Knox

- Como el método Knox basa su criterio de significación en la interacción espacio-tiempo, no es sensible para detectar clusters en una sola de estas dimensiones. Por ejemplo, agregaciones que ocurren en un corto período de tiempo pero separadas espacialmente en áreas distantes, con seguridad no serán detectadas con este método, por lo que es aconsejable el uso de otras técnicas para la detección de agregaciones temporales.
- La aproximación de Poisson no siempre es válida. Barton y David[29] dan una condición para que la distribución de Poisson aproxime bien la verdadera distribución del número de pares cercanos en el espacio y en el tiempo (A); debe cumplirse que los totales A+B y A+C sean pequeños en comparación con el total N. Dichos autores también proporcionan una aproximación Gaussiana basada en la media y varianzas aproximadas de A, que se recomienda utilizar cuando no se cumpla la condición anterior.
- Debe tenerse en cuenta que los resultados obtenidos con este método dependen fuertemente de los valores críticos *e* y *t*, por lo que es recomendable prestar un poco de atención a la hora de elegirlos y, en ocasiones, se aconseja probar con varios conjuntos de estos valores[30]. Una posible solución para seleccionar las distancias

críticas se basa en los percentiles, y tiene la ventaja de que así las distancias son independientes de la escala de medida.

Ejercicio

Se han detectado 13 casos de una enfermedad en cierta región durante el año 1997. Los investigadores presumen la existencia de una agregación espacio-temporal, ya que han observado la distribución de los casos en un mapa cuadriculado y en una línea temporal asociada, cuestión que les hace pensar en tal posibilidad. Para verificar tal hipótesis han definido dos valores críticos, uno espacial e=8 y otro temporal t=8 ¿hay evidencia de agregación espacio-temporal entre los casos? Las fechas de ocurrencia de los casos y sus coordenadas espaciales pueden verse en la Tabla 11; además, los datos se encuentran en el archivo KNOX.xls de los ejemplos de Epidat 3.1.

Tabla 11. Coordenadas espaciales y fecha de ocurrencia de los casos registrados.

CASO	FECHA	COORDX	COORDY
1	05/01/1997	1	1
2	05/01/1997	10	4
3	06/01/1997	3	2
4	12/01/1997	3	2
5	13/01/1997	2	1
6	16/01/1997	1	1
7	21/01/1997	4	4
8	29/01/1997	2	2
9	03/02/1997	5	3
10	04/02/1997	2	5
11	05/02/1997	3	4
12	11/02/1997	4	3
13	11/02/1997	8	3

Manejo del submódulo y solución del ejercicio.-

Los datos pueden introducirse desde el teclado (entrada manual) o importarse en formato Dbase, Access o Excel (entrada automática). Si la entrada es manual, al indicar el número de casos el programa genera una tabla con tantas filas como casos, y el usuario debe rellenar tres columnas correspondientes a las coordenadas espaciales X e Y y a la fecha de ocurrencia de cada caso. En el caso de entrada automática, Epidat necesita que los datos tengan una estructura como la mostrada en la Tabla 11, es decir, se necesitan tres variables que identifiquen los campos coordenada espacial X, coordenada espacial Y y fecha de ocurrencia.

En ambos casos, el usuario debe definir, además, las distancias temporal y espacial críticas.

Resultados con Epidat 3.1

```
Vigilancia: Detección de clusters, agregaciones espacio-temporales

Archivo de trabajo: C:\Archivos de programa \Epidat 3.1 \Ejemplos
\Vigilancia \KNOX.xls
Campo que identifica:
  Coordenada espacial X: COORDX
  Coordenada espacial Y: COORDY
  Fecha: FECHA
Método                    : Knox
Distancia espacial crítica: 8
Distancia temporal crítica: 8
```

```
Número de casos:       13

                          TIEMPO           Total
                        Si       No
     ESPACIO       -------  -------  -------
                Si      26       47       73
                No       2        3        5
                ----------  -------  -------
             Total      28       50       78

N° esperado de pares cercanos en espacio y tiempo: 26,2051
Probabilidad de observar al menos 26 pares cercanos en espacio y tiempo:
0,5420
```

La probabilidad de observar 26 o más pares de casos cercanos en espacio y tiempo es elevada (0,5420); por tanto, no hay evidencia de clúster espacio-temporal en los casos del ejemplo.

OTRAS AGREGACIONES: Método REMSA

Este método, debido a Aldrich[31], ha sido diseñado para la detección de agregaciones de diversa índole, como pueden ser aquellas atribuibles a la concurrencia de múltiples características: biológicas (sexo, raza, edad, ...), socio-económicas (ocupación, nivel de escolaridad, ...), espaciales (residencia urbana, rural, ...), temporales (fecha en que ocurre el evento en estudio: mes en que se adquiere una enfermedad o fallece la persona, ...), etc. Por ejemplo, puede observarse que existe un número inusualmente alto de casos en los individuos de una región que enfermaron y comparten ciertas características: son todos del sexo masculino, de raza blanca, viven en zonas urbanas metropolitanas y acudieron a la consulta médica en el mes de julio, ya que presentaron síntomas en ese mes.

El método REMSA explora si una enfermedad se encuentra aleatoriamente distribuida dentro de ciertos subgrupos de la población y, para ello, se basa en clasificar la población y los enfermos según varios atributos simultáneamente. En la Tabla 12, por ejemplo, se ha distribuido una población de T=100 habitantes por sexo, raza y residencia; entre paréntesis se ha escrito el número de casos de cierta dolencia que se adjudica a cada celda, de un total de N=8 casos observados en esa comunidad.

Tabla 12. Clasificación de la población y los casos según los factores biológicos sexo y raza y el factor espacial residencia.

		Sexo							
		Masculino (1)				**Femenino (2)**			
		Residencia				**Residencia**			
		1	**2**	**3**	**4**	**1**	**2**	**3**	**4**
Raza	**Otra (2)**	2 (1)	11 (0)	5 (0)	7 (0)	12 (1)	4 (0)	5 (0)	3 (2)
	Blanca (1)	11 (3)	6 (1)	12 (0)	4 (0)	6 (0)	2 (0)	4 (0)	6 (0)

Sobre la base de dicha clasificación, se calcula la probabilidad que tiene un individuo de la población de pertenecer a una celda dada, es decir, de poseer determinados atributos. Así, en este ejemplo, la probabilidad que tiene un individuo de ser de raza blanca, sexo masculino y residir en el área 1, se estima en 11/100, es decir, 0,11.

La hipótesis nula de que no hay ningún tipo de agregación en los casos se traduce en que éstos se distribuyen igual que la población respecto a las características de la tabla. Entonces puede asumirse que el número de casos observados en la celda i-ésima sigue una distribución binomial de parámetros N (tamaño de la población) y p_i (probabilidad de la celda) y, bajo esa hipótesis, se calcula la probabilidad de que el número de casos observados en la celda i (N_i) sea mayor de lo esperado:

$$Valor\ p = \Pr\{Bin(N, p_i) \geq O_i\}$$

Esta técnica puede ser útil en ocasiones en las cuales se conoce poco sobre la epidemiología de la enfermedad. Como un instrumento de exploración puede brindar elementos importantes sobre las características compartidas por los casos, como similitudes demográficas (edad, sexo, etc), ocupacionales, y de otra índole, que pueden ser de gran utilidad en la génesis de hipótesis etiológicas.

Limitaciones del método REMSA

- Se requiere que los datos sean clasificados según atributos o categorías nominales y, por tanto, las variables continuas deben ser categorizadas, lo que puede implicar una pérdida de capacidad explicativa. Además, es necesario conocer la distribución de la población y de los casos en las categorías definidas.

Ejercicio

Con los datos de la tabla 12, explorar si en algún subgrupo de población definido por el sexo, la raza y la residencia existe evidencia de agregación de casos de la enfermedad. Estos datos se encuentran en el archivo REMSA.xls de los ejemplos de Epidat 3.1.

Manejo del submódulo y solución del ejercicio.-

Los datos pueden introducirse desde el teclado (entrada manual) o importarse en formato Dbase, Access o Excel (entrada automática). Si la entrada es manual, debe indicarse el número de variables nominales y el programa genera, en primer lugar, una tabla con tantas filas como variables en la que deben rellenarse dos columnas, una con los nombres de las variables (por defecto, el programa utiliza las etiquetas V-i), y otra con el número de categorías de cada variable. Con esta información, Epidat genera una segunda tabla con $C_1 \times C_2 \times ... \times C_k$ filas, donde k es el número de variables y C_i el número de categorías de la variable i; en esta tabla cada fila representa un subgrupo de población definido por una de las categorías de cada variable, y las columnas que debe rellenar el usuario son dos: la población y el número de casos (como ejemplo véase la Tabla 13).

En el caso de entrada automática, los datos se cargan desde un fichero en formato Dbase, Access o Excel, con una configuración como la mostrada en la Tabla 13.

Tabla 13. Formato de tabla preparada para importar datos desde Epidat 3.1 para la aplicación del método REMSA.

SEXO	RAZA	RESIDENCIA	CASOS	POBLACION
1	1	1	3	11
1	1	2	1	6
1	1	3	0	12
1	1	4	0	4
1	2	1	1	2
1	2	2	0	11
1	2	3	0	5
1	2	4	0	7
2	1	1	0	6
2	1	2	0	2
2	1	3	0	4
2	1	4	0	6
2	2	1	1	12
2	2	2	0	4
2	2	3	0	5
2	2	4	2	3

Resultados con Epidat 3.1

```
Vigilancia: Detección de clusters, otras agregaciones

Archivo de trabajo: C:\Archivos de programa \Epidat 3.1 \Ejemplos
\Vigilancia \REMSA.xls
Campo que contiene:
  Casos: CASOS
  Población: POBLAC
  Variables nominales:
  SEXO
  RAZA
  RESIDENCIA

SEXO    RAZA    RESIDENCIA    Casos  Probabilidad  Valor p
-----------  ----------  ----------  ----------  --------------------
   1     1           1           3         0,11       0,04873
   1     1           2           1         0,06       0,39043
   1     2           1           1         0,02       0,14924
   2     2           1           1         0,12       0,64037
   2     2           4           2         0,03       0,02234

Valor p: probabilidad de que el n° de casos esperado sea >= al observado.
```

Se observa que en las cinco celdas donde hay casos, Epidat calcula la probabilidad p_i de que un individuo pertenezca a esa celda, y luego presenta la probabilidad de obtener un número de casos igual o mayor al observado.

Si se toma un nivel de significación de 0,05, puede asumirse que en las líneas primera y última se da una agregación inusual de casos, puesto que la probabilidad de observar 3 ó más casos (0,04873) y 2 ó más casos (0,02234), respectivamente, es pequeña como para pensar que estos resultados se han producido por azar. Esto es, en la celda correspondiente a los hombres de raza blanca y que residen en el área 1 y en la celda referida a las mujeres de otra raza y residencia en el área 4, hay evidencia de que se produjo una agregación de casos de la dolencia en cuestión.

Bibliografía

1.- Raubertas RF. Spatial and temporal analysis of disease occurrence for detection of clustering. Biometrics. 1988;44:1121-9.

2.- Rodrigues LC, Marshall T, Murphy M, Osmond C. Space time clustering of births in SIDS: Do perinatal infections play a role ? Int J epidemiol. 1992;4:714-9.

3.- Chen R, Mantel N, Klingberg MA. A study of tree techniques for time-space clustering in Hodgkin's disease. Stat Med. 1984;3:173-84.

