

# INTRODUCCIÓN DE FACTORES DE RIESGO EN LOS MÉTODOS DE KNOX Y GRIMSON PARA EL ESTUDIO DE CONGLOMERADOS ESPACIO-TEMPORALES

GLADYS CASAS CARDOSO\*      RICARDO GRAU ABALO†  
GLADYS CARDOSO ROMERO‡

*Recibido/Received: 20 May 2003*

---

## Resumen

No existen métodos estadísticos de detección de conglomerados de enfermos que puedan aplicarse en todas las situaciones, por tanto se hace necesario conocer con exactitud cuando se puede confiar en los resultados que ellos arrojan. En este trabajo se simulan casos de enfermos en el espacio por medio de dos distribuciones uniformes y en tiempo siguiendo el modelo de epidemia simple y se analiza como responden ante ellos los métodos de Knox y Grimson. Se estudia además el comportamiento de ambos tests al incorporarles factores de riesgo. Todos los cálculos fueron hechos con ayuda del paquete Mathematica.

**Palabras clave:** Clases de enfermedades, interacción espacio-tiempo, método de Knox, método de Grimson, simulación.

## Abstract

There are not infallible disease clusters methods, so it is necessary to know exactly when their results are reliable. In this paper space data were simulated using two uniform distributions and time data, following the pattern of simple epidemic model. Behaviour of both tests: Knox and Grimson were studied. Besides the conduct of both tests with risks factors were analysed. Numerical computing were done with Mathematica software.

---

\*Departamento de Matemática, Facultad de Matemática Física y Computación, Universidad Central "Marta Abreu" de las Villas, Carretera a Camajuaní, km 5 1/2, Santa Clara, Villa Clara, Cuba. E-Mail: gladita@cei.uclv.edu.cu

†Misma dirección que G. Casas. E-Mail: rgrau@uclv.edu.cu

‡Misma dirección que G. Casas. E-Mail: gcardoso@uclv.edu.cu

**Keywords:** Disease clusters, space-time interaction, Knox’s method, Grimson’s method, simulation.

**Mathematics Subject Classification:** 62P10.

## 1 Introducción

Las enfermedades infecciosas fueron en una época lejana la causa fundamental de mortalidad en grandes grupos de humanos y aún en nuestros días atacan a millones de personas en países subdesarrollados. En la actualidad, con el incremento de los viajes por aire, la población de todo el mundo está en riesgo de infestación por enfermedades contagiosas como la influenza por sólo mencionar alguna.

Desde el siglo XIV la modelación matemática de procesos infecciosos ha venido enriqueciendo el conocimiento epidemiológico y ha representado una gran ayuda en la promoción de acciones certeras para la salud pública. Algunos autores incluso plantean que resulta muy difícil comprender la dinámica de una enfermedad infecciosa sin la estructura formal de un modelo matemático.

Los avances de la Computación en las últimas décadas han contribuido notablemente al fortalecimiento la matemática y esta a su vez ha sido utilizada como una poderosa herramienta de trabajo por otras ciencias entre las que se encuentra la medicina. Varios modelos de epidemias han mostrado su utilidad en la prevención y en la aplicación de diversos procedimientos de control de enfermedades; sin embargo no existen modelos matemáticos infalibles, todos tienen sus ventajas y limitaciones y ambas deben conocerse a fin de evitar aplicaciones incorrectas que pueden ser sumamente perjudiciales en un escenario epidemiológico.

Por razones obvias no es posible efectuar experimentos con enfermedades en humanos. Los datos que se tienen son los recopilados de las epidemias naturales que han ocurrido en años anteriores, pero muy frecuentemente estos son imprecisos o están incompletos. Es por ello que los modelos matemáticos y la simulación computarizada de datos juega un papel fundamental en la realización de experimentos teóricos, en la estimación de sus parámetros y en la validación de pruebas estadísticas relacionadas con tales afecciones.

Uno de los problemas más importantes al que se enfrentan los epidemiólogos es la detección temprana de focos epidémicos. Si se detecta la presencia de un conjunto de enfermos superior a lo esperado que están cerca geográfica y temporalmente puede pensarse en el incipiente origen de una epidemia, [1]. Muchas son las técnicas desarrolladas con estos fines [2], entre ellas se encuentran la de Knox y la de Grimson.

## 2 Detección de conglomerados de enfermos con factores de riesgo

Es tarea de los epidemiólogos no solo detectar a tiempo un foco epidémico, sino también encontrar las razones que lo han provocado. El análisis clásico de los conglomerados arroja poca información acerca de los motivos que los propician, sólo indica cuando se está o no

en presencia de un exceso de casos diagnosticados. En la práctica no siempre se conocen las causas por las que se reporta la agrupación y por lo general no se tienen conocimientos suficientes del proceso para poder modelarlo, es por ello que se recomienda realizar un estudio combinado con factores de riesgo, [3].

Cada uno de los factores a considerar se tratará como una variable dicotómica: el valor 1 indicará la presencia de la característica que el mismo representa y 0 indicará su ausencia. Dicotomizar un factor de riesgo no es una tarea sencilla y en ocasiones resulta ser subjetiva por lo que se recomienda consultar la opinión de varios especialistas.

Existen dos vías diferentes de incorporar factores de riesgo a los métodos clásicos de detección de aglomeraciones. La primera es una vía general que puede aplicarse a cualquier técnica de detección de conglomerados. No modifica matemáticamente el test en sí, sino que su labor consiste en ir restringiendo los datos de entrada de manera que se consideren solo aquellos casos en los que estén presentes todos los factores seleccionados.