4.- Glaser SL. Spatial clustering of Hodgking's diseases in San Francisco by area. Am J Epidemiol. 1990;132 Suppl:167-77.

5.- Ross A.; Scott D. Point pattern analysis of the spatial proximity of residences prior to diagnosis of persons with Hodgkin's Disease. Am J Epidemiol. 1990;132 Suppl:53-61.

6.- Alexander FE, Mckinney PA, Willians J, Ricketts TJ, Cartwright RA. Epidemiological evidence for the 'two-disease hypotesis' in Hodkin's disease. Int J Epidemiol. 1991;20:354-61.

7.- Caldwell GG. Twenty two years of cancer cluster investigations at the Center for Disease Control. Am J Epidemiol. 1990;132 Suppl:43-7.

11.- CLUSTER 3.1 Software System for Epidemiologic Analysis. Instruction Manual. February 1993. U.S. Department of Health & Human Services. Public Health Service. Agency for Toxic Substances and Disease Registry. Atlanta, Georgia.

12.- Chen R, Mantel N, Connelly RR, Isacson P. A monitoring system for chronic diseases. Meth Inform Med. 1982;21:86-90.

13.- Chen R. A surveillance system for congenital malformations. J Amer Stat Ass. 1978;73:323-7.

14.- Chen R. Statistical techniques in birth defects surveillance systems. Contr Epidemiol Biostat. 1979;1:184-9.

15.- Chen R, McDowell M, Terzian E, and Weatherall J. Eurocat Guide to Monitoring Methods for Malformation Registers, EEC Concerted Action Project. Bruxelles, 1983.

16.- Chen R. Revised values for the parameters of the Sets technique for monitoring the incidence rate of a rare disease. Meth Inform Med. 1986;25:47-9.

17.- Naus JI. The Distribution of the size of the maximum cluster of points on a line. J Am Stat Assoc. 1965;60:532-38.

18.- Naus JI. Approximations for distributions of scan statistics. J Am Stat Assoc. 1982;77:177-83.

19.- Glaz J. Approximations for the tail probabilities and moments of the scan statistic. Stat Med. 1993;12:1845-52.

20.- Hardy RJ, Buffler PA, Prichard HM, Schroeder GD, Cooper S, Crane M, Doak C. Monitoring for health effects of low-level radioactive waste disposal: a feasibility study. Report submitted to Texas Low-level Radioactive Waste Disposal Authority. Texas Dept. of Health, 1983.

21.- Hardy RJ, Schroeder GD, S, Cooper SP, Buffler PA, Prichard HM, Crane M. A surveillance system for assessing health effects from hazardous exposures. Am J Epidemiol. 1990;132(1):532-42.

22. Weatherall JAC, Haskey JC. Surveillance of malformations. Br Med Bull. 1976;32:39-44.

23.- Page ES. Continuous inspection schemes. Biometrika. 1954;41:100-15.

24.- Grimson RC, Wang KC, Johnson PWC. Searching for hierarchical clusters of disease: spatial patterns of sudden infant death syndrome. Soc Sci Med. 1981;15:287-93.

25.- Grimson RC, Rose RD. A versatile test for clustering and a proximity analysis of neurons. Meth Inform Med. 1991;30**:**299-303.

26.- Ohno Y, Aoki K, Aoki N. A test of significance for geographic clusters of disease. Int J Epidemiol. 1979;8:273-81.

27.- Bender AP, Willians AN, Johnson R, Jagger HG. Appropriate public health responses to cluster: the art of being resposibly responsive. Am J Epidemiol. 1990;132 Suppl:48-52.

28.- Knox G. The detection of space-time interactions. Appl Stat. 1964**;**13:25-9.

29.- Barton DE, David RN. The random intersection of two graphs. In: David FN, editor. Research papers in Statistics. New York: Wiley; 1966. p. 445-59.

30.- Smith P. Current assessment of "case clustering "of lymphomas and leukaemias. Cancer. 1978;42:1026-34.

31.- Aldrich TE. Detecting space-time aggregation of rare events. Am J Epidemiol. 1984;120:464.

GRÁFICO DE LÍMITES HISTÓRICOS

Conceptos generales

El análisis temporal de los datos de vigilancia epidemiológica puede hacerse usando diferentes métodos; el que aquí se presenta es uno de los más sencillos, basado en la comparación entre el número de casos declarados durante un periodo de cuatro semanas (cuatrisemanas) y una distribución histórica formada por 15 valores del quinquenio anterior. Estos 15 valores incluyen los casos declarados en las cuatrisemanas anterior, igual y posterior a la cuatrisemana que se quiere analizar (t) en los 5 años previos al analizado (a):

	t-1	t	t+1
Año a		X_0	
Año a-1	X_1	X_6	X_{11}
Año a-2	X_2	X_7	X_{12}
Año a-3	X_3	X_8	X_{13}
Año a-4	X_4	X_9	X_{14}
Año a-5	X_5	X_{10}	X_{15}

La elección de un periodo cuatrisemanal trata de evitar las fluctuaciones semanales en los datos, que reflejan más prácticas irregulares de notificación desde los puntos declarantes que cambios reales en la incidencia de la enfermedad. Por otro lado, la amplitud del periodo histórico elegido permite controlar el efecto estacional.

La desviación, tanto positiva como negativa, en la razón entre el número de casos declarados en la cuatrisemana a analizar y la media del periodo histórico indica alejamiento del patrón base. Esta razón se representa en escala logarítmica de tal manera que, cuando es igual a 1, no produce efecto en la gráfica. Para distinguir aquellas observaciones que merecen una atención especial se establecen los límites históricos construidos a partir de la media ± dos desviaciones estándar de los datos históricos.

En resumen, este método analítico compendia en un sencillo gráfico los casos notificados cuatrisemanalmente para diferentes procesos o enfermedades, así como los cambios en la incidencia respecto a sus datos históricos, lo cual permite identificar rápidamente aquellos procesos que necesitan un análisis más complejo.

Recomendaciones

- Cambios en el sistema de vigilancia pueden influir en la tendencia temporal de los datos, como, por ejemplo, modificaciones en la definición de caso utilizada en la vigilancia, nuevas técnicas diagnósticas, cambios en los criterios de notificación, etc. Todo esto habrá de tenerse en cuenta a la hora de interpretar los datos.
- También hay que tener en cuenta que un brote que afecte a la distribución histórica del quinquenio anterior, disminuirá la sensibilidad del método para detectar cambios en la incidencia de la semana que se analiza, por lo que sería aconsejable en este caso excluir los casos asociados al brote.

Ejercicio

La Tabla 3 contiene los casos cuatrisemanales de brucelosis declarados al Sistema de Enfermedades de Declaración Obligatoria de Galicia durante el período 1997-2002.

Se desea analizar la situación de esta enfermedad en las cuatrisemanas 6 y 8 del año 2002. ¿Qué indican los resultados observados?

Tabla 3. Casos cuatrisemanales de brucelosis declarados en Galicia entre 1997 y 2002.

	Año					
Cuatrisemana	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**
1	1	2	3	0	1	1
2	1	0	1	0	2	1
3	2	2	0	2	2	2
4	2	0	2	0	4	3
5	**2**	**4**	**1**	**0**	**1**	4
6	**3**	**1**	**3**	**1**	**2**	**7**
7	**3**	**4**	**1**	**1**	**2**	5
8	1	4	0	0	0	2
9	1	1	2	0	1	3
10	2	1	0	1	4	2
11	1	1	0	0	1	3
12	1	1	1	0	2	1
13	2	6	0	2	0	0

Manejo del submódulo y solución del ejercicio

Este submódulo permite representar el gráfico de límites históricos para una o varias enfermedades y los datos pueden introducirse tanto desde el teclado como importarse en formato Dbase, Excel o Access.

Cuando la entrada de datos se realiza de forma manual, hay que indicar al programa cuáles son el año y la cuatrisemana actual, indicar el número de enfermedades que se analizan y describirlas mediante una etiqueta que se utilizará en el gráfico. En "Datos y resultados" Epidat 3.1 define automáticamente una tabla con 16 filas: la primera corresponde a la cuatrisemana actual y las restantes corresponden al periodo histórico para esa cuatrisemana. Por ejemplo, los datos que se deberán introducir para analizar la cuatrisemana 6 de 2002 en el ejercicio serán los que aparecen sombreados en tono gris en la Tabla 3, datos que Epidat 3.1 pide en orden cronológico inverso, es decir, del más reciente al más antiguo.

Para realizar los cálculos a partir de datos procedentes de archivos importados, Epidat 3.1 necesita que éstos tengan una estructura determinada. Concretamente, en este submódulo la tabla debe tener al menos 4 variables: una que identifique la enfermedad, dos para definir el período temporal, año por un lado y cuatrisemana por otro, y otra que contenga el número de casos declarados (véase la Tabla 4). Por otra parte, es necesario que la tabla incluya todos las cuatrisemanas de al menos los 5 años previos al que se analiza, denotado en Epidat 3.1 "año actual", para poder definir el período histórico correctamente. Una vez identificadas las variables, hay que indicar al programa cuales son el año y la cuatrisemana actual y cargar los datos. La tabla con las etiquetas de las enfermedades se cubre automáticamente con los diferentes valores que toma la variable enfermedad en el archivo, pero se pueden modificar manualmente. Estas etiquetas son las que aparecerán en el gráfico.

Para analizar una cuatrisemana diferente del mismo archivo sólo hay que cambiar el valor de la cuatrisemana actual y volver a cargar los datos; la tabla con la información del período histórico se actualiza automáticamente.

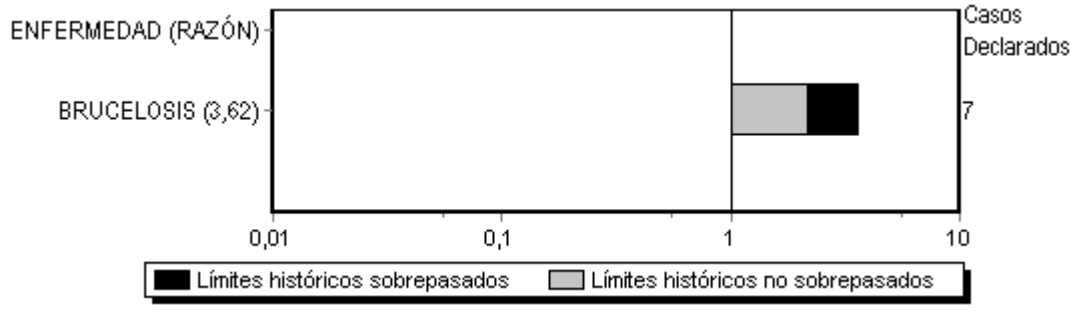
Tabla 4. Formato de tabla preparada para importar datos desde Epidat 3.1 en el submódulo de Límites históricos.

ENFERMEDAD	AÑO	CUATRISEMANA	CASOS
Brucelosis	1997	1	1
Brucelosis	1997	2	1
...	...	...	...
Brucelosis	1997	13	2
Brucelosis	1998	1	2
Brucelosis	1998	2	0
...	...	...	...
Brucelosis	1998	13	6
...	...	...	...
Brucelosis	2002	1	1
Brucelosis	2002	2	1
...	...	...	...
Brucelosis	2002	13	0

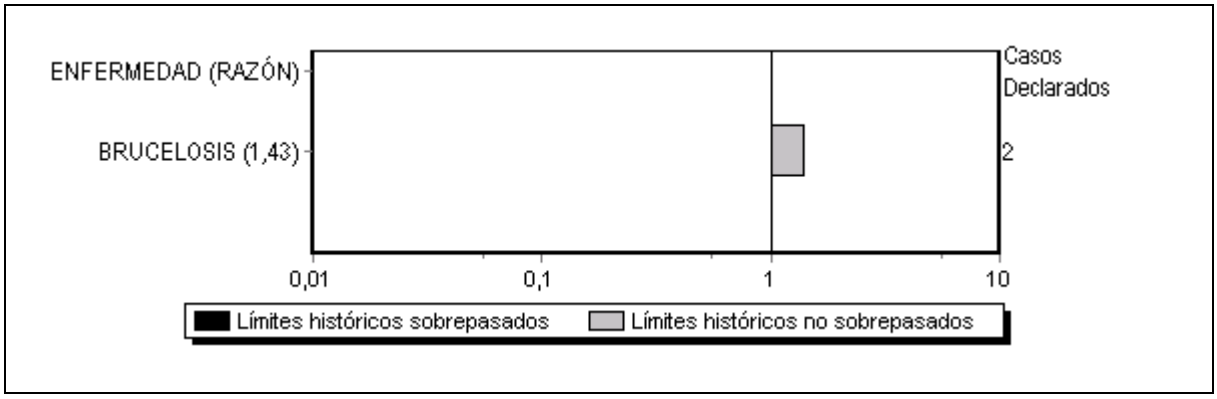
Los datos del ejercicio se encuentran en el archivo EDOS97-02.xls, incluido en Epidat 3.1, en la hoja *Brucelosis*. Este archivo contiene otra hoja llamada *Varias* que incluye la misma información para varias enfermedades.

Resultados del ejercicio con Epidat 3.1

```
Vigilancia: Gráfico de Límites Históricos

Archivo de trabajo: C:\Archivos de programa \Epidat 3.1 \Ejemplos
\Vigilancia en salud pública\EDOS97-02.xls
Campo que identifica:
  Enfermedad: ENFERMEDAD
  Año: AÑO
  Cuatrisemana: CUATRISEMANA
  Casos: CASOS

Número de enfermedades: 1
Año actual: 2002


Cuatrisemana actual: 6
```



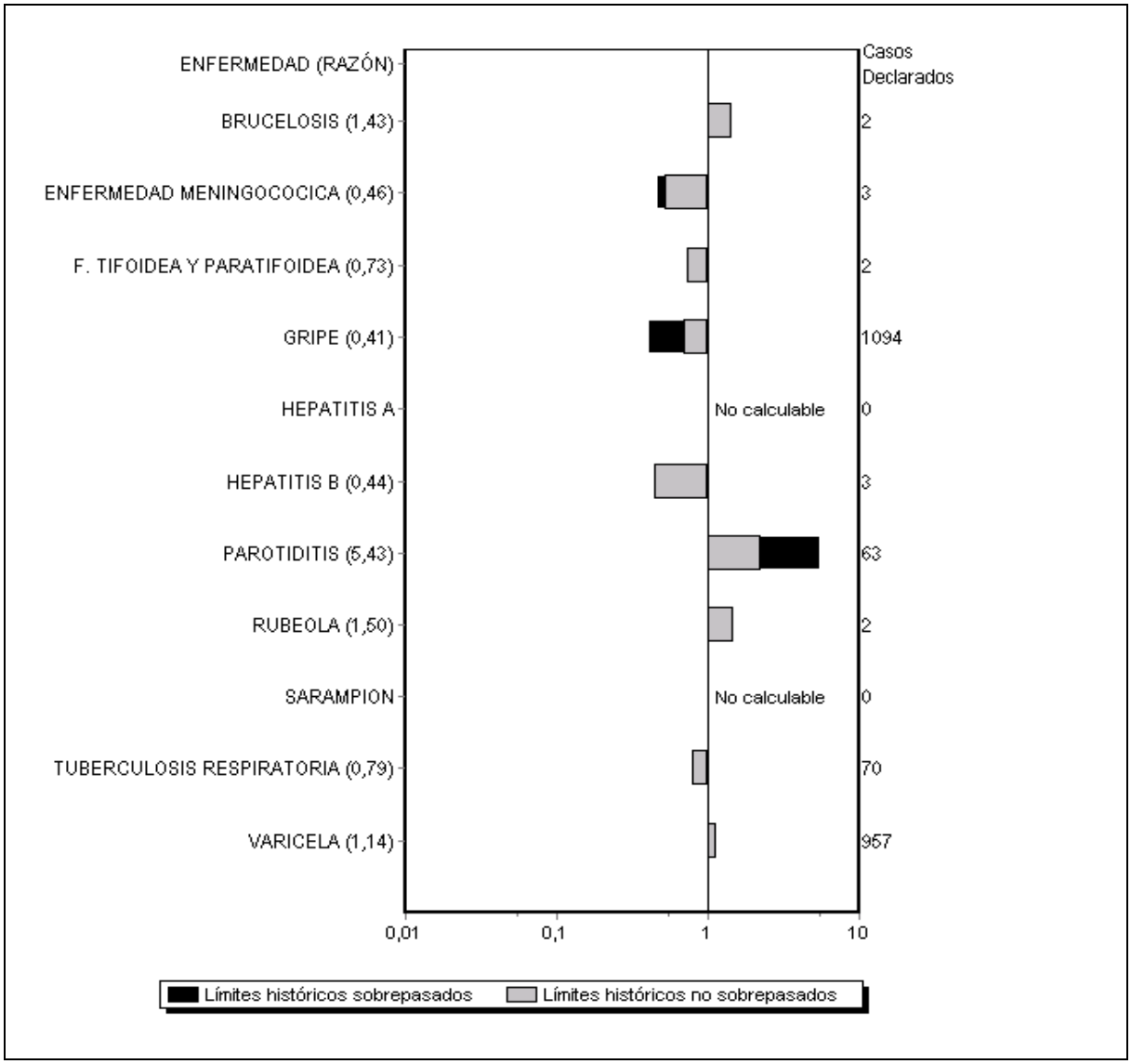

```
Cuatrisemana actual: 8
```



Los resultados observados muestran que para la cuatrisemana 6 se sobrepasaron los límites históricos, lo cual indica que deberá realizarse un análisis más profundo para saber qué pudo haber determinado esta desviación por encima del patrón base (por ejemplo, un brote).

Si se utilizan los datos de la hoja *Varias* incluida en el archivo del ejercicio, Epidat 3.1 muestra el siguiente gráfico de límites históricos para la cuatrisemana 8 del año 2002, en el que puede verse que tan sólo en el caso de la parotiditis se han sobrepasado los límites históricos. En el caso de la gripe y de la enfermedad meningocócica se puede observar el efecto contrario; es decir, el número de casos notificados en esta cuatrisemana es inferior al esperado.

```
Vigilancia: Gráfico de Límites Históricos

Archivo de trabajo: C:\Archivos de programa \Epidat 3.1 \Ejemplos \
Vigilancia en salud pública \ EDOS97-02.xls
Campo que identifica:
  Enfermedad: ENFERMEDAD
  Año: AÑO
  Cuatrisemana: CUATRISEMANA
  Casos: CASOS

Número de enfermedades: 11
Año actual: 2002
Cuatrisemana actual: 8
```



Bibliografía

1. Janes GR, Hutwagner LC, Cates W Jr, Stroup DF, Williamson GD. Descriptive Epidemiology: Analizing and Interpreting Surveillance Data. En: Steven M, Teutsch R, Churchill E. Editores. *Principles and Practice of Public Health Surveillance*. Oxford: Oxford University Press; 2000: 112-67.

GRÁFICO DE DISTRIBUCIÓN ESPACIAL

Conceptos generales

El análisis de la distribución espacial de los datos con vistas a la vigilancia epidemiológica permite identificar aquellas áreas/distritos/zonas, que declaran un mayor número de casos de los que cabría esperar bajo condiciones "normales". Incluso cuando los datos globales correspondientes a cierto proceso descienden, este tipo de análisis puede identificar áreas con tasas elevadas, lo cual permitiría políticas de intervención de salud pública más ajustadas. El ejemplo más clásico y conocido de una intervención basada en un análisis espacial fue el realizado por John Snow durante la epidemia de cólera[] ocurrida en Londres a finales del siglo XIX.

En el gráfico de distribución espacial se señalan aquellas áreas que declararon un número de casos superior al esperado si dichas áreas tuviesen la tasa de incidencia del área de referencia. Para ello se utiliza la siguiente metodología:

1. Si la población del área es mayor que el número de casos declarados multiplicado por cien, se marcarían como áreas de incidencia superior a la esperada, con un nivel de confianza del 95%, aquellas en las que la desviación estándar observada (S) es superior a 1,96, donde S viene dada por la fórmula siguiente:

$$S = (t - t_R)\sqrt{\frac{P_R}{t_R} - (t_R)^2}$$

2. Si la población del área es menor que el número de casos declarados multiplicado por cien, se marcarían como áreas de incidencia superior a la esperada con un nivel de confianza del 95%, aquellas con un valor de χ^2 superior a 3,84, siendo:

$$\chi^2 = \frac{(O - E)^2}{E} \text{ con } E = P t_R$$

En las expresiones anteriores t es la tasa de incidencia, P el tamaño de la población, O es el número de casos observados, E representa el número de casos esperados, y el subíndice R indica el área de referencia.

Una vez obtenidos los resultados se construye el gráfico de la siguiente forma:

Las filas representan enfermedades y las columnas áreas geográficas; se resaltan los cuadros donde la tasa de incidencia es superior a la esperada.

Recomendaciones

- El tamaño del área en este tipo de análisis debería estar determinado por el tipo de condición implicada. Así, para condiciones raras, pueden ser necesarias áreas más amplias, mientras que para eventos que ocurren con alta frecuencia o en situación de brote, puede ser necesario definir áreas más pequeñas.
- Este método analítico muestra aquellas áreas geográficas con una incidencia superior a la esperada para diferentes procesos o enfermedades, lo cual permite identificar rápidamente aquellas áreas que necesitan un análisis más complejo.

[] En 1854, el epidemiólogo John Snow documentó los casos de cólera que ocurrían en Londres y mostró que el cólera era mucho más frecuente entre los clientes de una compañía que se proveía de agua proveniente de la parte baja del río Támesis. Luego de examinar los datos, Snow concluyó que el problema se centraba en la bomba de agua de Broad Street; una vez anulada la bomba, la epidemia se detuvo.

- La naturaleza infecciosa de los datos de vigilancia epidemiológica así como su capacidad de transmisión hacen que éstos tiendan a la agregación espacial, que no siempre coincide con los límites administrativos predefinidos para las áreas geográficas.

Ejercicio

La siguiente tabla contiene los casos notificados por provincia durante el año 2002 al Sistema de Enfermedades de Declaración Obligatoria de Galicia:

Área	Población	Parotiditis	Rubéola	Sarampión
Galicia	2.732.412	1.161	16	6
A Coruña	1.108.513	884	12	2
Lugo	365.781	149	0	0
Ourense	345.479	34	1	3
Pontevedra	912.639	94	3	1

Se desea realizar un gráfico de distribución espacial de acuerdo con las áreas definidas, ¿qué indican los resultados observados?

Manejo del submódulo y solución del ejercicio

Este submódulo permite representar el gráfico de distribución espacial para una o varias enfermedades en diferentes áreas geográficas, tomando una de ellas como referencia, y los datos pueden introducirse desde el teclado o importarse en formato Dbase, Excel o Access.

Cuando la entrada de datos se realiza de forma manual, hay que precisar el número de áreas geográficas e indicar cuál de ellas es la de referencia, definir el número de enfermedades que se analizan y describirlas mediante una etiqueta. En el gráfico, las etiquetas no pueden superar los 20 caracteres, por lo que aparecerán cortadas aquellas con una longitud mayor. Por último, en "Datos y resultados" hay que cubrir la tabla con la población de cada área y el número de casos declarados por enfermedad y área.

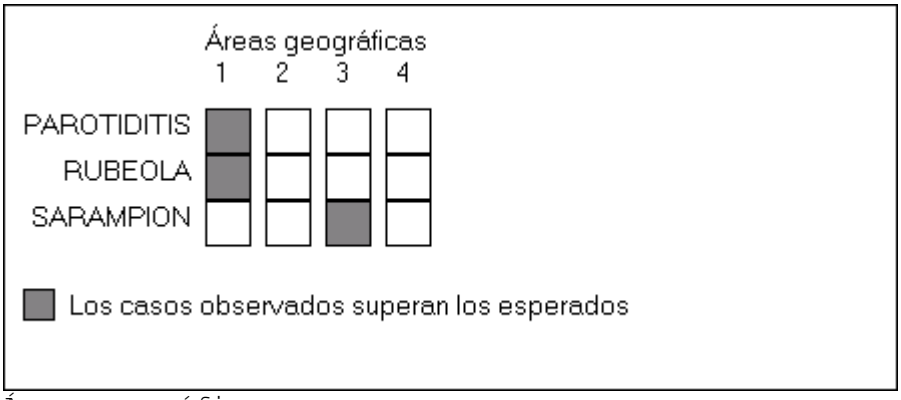
Para realizar los cálculos a partir de datos procedentes de archivos importados, Epidat 3.1 necesita que éstos tengan una estructura determinada. Concretamente, en este submódulo la tabla debe tener al menos 3 variables: una que identifique las áreas geográficas, otra con los respectivos tamaños poblacionales, y una variable para cada enfermedad que se incluya en el gráfico, que ha de contener los casos declarados por área. Una vez identificadas las variables, hay que indicar al programa cuál es el área de referencia, que por defecto Epidat 3.1 toma como la primera, y cargar los datos. La tabla con las etiquetas de las enfermedades se cubre automáticamente con los nombres de las variables que identifican las enfermedades en el archivo, pero se pueden modificar manualmente. Estas etiquetas son las que aparecerán en el gráfico.

Los datos del ejercicio se encuentran, con esa misma estructura, en el archivo PROVINCIAS-2002.xls incluido en Epidat 3.1.

Resultados del ejercicio con Epidat 3.1

```
Vigilancia: Gráfico de Distribución Espacial

Archivo de trabajo: C:\Archivos de programa \Epidat 3.1 \Ejemplos
\Vigilancia en salud pública \PROVINCIAS-2002.xls
Campo que identifica:
  Área: AREA
  Población: POBLACION
```

```
  Enfermedades:
                PAROTIDITIS
                RUBEOLA
                SARAMPION

Número de enfermedades: 3
Área de referencia: Galicia
Gráfico de distribución espacial
```




```
Áreas geográficas
1-A Coruña
2-Lugo
3-Ourense
4-Pontevedra
```

Los resultados observados muestran que en el área geográfica correspondiente a la provincia de A Coruña la incidencia de parotiditis y rubéola es superior a la esperada en ese periodo, lo mismo que para el sarampión en el área geográfica correspondiente a Ourense.

Bibliografía

1. Janes GR, Hutwagner LC, Cates W Jr, Stroup DF, Williamson GD. Descriptive Epidemiology: Analizing and Interpreting Surveillance Data. En: Steven M, Teutsch R, Churchill E. Editores. *Principles and Practice of Public Health Surveillance*. Oxford: Oxford University Press; 2000: 112-67.
2. Atkinson P, Molesworth A. Geographical analysis of communicable disease data. En: Elliot P, Wakefield J. Editores. *Spatial Epidemiology methods and applications*. Oxford: Oxford University Press; 2000: 253-65.

CORREDORES ENDÉMICOS

Conceptos generales

Uno de los principales objetivos de los sistemas de vigilancia es generar información que permita identificar precozmente cambios en los patrones de la morbilidad de enfermedades de importancia para la salud pública. Por ello, además de monitorizar cambios en los factores condicionantes de la morbilidad como, por ejemplo, el deterioro de las condiciones de vida de la población por reducción de su acceso a servicios básicos, incremento de vectores o reservorios, cambios en los niveles de resistencia a pesticidas, entre otros, se debe identificar lo más precozmente posible un incremento, más allá de lo esperado, en el número de casos o de sus tasas de incidencia.

Para ello un instrumento útil es el denominado "canal endémico" o "corredor endémico" (CE), que es la representación gráfica de la incidencia actual sobre la histórica, la cual alerta ante una incidencia superior a la esperada.

El empleo de los corredores endémicos es preferible para enfermedades endémicas de corto período de incubación, y evolución aguda, aunque no se excluye su uso para vigilar otras, e inclusive el comportamiento de la mortalidad. Pueden utilizarse para vigilar el comportamiento de una vía de transmisión o un microorganismo en particular.

Si bien el uso de los corredores endémicos es útil en los niveles nacionales y regionales, su utilización en los niveles locales puede facilitar la detección de pequeños brotes, a veces no identificables si se trabaja con datos agregados de zonas geográficas amplias.

Método de construcción del corredor endémico

Existen varios métodos para la construcción de un canal endémico, todos basados en determinar para cada periodo (semanas o cuatrisemanas) una medida de tendencia central y sus valores mínimos y máximos, con la finalidad de definir zonas de seguridad o alerta. El método más adecuado y más ampliamente difundido es el que utiliza el promedio de las medias geométricas de las tasas de incidencia de los últimos años. La ventaja de la media geométrica frente a la aritmética es que reduce la influencia que puede tener algún brote o epidemia que haya ocurrido en el periodo que se está considerando para el cálculo.

Para construirlo se deben utilizar las tasas de incidencia y no el número de casos, aunque se trate de una población pequeña, a fin de evitar la influencia que puede tener cambios importantes en el tamaño de la población en el periodo tomado como referencia.

En cuanto al número de años de la serie histórica, es aconsejable limitarlo a los últimos 5, 6 ó 7, ya que un mayor número de años puede mejorar el modelo de predicción, pero las condiciones que mantienen la endemia verosímilmente pudieran haber variado en ese lapso, así como los mecanismos de notificación y registro.

Lo ideal es construir canales endémicos semanales, ya que los mensuales limitan la posibilidad de detectar oportunamente los brotes y por lo tanto la implementación de medidas oportunas de control.

El método utilizado en Epidat 3.1 para la construcción del corredor endémico es el de la media geométrica. Se prefiere esta metodología debido a que la media geométrica es la medida de tendencia central ideal para distribuciones que no tienen una distribución normal, está especialmente indicada para valores aberrantes (por ejemplo, brotes con un número mucho más alto de lo esperado o habitual, o problemas de subregistro que pueden dar valores muy por debajo de los reales), ya que ellas se "diluyen" y no distorsionan la serie histórica.

Pasos para la construcción del corredor endémico:

- Para cada periodo (semanas o cuatrisemanas) se transforman logarítmicamente los valores de la tasas de incidencia de los últimos 5 a 7 años. Esta transformación comprime los valores altos y estira los bajos.
- En cada periodo se debe calcular el promedio de los valores transformados y su intervalo de confianza.
- Los estadísticos obtenidos en el paso anterior se deben convertir a sus valores originales calculando su antilogaritmo.
- En el gráfico se representan las series, semanales o cuatrisemanales de un año, de los casos observados, el promedio del período histórico (5, 6 ó 7 años previos) y los límites de su intervalo de confianza. Esto da lugar a cuatro zonas en el gráfico, de

abajo hacia arriba: **zona de éxito** (por debajo por debajo del límite inferior), **zona de seguridad** (entre el límite inferior y los casos esperados, sombreada en gris claro), zona de alerta (entre los casos esperados y el límite superior, sombreada en gris oscuro) y **zona epidémica** (por encima del límite superior).

El corredor endémico acumulativo se construye de igual modo pero utilizando la serie acumulada de casos. Este gráfico facilita la vigilancia de sucesos endémicos de baja incidencia.

Recomendaciones

- En el caso de la vigilancia de una enfermedad de baja incidencia, o de una población pequeña, o si se vigilan intervalos cortos de tiempo, la variabilidad aleatoria desempeña un papel importante, por lo cual será menor la precisión de la predicción.
- Asimismo, la inestabilidad en la incidencia histórica provoca corredores muy dentados, con anchas bandas de alerta y seguridad.

Ejercicio

El archivo MENINGO91-02.xls contiene los casos cuatrisemanales de enfermedad meningocócica declarados al Sistema de Enfermedades de Declaración Obligatoria de Galicia durante el período 1991-2002. La siguiente tabla contiene las poblaciones de Galicia para cada año de ese período:

Año	Población
1991	2.752.322
1992	2.740.615
1993	2.731.958
1994	2.726.150
1995	2.722.967
1996	2.721.540

Año	Población
1997	2.721.539
1998	2.724.809
1999	2.730.391
2000	2.732.412
2001	2.732.923
2002	2.737.370

Se desea analizar la situación de esta enfermedad en los años 1996 y 2002.

Manejo del submódulo y solución del ejercicio

Este submódulo permite representar gráficamente, para una enfermedad, corredores endémicos por semanas o cuatrisemanas, y corredores endémicos acumulativos. Los datos pueden introducirse desde el teclado o importarse en formato Dbase, Excel o Access.

Cuando la entrada de datos se realiza de forma manual, hay que indicar al programa cuál es el número de años del período histórico (5, 6 ó 7) y cual es el año actual, introducir los tamaños poblacionales para cada uno de los años en estudio y cargar los casos declarados a lo largo del período histórico.

Para realizar los cálculos a partir de datos procedentes de archivos importados, Epidat 3.1 necesita que éstos tengan una estructura determinada. Concretamente, en este submódulo la tabla debe tener al menos 3 variables: dos para definir el período temporal, año por un lado y semana o cuatrisemana por otro, y otra que contenga el número de casos declarados de la enfermedad en estudio (véase la Tabla 5). Una vez identificadas las variables, hay que indicar al programa cuál es el número de años del período histórico (5, 6 ó 7) y cuál es el año actual y cargar los datos. La tabla con los tamaños poblacionales anuales del área geográfica que se analiza debe cubrirse siempre de forma manual.

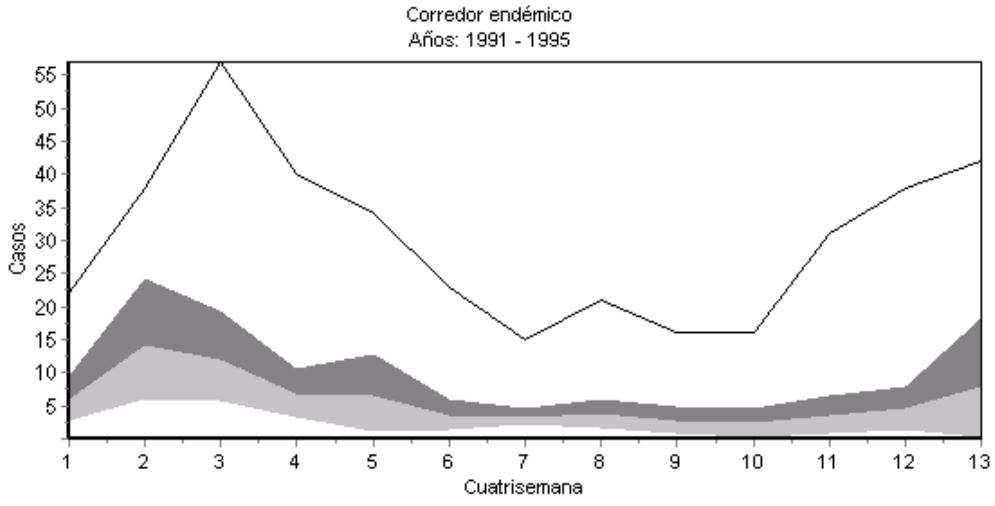
Para analizar un año diferente del mismo archivo solo hay que cambiar el valor del año actual y volver a cargar los datos.

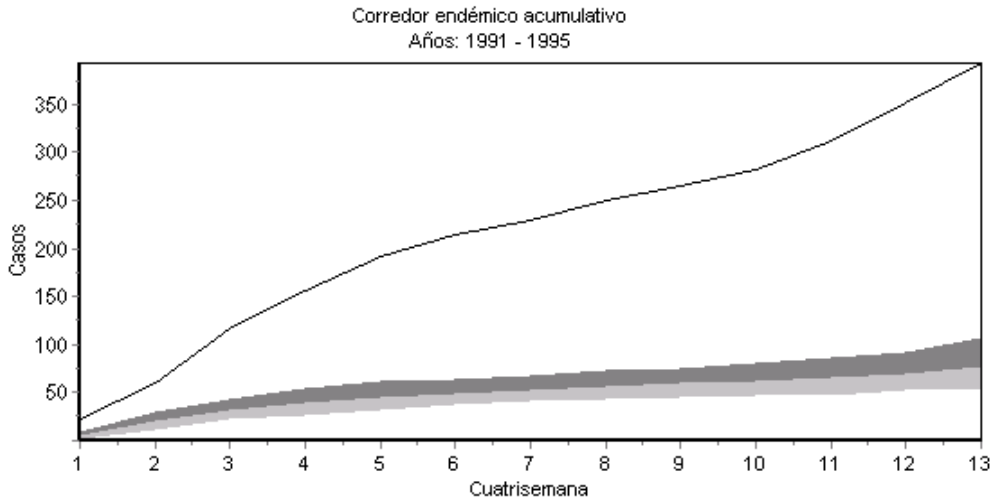
Tabla 5. Formato de tabla preparada para importar datos desde Epidat 3.1 en el submódulo de Corredores endémicos.

AÑO	CUATRISEMANA	CASOS
1991	1	9
1991	2	21
...	...	...
1991	13	11
1992	1	3
1992	2	20
...	...	...
1992	13	20
...	...	...
2002	1	16
2002	2	10
...	...	...
2002	13	5

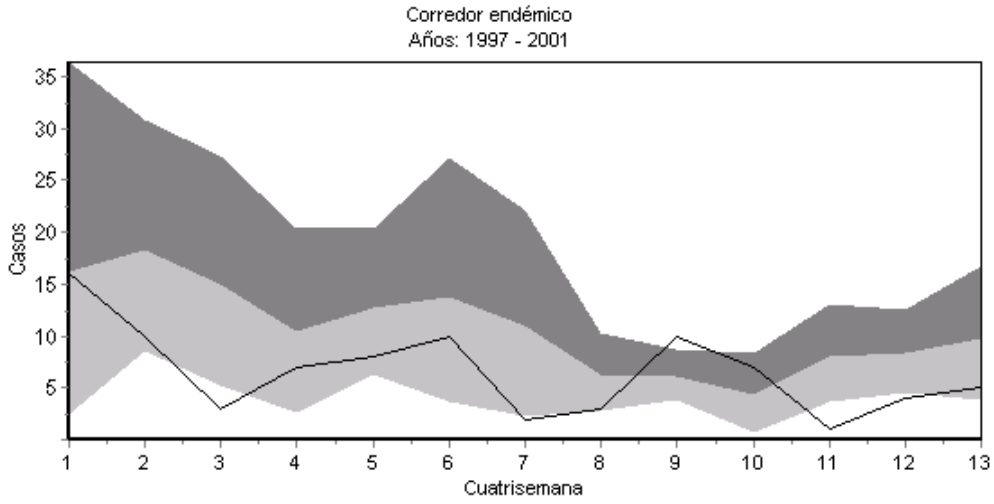
Resultados del ejercicio con Epidat 3.1

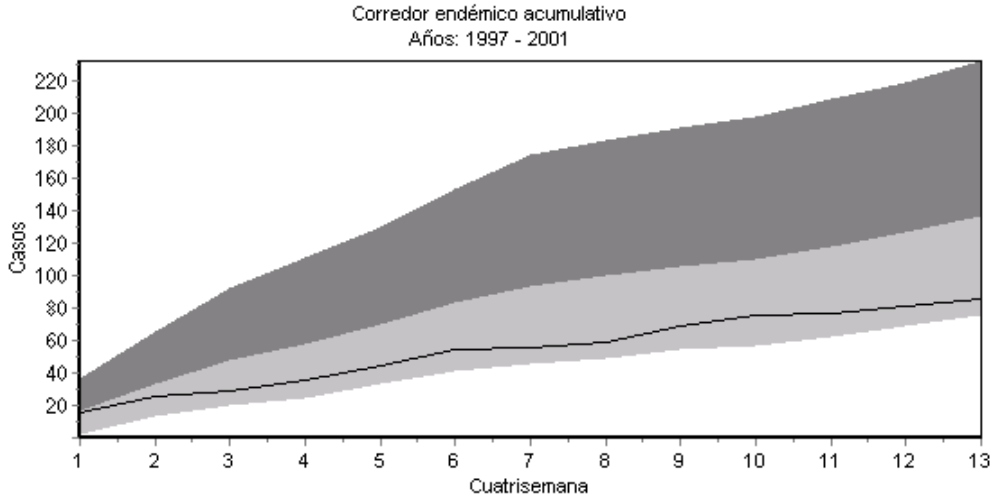
```
Vigilancia: Gráfico de corredores endémicos

Archivo de trabajo: C:\Archivos de programa \Epidat 3.1 \Ejemplos
\Vigilancia en salud pública \MENINGO91-02.xls
Campo que identifica:
  Año: AÑO
  Período: CUATRISEMANA
  Casos: MENINGO

Nivel de confianza:  95,0%

Número de años: 5
Año actual    : 1996
```





Año actual : 2002





Los corredores obtenidos para cada uno de los dos años, 1996 y 2002, son bien diferentes. En el primer caso, la serie cuatrisemanal de casos declarados se encuentra por encima del límite superior del corredor (es decir, en la zona epidémica). Esto es debido a que en el año 1996 hubo en Galicia una epidemia de enfermedad meningocócica que se manifiesta en que la incidencia de la enfermedad en ese año estuvo muy por encima de lo esperado según la

evolución de los 5 años previos. Esto ya no se observa en el año 2002, donde la enfermedad evoluciona en la zona de seguridad, salvo en las cuatrisemanas 9 y 10 en las que se produce un pico de incidencia.

Bibliografía

1. Bortman M. Elaboración de corredores o canales endémicos mediante planillas de cálculo. [Informe especial]. *Rev Panam Sal Pub* 1999; 5(1): 1-8.
2. Gómez BC. Corredores endémicos con media geométrica y su intervalo de confianza: Una nueva y eficiente alternativa para la vigilancia. *Reporte Técnico de Vigilancia* 2000; 5(4).

ONDAS EPIDÉMICAS

Conceptos generales

Modelos Susceptibles-Infecciosos–Inmunes: generación de una onda epidémica[1-3]

La dinámica de difusión de una enfermedad infecciosa en una población puede modelarse mediante la descripción del flujo de individuos entre los tres estados: susceptibles, infecciosos e inmunes o recuperados (modelos SIR). Este flujo puede describirse fácilmente mediante un sistema de ecuaciones diferenciales que permiten estimar dos parámetros:

- El coeficiente de transmisión β: parámetro resultado de multiplicar la tasa de contacto entre individuos (c) y la probabilidad de transmisión del agente infeccioso en cada contacto (p): $\beta = p \times c$
- La tasa de recuperación γ, es decir, la velocidad con la que los individuos infecciosos pasan al estado de inmunes.

Tales ecuaciones parten de dos asunciones principales:

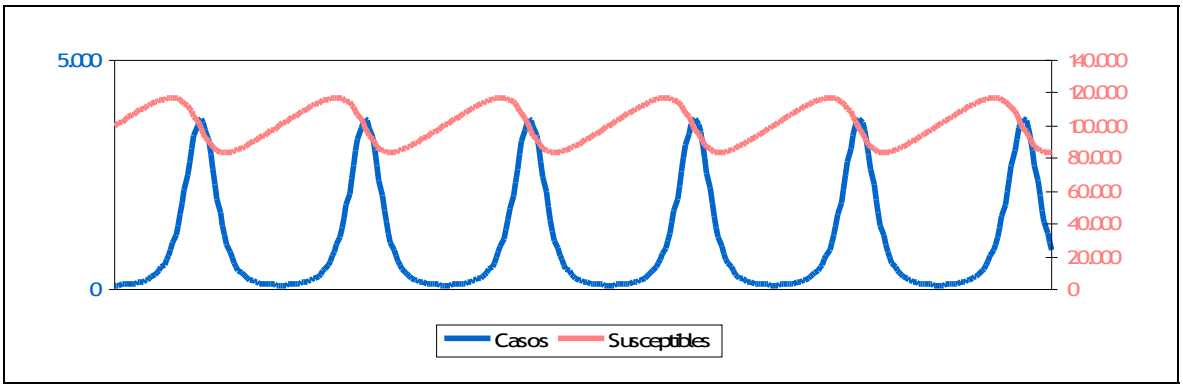
- La población se mezcla de manera homogénea.
- La población es cerrada, es decir, no se producen entradas (nacimientos o inmigraciones) ni salidas (muertes o emigraciones) de individuos.

Tales supuestos se cumplen raramente en la realidad, pero definen un modelo simple y de fácil comprensión, al que pueden añadirse parámetros que permitan reproducir el comportamiento de la realidad en estudio con mayor fidelidad.

De hecho, este modelo simple explica por qué las enfermedades infecciosas se manifiestan en forma de ondas epidémicas. La aparición de una epidemia requiere de la superación de un nivel umbral de individuos susceptibles en la población. Superado éste, se produce una epidemia, que se difunde hasta agotar los susceptibles. Cuando éstos comienzan a escasear, la probabilidad de que un individuo infectado entre en contacto con un susceptible es cada vez menor, y la epidemia acaba por extinguirse cuando la mayoría de los susceptibles han pasado del estado de infecciosos al de inmunes. Después de cada onda epidémica, quedan siempre individuos susceptibles en la población. También quedan individuos infectados pero, dado que representan una proporción muy pequeña de la población, la probabilidad de contacto entre infeccioso y susceptible es muy baja y, la enfermedad sólo se transmite ocasionalmente.

La incorporación de nuevos susceptibles a la población (nacimientos o inmigración) permite que se alcance de nuevo el nivel de susceptibles necesario para que se desencadene una epidemia, iniciándose de nuevo el ciclo (Figura 1).

Figura 1. Modelación del flujo de individuos susceptibles, infecciosos e inmunes frente a una enfermedad infecciosa en una población.



Análisis de la onda epidémica

Las ondas epidémicas se componen de una sucesión de curvas sinusoidales o curvas epidémicas. La curva epidémica se puede describir con las medidas de tendencia central y de dispersión que dan una idea del comportamiento de la epidemia y, por tanto, pueden utilizarse para caracterizar el fenómeno de propagación del agente infeccioso en la población de estudio[4,5]:

- El **retardo medio ponderado** ($\bar{t}$) es un estimador del tiempo medio de infección individual; es decir, del intervalo temporal que en promedio transcurrirá desde la aparición del primer caso en la población hasta que un individuo al azar llegue a infectarse. Cuanto mayor sea su valor, más larga será la duración de la epidemia.

$$\bar{t} = \frac{\sum_{i=1}^{N} t_i n_i}{\sum_{i=1}^{N} n_i}$$

  donde N es el número total de períodos desde que comienza la onda epidémica hasta que finaliza, t_i es el período temporal i-ésimo y n_i es el número de casos registrados en el tiempo t_i.

- La **desviación estándar** (s) mide la dispersión en la aparición de los casos con respecto al tiempo. Así, cuanto mayor sea el espaciamiento temporal entre la aparición de casos sucesivos, más tiempo tardará en propagarse la onda epidémica.

$$s = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \left(t_i - \bar{t}\right)^2}$$

- La **asimetría** (A) refleja las diferencias en la forma de aparición de los casos antes y después de alcanzarse el pico de la epidemia:
  - Si se produce un gran incremento en el número de casos al inicio del brote y éste se extingue lentamente, la asimetría es positiva.
  - Si los casos aparecen diseminados en el tiempo al inicio de la epidemia, alcanzándose lentamente el pico de ésta, la asimetría es negativa.

$$A = \frac{m_3}{s^3} \quad \text{donde} \quad m_3 = \frac{\sum_{i=1}^{N} (t_i - \bar{t})^3}{N}$$

- La **curtosis** (C) representa la concentración temporal en la aparición de casos, esto es, el apuntamiento de la curva epidémica (con respecto a una curva normal):
- Si los casos aparecen agregados en torno a la fecha de aparición del caso medio, la curtosis tendrá un valor próximo a tres.
- Si, por el contrario, los casos se generan de forma diseminada en el tiempo el valor de la curtosis será próxima a cero.

$$C = \frac{m_4}{s^4} - 3 \quad \text{donde} \quad m_4 = \frac{\sum_{i=1}^{N} (t_i - \bar{t})^4}{N}$$

Dado que las diferentes poblaciones no son homogéneas en cuanto al número y distribución de susceptibles, sólo son directamente comparables entre sí aquellos parámetros adimensionales, como la asimetría y la curtosis. Para comparar ondas epidémicas con unidades de medida diferentes se puede utilizar el coeficiente de variación (CV), calculado como el cociente entre la desviación estándar y la media, multiplicado por 100:

$$CV = \frac{s}{\bar{t}} \times 100$$

Velocidad de difusión de una onda epidémica

Otra forma de caracterizar el comportamiento de una curva epidémica es estimar la velocidad con la que aparecen los casos. Esta velocidad puede calcularse a partir de los datos que proporciona sistemáticamente la vigilancia epidemiológica, definiendo una variable P_t que represente la proporción de casos acumulados hasta el instante t. P_t responde a un modelo logístico (Figura 2) cuya ecuación es:

$$P_t = \frac{1}{1 + e^{a - bt}}$$

que expresa una relación lineal entre el tiempo y el logit de P_t:

$$Ln\left(\frac{1}{P_t} - 1\right) = a - bt$$

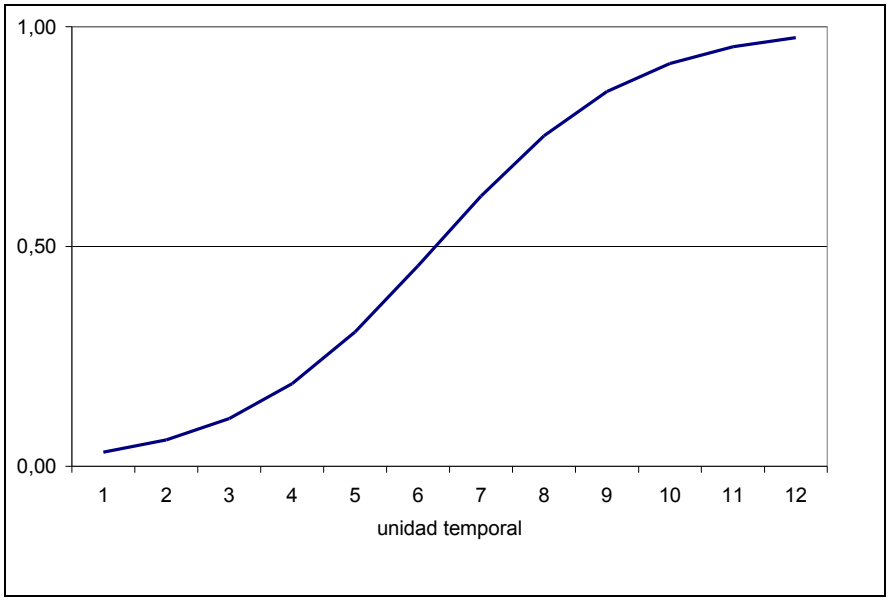
Interesa conocer el valor del parámetro b, llamado **coeficiente de difusión**, y que corresponde a la pendiente o tasa de cambio, por unidad temporal, del término:

$$Ln\left(\frac{1}{P_t} - 1\right)$$

En base a esta ecuación, el coeficiente de difusión puede interpretarse como la proporción de casos (respecto al total que integra la curva epidémica) que aparecen por unidad de tiempo.

La derivada de la variable P_t, que corresponde a la función de densidad de la variable "casos", es la velocidad de difusión de la onda.

Figura 2. Proporción acumulada de casos: modelo logístico de generación y extinción de la epidemia.



El valor del coeficiente de difusión es mayor para aquellas ondas de mayor velocidad de expansión que, generalmente, son las correspondientes a brotes de periodo epidémico breve. Dado que es adimensional, el coeficiente de difusión puede utilizarse para comparar la velocidad de una onda en distintas poblaciones. Sin embargo, su valor está determinado por la proximidad temporal en la aparición de casos, independientemente de su número. Por tanto, las ondas mas rápidas son aquellas integradas por pocos casos que han aparecido muy próximos en el tiempo, es decir, en un lapso reducido.

Utilidades del método e interpretación de resultados

El estudio de las ondas epidémicas puede aplicarse al estudio de brotes, o a la caracterización del fenómeno de difusión de una enfermedad en una población, a partir de los datos de incidencia proporcionados por los sistemas de vigilancia epidemiológica de manera sistemática.

Los parámetros de tendencia central y de difusión de la onda epidémica y el coeficiente de difusión deben considerarse de forma conjunta para interpretar cómo se ha producido el fenómeno de difusión de la enfermedad considerada, dado que proporcionan información complementaria.

Tasa de propagación

En el caso de las enfermedades infecciosas de transmisión persona a persona, se puede calcular, a partir del número de casos notificados de manera sistemática, la tasa de propagación. Se trata de un estimador de la intensidad de transmisión de la enfermedad en la población, que se calcula como la razón de casos secundarios y casos primarios en un instante determinado.

En concreto, la tasa de propagación para cada período t se estima como el cociente entre dos medias móviles (calculadas con 3 casos): la correspondiente al período $t+k$, siendo k el período medio de incubación, entre la correspondiente al período t. Se asume, por tanto, que los casos notificados en un instante son secundarios a los notificados en el instante precedente, si el intervalo de tiempo transcurrido es igual al periodo de incubación.

La representación gráfica de la curva de la tasa de propagación en función del tiempo permite describir el comportamiento de ésta y establecer en qué momento la transmisión del agente infeccioso en estudio está aumentando en la población.

Recomendaciones y advertencias

- Para el cálculo de los parámetros de difusión de una onda epidémica es fundamental definir las fechas de inicio y final de ésta y, consecuentemente, el número de casos que la componen. Tales fechas son relativamente fáciles de establecer en el estudio de brotes, pero no ocurre así cuando se estudian las series temporales aportadas por la vigilancia epidemiológica. Es necesario definir un valor a partir del cual el incremento en el número de casos en dos períodos consecutivos indique el inicio de la onda y, para ello, la razón entre el número de casos de dos años consecutivos puede servir como indicador. Sin embargo, no es posible definir un valor para esta razón superado el cual pueda afirmarse que se ha iniciado una situación epidémica, pues éste puede variar mucho entre las distintas poblaciones. Una alternativa para la definición del inicio de una onda epidémica se puede consultar en Guatire y Richardson[6].

- El coeficiente de difusión puede considerarse el mejor estimador de la velocidad de difusión de la onda. Sin embargo, el cálculo de los restantes parámetros de difusión incorpora en el análisis comparativo el número de casos que integran la onda y, por tanto, describe la forma de ésta. Para estudiar la dinámica de transmisión de una enfermedad infecciosa en una población, es conveniente calcular y valorar todos los parámetros conjuntamente, pues aportan informaciones complementarias.
- El cálculo de los parámetros de difusión de una onda sólo puede realizarse a posteriori, es decir, una vez la onda se ha extinguido en una población. Este método no tiene, por tanto, valor predictivo.
- Tanto el cálculo de los parámetros de difusión de una onda epidémica como el de la tasa de propagación, puede partir del número de casos incidentes o de la tasa de incidencia. La elección de una u otra variable depende del objetivo del estudio y de la información de que se disponga. La utilización de tasas de incidencia permite comparar entre curvas de distintas poblaciones.

Ejercicio

Se quiere estudiar un brote epidémico de una enfermedad de transmisión persona a persona cuyo periodo de incubación fue de 3 días. El brote se ha iniciado el 14 de abril y ha finalizado el 27 de mayo de 1998 y los datos están disponibles en el archivo BROTE.xls incluido en Epidat 3.1.

Se desea:

1. Representar gráficamente la curva epidémica del brote.
2. ¿Cuáles son las medidas de tendencia central y dispersión de la variable que definen la curva epidémica? ¿Cuál es su coeficiente de difusión?.

Manejo del submódulo y solución del ejercicio

Este submódulo permite representar gráficamente la curva epidémica de una enfermedad y estima los parámetros que miden la difusión de la onda. Los datos pueden introducirse desde el teclado o importarse en formato Dbase, Excel o Access.

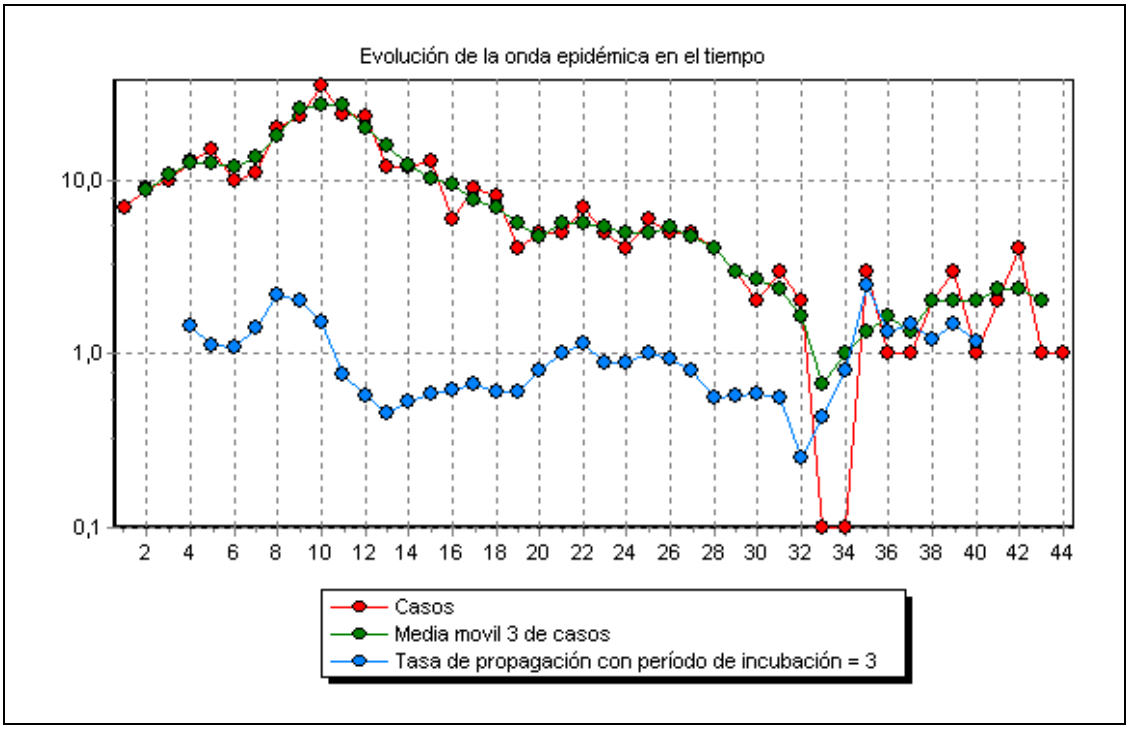
Cuando la entrada de datos se realiza de forma manual, hay que indicar al programa cuál es la unidad temporal empleada (días, semanas, cuatrisemanas, meses o años), el período medio de incubación, el total de períodos analizados, y definir la fecha de inicio del período. Con esta información se genera una tabla en la que deben introducirse los casos ocurridos en cada período.

Para realizar los cálculos a partir de datos procedentes de archivos importados, Epidat 3.1 necesita que estos tengan una estructura determinada. Concretamente, en este submódulo la tabla debe tener una variable que contenga el número de casos de la enfermedad en estudio, y una o dos variables, dependiendo del caso, para definir el período temporal, por ejemplo, año y semana (véase el archivo BROTE.xls). Una vez identificadas las variables, se realiza la carga de datos.

Resultados con Epidat 3.1 con los datos del brote

```
Vigilancia: Ondas epidémicas

Archivo de trabajo: C:\Archivos de programa \Epidat 3.1 \Ejemplos
\Vigilancia en salud pública \BROTE.xls
```

```
Campo que contiene:
  Casos observados: CASOS
  Fecha: FECHA


   Período medio de incubación:       3 Días
            Número de períodos:      44
            Nivel de confianza:    95,0%


Medidas no adimensionales        Valor
-------------------------   ----------
Tiempo medio de infección      13,9027
      Desviación estándar       9,3647


   Medidas adimensionales        Valor
-------------------------   ----------
 Coeficiente de asimetría       1,2257
  Coeficiente de curtosis       4,1411
 Coeficiente de variación      67,3594


Coeficiente de difusión
     Valor        IC (95%)
---------- ----------------------
    0,1678   0,1623   0,1733
```



Bibliografía

1. Anderson RM, May RM. *Infectious Diseases of Humans: dynamics and control*. Oxford: Oxford University Press; 1991.

2. Giesecke J. *Modern Infectious Disease Epidemiology*. Ed Arnold; 1994.

3. Halloran ME. Chapter 4. En: Weber T. Editor. *Epidemiologic methods for the study of infectious diseases*. Oxford: Oxford University Press; 2001: 56-85.

4. Cliff A, Haggett P. Methods for the measurement of epidemic velocity from time-series data. *Int J Epidemiol* 1982; 11(1): 82-9.

5. Cliff A, Haggett P. Disease diffusion: the spread of epidemics as a spatial process. Chapter 4. En: McGlashan ND. Editor. *Medical Geography: Techniques and field studies*. London. p. 94-122.

6. Guatire L, Richardson S. A time series construction of an alert threshold with application to *S. Bovis morbificans* in France. *Stats Med* 1991; 10: 1493-509.

EFECTIVIDAD VACUNAL

Conceptos generales

Toda intervención vacunal en una población, sea continuada como la de los programas de vacunaciones, u ocasional, como cuando se impulsan campañas, ha de producir un efecto en la frecuencia de la enfermedad para cuyo control se realiza. Dicho muy sucintamente, la intervención vacunal ha de hacer que la incidencia de la enfermedad sea menor que la que habría sido si no se hubiese intervenido; y esa diferencia entre lo que es y lo que habría sido se denomina **efectividad vacunal** o **efectividad de la vacunación**.

Pero definir concretamente en qué consiste el efecto de una intervención vacunal no es sencillo; y no lo es, tanto por las características epidemiológicas de la mayoría de las enfermedades hoy día controladas mediante vacunación, que son a las que se ceñirá cuando en adelante se hable de "enfermedad", como por las propiedades de ciertas vacunas, que no sólo disminuyen o anulan la susceptibilidad en los individuos que las reciben, sino que son capaces también de interferir en la transmisión de la infección (es decir, hacen que los individuos vacunados transmitan la infección menos fácilmente que los no vacunados).

Recuérdese que, dicho muy simplemente, un miembro de una población adquiere una enfermedad transmisible si: (1) es susceptible y (2) estuvo expuesto a la infección; de modo que, cuantos más miembros de la comunidad estén en condiciones de transmitir la infección, más fácil será que un susceptible se exponga a ella. Así pues, debido a que la exposición de los susceptibles de una comunidad depende de la prevalencia de personas infectadas que en ella hay, la incidencia de enfermedad en un momento concreto en dicha comunidad depende de la que hubo en un momento anterior.

En un artículo publicado en 1991, Halloran y Struchiner[1] muestran cómo, en situaciones como ésta en la que hay dependencia de sucesos, si existe una intervención que puede afectar tanto al suceso (susceptibilidad/enfermedad) como a la relación de dependencia (transmisión de la infección), se pueden describir cuatro efectos de la intervención: directo, indirecto, total y poblacional, para los que dieron unas definiciones, que aquí se recogen adaptadas al campo de la vacunación:

El **efecto directo** de una vacuna en un individuo vacunado es la diferencia entre el resultado de enfermedad observado en el individuo vacunado y el que se habría observado en ese mismo individuo si no se hubiese vacunado, permaneciendo igual el resto de las cosas.

El **efecto indirecto** de una intervención vacunal en un individuo es la diferencia entre el resultado de enfermedad observado en un individuo no vacunado que es miembro de la comunidad en que se realizó la intervención vacunal, y el que se habría observado en el individuo –también no vacunado– de haber pertenecido a una comunidad comparable pero en la que no se realizó la intervención.

El **efecto total** de una intervención vacunal en un individuo es la diferencia entre el resultado de enfermedad observado en un individuo vacunado que es miembro de la comunidad en la que se realizó la intervención vacunal y el que se habría observado en el individuo –también vacunado– de haber pertenecido a una comunidad comparable pero en la que no se realizó la intervención.

El **efecto poblacional** de una vacuna y una intervención vacunal es la diferencia entre el resultado de enfermedad observada en un individuo medio de la comunidad donde se realizó la intervención vacunal, y el observado en un individuo medio de una comunidad donde no realizó la intervención.

De las definiciones anteriores se infiere que los tres primeros efectos ocurren en individuos: el directo debido a la vacuna, que es capaz de reducir o anular la susceptibilidad a la

infección en los que la reciben; el indirecto debido a la propia intervención vacunal, que disminuye las posibilidades de exposición de los susceptibles no vacunados; y el total debido a la combinación de la vacuna y la intervención vacunal, en unos por la reducción de la susceptibilidad que provoca la vacuna y en otros –los fallos vacunales– porque disminuyen las posibilidades de exposición debido a la intervención. El efecto restante, el poblacional, que ocurre en la población misma debido tanto a la vacuna como a la intervención vacunal, es el que tiene verdadero interés para la salud pública cuando la vacuna que se emplea es capaz de producir un efecto indirecto. Y tanto más, cuanto la intensidad del efecto indirecto depende de la extensión que alcance la intervención, en este caso de la cobertura vacunal.

Así mismo, por estar construidas como contrafácticos, de las definiciones se deducen con facilidad las bases del diseño de los estudios con los que se puede estimar cada uno de los efectos. Para estimar efectos directos se compararán los resultados de enfermedad observados en individuos vacunados y no vacunados de una misma población con intervención vacunal, y para estimar el resto de efectos se comparan los resultados de enfermedad observados en una población con intervención vacunal con los observados en una población que carece de ella, que aquí se llamará control. Más concretamente, en todos los casos un elemento de la comparación es la población control, y el otro procede de la población con intervención: los no vacunados, para estimar el efecto indirecto; los vacunados, para el total; y todos, vacunados y no vacunados, para el poblacional.

Recomendaciones

Si bien las propias definiciones de los efectos orientan sobre el diseño de los estudios con que han de ser estimados, no ocurre lo mismo con las cláusulas de identidad entre individuos y la comparabilidad de poblaciones que en ellas se prescriben para evitar que esas estimaciones resulten sesgadas. En este sentido, además de las precauciones habituales en los estudios observacionales[2,3], hay que adoptar algunas debidas al tipo de intervención realizada.

En los individuos habrá que poder asumir que los factores que determinan la susceptibilidad a la infección se distribuyen de igual modo en vacunados y no vacunados (asunción que se suele referir como "vacunación aleatoria"); y habrá que asumir también que, para vacunados y no vacunados, la exposición a la infección es semejante (asunción conocida como "mezcla aleatoria" cuando la transmisión se reduce al contacto entre personas). En general, es esto último lo más difícil de aceptar, puesto que las diferencias en susceptibilidad suelen poder corregirse mediante estratificación (por ejemplo, las asociadas a la edad, al número de dosis recibidas o a la pérdida de inmunidad por el paso del tiempo), mientras no suele haber fácil remedio para las diferencias en exposición, que pueden ser debidas, por ejemplo, a que dentro de la población hay amplios grupos con niveles de cobertura vacunal diferente y por ello, si la vacuna produce efectos indirectos, con intensidades también diferentes en la interrupción de la transmisión de la infección, habrá una sobreexposición de los no vacunados; o, al contrario, puede producirse una sobreexposición de vacunados si el haberse vacunado produce una falsa sensación de protección; o, simplemente, porque de forma "aleatoria" la muy baja prevalencia de infección en la población permitió que la exposición se repartiese de modo muy diferente en vacunados y no vacunados[4].

De todas formas, siempre que la transmisibilidad de la infección sea elevada, existe un modo de estimar los efectos que ocurren en los individuos condicionado a su exposición a la infección: consiste en aprovechar aquellos lugares donde se reúnen pocas personas pero que mantienen relaciones muy estrechas, como ocurre en hogares o aulas, de modo que todos ellos habrán estado expuestos a la infección. Los estudios en estos lugares de mezcla de tamaño pequeño permitirán, pues, asumir que la exposición de vacunados y no vacunados es semejante y estimar los efectos de la vacuna sobre la susceptibilidad, sobre la transmisibilidad y sobre ambas combinadas. Este método ofrece, por tanto, información muy

útil sobre el comportamiento de la vacuna, y aunque con él no se pueden estimar efectos en poblaciones, sus resultados mejoran la interpretación dada a los de estudios poblacionales en los que hay problemas para asumir la igualdad de exposición en vacunados y no vacunados.

Por su parte, la comparación de poblaciones hace referencia a que ambas, la que fue objeto de intervención y la que sirve de control, tengan experiencias de enfermedad semejantes, que a su vez sean reflejo de distribuciones también semejantes de los factores que determinan su transmisión en la población; y hace referencia también a que entre ellas, hablando en términos de la transmisión de la enfermedad, no hay relación alguna. Para disponer de una población que reúna estas características se puede recurrir a alguna muy poco comunicada con la población objeto de intervención, o, lo que es más común, a esta misma población antes y después de la intervención. En este último caso es necesario elegir periodos que sean suficientemente largos como para garantizar la estabilidad en el comportamiento de la enfermedad, y suficientemente cortos para que entre el antes y el después no hayan variado los factores sociales que determinan la transmisión de la enfermedad en la población.

Un modo alternativo para obtener una población control es construir con modelos matemáticos lo que habría sido el comportamiento de la enfermedad en la población si no se hubiese realizado la intervención. Pero esta metodología sobrepasa los límites de Epidat 3.1.

Parámetros

Son varios los parámetros con los que se pueden estimar cada uno de estos cuatro efectos, de los que Epidat 3.1 incluye, para estimaciones no condicionadas a la exposición, la tasa de incidencia y la incidencia acumulada. Las efectividades asociadas a cada uno de los efectos se calculan como indica la Tabla 6, que fue tomada de Halloran et al[5], y se interpretan como sigue: la efectividad directa, como la proporción de enfermedad evitada en vacunados por estar vacunados; la efectividad indirecta, como la proporción de enfermedad evitada en no vacunados por residir en una comunidad con intervención vacunal; la efectividad total, como la proporción de enfermedad evitada en vacunados por estar vacunados y residir en una comunidad con intervención vacunal; y la efectividad poblacional, como la proporción de enfermedad evitada en una comunidad por la intervención vacunal.

Y el parámetro que Epidat 3.1 incluye para las estimaciones condicionadas a la exposición es la tasa de ataque secundario, con la que se mide, del modo que se indica también en la Tabla 6, el efecto de la vacuna sobre la susceptibilidad, la transmisibilidad y ambas combinadas, que tienen una interpretación semejante a las anteriores efectividades directa, indirecta y total, respectivamente.

Los intervalos de confianza los calcula Epidat 3.1 restando de 1 los valores de los límites del intervalo de confianza para el riesgo relativo, la razón de tasas de incidencia o el cociente entre las tasas de ataque secundario, según el caso.

Tabla 6. Las diferentes efectividades vacunales.

Parámetro	Directa	Indirecta	Total	Poblacional
Incidencia acumulada (I)	$1-I_{I1}/I_{I0}$	$1-I_{I0}/I_c$	$1-I_{I1}/I_c$	$1-(fI_{I1}+(1-f)I_{I0})/I_c$
Tasa de incidencia (T)	$1-T_{I1}/T_{I0}$	$1-T_{I0}/T_c$	$1-T_{I1}/T_c$	$1-(fT_{I1}+(1-f)T_{I0})/T_c$

	Susceptibilidad	Transmisión	Combinada
Tasa de ataque secundario (S)	$1–S_{01}/S_{00}$	$1–S_{10}/S_{00}$	$1–S_{11}/S_{00}$
Subíndices en T e I: población (intervención -I- y control -C-) y status vacunal. Subíndices en S: estatus vacunal del caso índice y de los expuestos. Subíndices del status vacunal: vacunado (1) y no vacunado (0). Cobertura vacunal = f Por ejemplo, S_{01} es la tasa de ataque en hogares con caso índice no vacunado y expuestos vacunados.			

Ejercicio

Durante las últimas décadas, la enfermedad E producía epidemias bianuales en las comunidades A y B. Ambas comunidades, aunque se podían considerar aisladas una de otra en términos de la transmisión de E, mostraban en cambio distribuciones semejantes de los factores que determinan dicha transmisión. También ambas contaban con un sistema de vigilancia semejante, que permitió describir lo que se consideraba un patrón constante de la distribución de la incidencia por edad de E durante las epidemias: era de un 10% en los que se encontraban en los dos primeros años de la vida, y de un 50% entre los que se encontraban en el 3º y 4º, niños éstos que acudían todos a centros de preescolar.

Para controlar la enfermedad, las dos comunidades decidieron incluir en el calendario de vacunaciones una vacuna frente a E, la comunidad A la vacA® y la B la vacB® (en ambos casos, la única dosis de vacuna se administra a los 12 meses de edad y la eficacia estimada en ensayos clínicos es del 80%); y, pasados unos años en los que no se modificaron los determinantes sociales de la transmisión de E, en las comunidades se produjo una epidemia que fue utilizada para evaluar la efectividad de la vacunación. Para ello se emplearon dos estudios diferentes, uno en los centros de preescolar y el otro en los hogares de los niños.

Tabla 7. Resultado del estudio en preescolares al final de la epidemia, durante la cual no hubo ninguna intervención orientada a su control.

Alumnos sin antecedentes de E al inicio de la epidemia		Casos de E	
		Comunidad A	Comunidad B
Vacunados	8.000	800	400
No vacunados	2.000	1.000	500

Tabla 8. Resultado del estudio en hogares, tomando como expuestos los niños de menos de 2 años de edad.

	Comunidad A				Comunidad B			
	CI Vacunado		CI No vacunado		CI Vacunado		CI No vacunado	
	Casos	Exp.	Casos	Exp.	Casos	Exp.	Casos	Exp.
Vacunados	10	52	8	40	5	54	8	42
No vacunados	45	50	54	60	21	48	50	56

Nota: CI- Caso índice

Aunque tanto la razón de casos vacunados y no vacunados, 0,8, como la cobertura vacunal (80%) es igual en ambas comunidades, la simple inspección de la tabla sugiere que E se comportó de forma distinta en ambas comunidades. Asumiendo que ambos estudios tienen la misma calidad, explique las diferencias observadas en términos de efectividad vacunal. (En general, el término eficacia se emplea cuando el efecto de la vacuna se midió en condiciones "ideales", como las de un ensayo clínico, y el de efectividad se reserva para

efectos medidos en condiciones "reales", como las de un programa o campaña de vacunación).

Solución del ejercicio

En el estudio con los preescolares, las efectividades se estiman con independencia de la exposición y empleando la experiencia previa con E como población control: con una incidencia del 50%, el número de casos que hubieran ocurrido en la población de 10.000 alumnos, bajo la experiencia de la época prevacunal, sería de 5.000.

Resultados con Epidat 3.1

Efectividad vacunal: Población A (preescolares)

Tipo de modelo:	No condicionado a la exposición
Tipo de datos:	Incidencia acumulada
Nivel de confianza:	95,0%

Efectividad vacunal	Valor	IC (95%)	
Directa	0,8000	0,7805	0,8178
Indirecta	0,0000	-0,0703	0,0656
Total	0,8000	0,7845	0,8144
Poblacional	0,6400	0,6201	0,6589

Efectividad vacunal: Población B (preescolares)

Tipo de modelo:	No condicionado a la exposición
Tipo de datos:	Incidencia acumulada
Nivel de confianza:	95,0%

Efectividad vacunal	Valor	IC (95%)	
Directa	0,8000	0,7719	0,8246
Indirecta	0,5000	0,4519	0,5439
Total	0,9000	0,8893	0,9097
Poblacional	0,8200	0,8068	0,8323

Efectividad vacunal: Población A (hogares)

Tipo de modelo:	Condicionado a la exposición
Nivel de confianza:	95,0%

Efectividad vacunal	Valor	IC (95%)	
Susceptibilidad	0,7778	0,5846	0,8811
Trasmisibilidad	0,0000	-0,1333	0,1176
Conjunta	0,7863	0,6247	0,8784

Efectividad vacunal: Población B (hogares)

Tipo de modelo:	Condicionado a la exposición
Nivel de confianza:	95,0%

Efectividad vacunal	Valor	IC (95%)	
Susceptibilidad	0,7867	0,5994	0,8864
Trasmisibilidad	0,5100	0,3161	0,6489
Conjunta	0,8963	0,7598	0,9552

En este ejemplo, evidentemente hipotético, se presenta una situación que de ser real habría traído por la calle de la amargura al responsable del programa de vacunaciones de A. ¿Cómo -se preguntaría- usando una misma vacuna y habiendo conseguido la misma cobertura vacunal que B se tiene el doble de casos que ellos?

Después de realizar el estudio en preescolares, "perfectamente" representativo de cada una de las poblaciones, la respuesta empieza a tomar forma. En principio, ambas vacunas mostraron la eficacia prevista, un 80% (es el valor de las efectividades directas), pero mientras vacA® en nada afectó la transmisión de E (la efectividad indirecta en A es nula y la efectividad total igual a la directa), vacB® la redujo a la mitad (la efectividad indirecta en B es del 50% y la total del 90%, lo que muestra que los casos debidos a fallos vacunales –que son el complementario de la efectividad directa- se redujeron a la mitad). En consecuencia, la efectividad poblacional obtenida en B con vacB® es superior (un 18%) a la obtenida en A con vacA®. Pero además, esta efectividad se puede traducir en casos evitados (que es el objetivo de la intervención): en B se evitaron el 82% de los casos previstos (incidencia esperada=50%, incidencia observada = 9%), y en A sólo el 64% (incidencia observada = 18%).

Obsérvese que hasta ahora sólo se emplearon los resultados del estudio en preescolares ¿De qué sirve, además de para ilustrar las posibilidades de Epidat 3.1, haber realizado el estudio en hogares? Para tener una opinión acerca de la validez de las estimaciones proporcionadas por el estudio en preescolares, que es el único modo de cuantificar el efecto poblacional de la vacunación sin tener que recurrir a mayores asunciones. Y en este caso, los resultados de la efectividad frente a la susceptibilidad y a la transmisión alientan a suponer que los datos poblacionales fueron estimados en una población en que vacunados y no vacunados tuvieron una exposición semejante a E y con una población control adecuada en términos de aquellos factores que determinan en ella la transmisión de E.

Huelga decir, para finalizar, que el responsable del programa de vacunaciones en A ya tiene su respuesta, a la que sigue como siempre una nueva pregunta, ¿Cómo –le preguntan- elegiste la vacA®?

Bibliografía

1. Halloran ME, Struchiner CJ. Study Designs for Dependent Happenings. *Epidemiology* 1991; 2: 331-8.
2. Orenstein WA, Bernier RH, Hinman AR. Assessing Vaccine Efficacy in the Field. *Epidemiol Rev* 1998; 10: 212-41.
3. Chen T, Orenstein WA. Epidemiologic Methods in Inmunization Programs. *Epidemiol Rev* 1996; 18: 99-117.
4. Nourjah P, Frerischs RR. Minimum Attack Rate for Measuring Vaccine Efficacy. *Int J Epidemiol* 1995; 24: 834-41.
5. Halloran ME, Longini IM, Struchiner CJ. Design and interpretation of vaccine field studies. *Epidemiol Rev* 1999; 21(1): 73-88.