Si el conglomerado es verdadero y los factores caracterizan bien la enfermedad, entonces la significación de la agrupación deberá desaparecer en subgrupos homogéneos respecto a los mencionados factores, [3]. Para comprender mejor esta idea puede mencionarse un ejemplo: si todas las personas fumaran 20 cigarrillos diarios, los estudios estadísticos concluirían que el cáncer del pulmón es una enfermedad genética y en cierto sentido eso sería verdad, pues si todos los individuos están expuestos de la misma manera al agente necesario, la distribución de casos depende sólo de la susceptibilidad individual, [4].

Una de las desventajas de ese algoritmo es que con la introducción de nuevos factores la cantidad de datos disminuye. Con respecto a esto se plantea que si los factores están realmente relacionados con la enfermedad, entonces deben estar presentes en la mayoría de los enfermos, y dicha disminución no debe ser muy grande.

La segunda vía para incorporar factores de riesgo a las técnicas de detección de congregaciones es mucho más específica; es particular para cada test pues parte de cierta modificación de su fundamento matemático. A continuación se mostrarán las transformaciones específicas realizadas a los métodos de Knox y de Grimson respectivamente.

### 3 Método de Knox

El método de Knox, [5], [6] es quizás el primero reportado en la literatura para la detección de conglomerados espacio - temporales. Fue ideado por el Profesor E. G. Knox en el año 1964 y desde entonces ha sido ampliamente utilizado. Su técnica consiste en calcular distancias espaciales y temporales entre todos los pares de casos y establecer ciertas definiciones de adyacencia en espacio - tiempo. Si la distancia geográfica calculada es menor que una determinada distancia crítica prefijada de antemano, se considera que los casos están cerca, de otro modo son casos lejanos. El mismo análisis se hace con la dimensión temporal. Los resultados se exhiben en una tabla de contingencia como la mostrada en la tabla 1, donde:  $N$ : número total de casos en estudio,  $dec$ : distancia espacial crítica,  $dtc$ : distancia temporal crítica,  $A$ : representa los pares de casos que están próximos geográfica y temporalmente,  $N_S$ : representa el número total de pares de casos cercanos espacialmente, y  $N_T$ : representa el número total de pares de casos cercanos temporalmente.

Tiempo ↓ / Espacio →	Cerca ( $dist.esp < dec$ )	Lejos ( $dist.esp \geq dec$ )	Total
Cerca ( $dist.temp < dec$ )	<b>A</b>	<i>B</i>	$N_T = A + B$
Lejos ( $dist.temp \geq dec$ )	<i>C</i>	<i>D</i>	$C + D$
Total	$N_S = A + C$	$B + D$	$\frac{1}{2}N(N - 1)$

Tabla 1: Método de Knox.

Knox supone que si ambas dicotomías fueran altamente asimétricas, el número de pares que aparecen en la primera celda de la tabla **A** debe seguir una distribución de Poisson con valor esperado natural, [5], [10]:

$$E(A) = \frac{2N_S N_T}{N(N - 1)} = \lambda \quad (1)$$

Varios autores han confirmado la significación esencial de este resultado, [7]. O. Abe posteriormente realizó una revisión detallada del test y enriqueció la forma de calcular la probabilidad empírica de significación para algunos casos particulares, [8].

### 3.1 Vía particular de inclusión de factores de riesgo al método de Knox

La técnica es en esencia la misma descrita con anterioridad, solo que la tabla de contingencia se extiende en una dimensión más: la de los factores de riesgo. **A** representa ahora los pares de casos que están próximos geográfica y temporalmente y en los que están presentes todos los factores de riesgo considerados, el cálculo de su esperanza también se verá afectado y estará dado por:

$$E(A) = \frac{6N_S N_T N_F}{N(N - 1)(N - 2)} = \lambda \quad (2)$$

donde:  $N_F$ : pares de casos en los que todos los factores  $F_1, F_2, \dots, F_f$  considerados tienen el valor 1. También aquí es válido afirmar que **A** sigue una distribución de Poisson con parámetro  $\lambda$ .

La cantidad de datos permanece inalterable en este algoritmo, solo se restringe el valor de  $N_F$  que interviene en el cálculo de  $E(A)$ .

## 4 Método de Grimson

El método de Grimson, [9] parte de dividir la región espacio - temporal en estudio en otras más pequeñas a las que denomina celdas. Aquellas celdas cuyo número de casos excede el valor esperado de una distribución de Poisson o un umbral prefijado por un especialista, se denominan celdas marcadas. La hipótesis nula se describe en términos de una distribución aleatoria de esas celdas marcadas por toda la región estudiada. La hipótesis alternativa enuncia que existe una cantidad grande de celdas marcadas adyacentes. Dos celdas son adyacentes si ellas tienen un borde común.

Sean:  $c$ : número total de celdas,  $m$ : número total de celdas marcadas, y  $y_i$ : número de celdas adyacentes a la celdas  $i$ ,  $i = 1, 2, \dots, c$ .

Para el método de Grimson es importante calcular la media  $\bar{y}$  y la varianza  $V(y)$  del número de bordes por celdas.

El estadístico  $\mathbf{G}$  de Grimson representa el número observado de celdas marcadas adyacentes. Bajo la hipótesis nula, el valor esperado de  $\mathbf{G}$  y su varianza pueden calcularse mediante fórmulas específicas que aparecen en [9].

$\mathbf{G}$  se estandariza en la forma usual y se utiliza el test clásico de la distribución normal para saber si es significativamente mayor que su valor esperado. En particular si  $c$  es grande y  $m$  pequeño,  $\mathbf{G}$  sigue una distribución de Poisson.

El método de Grimson tiene la particularidad de poder usarse en la detección de aglomeraciones espaciales, temporales y espacio - temporales. A continuación se mostrará una forma de reducir el caso espacio - temporal al espacial.

**Proposición 1** Sean:  $e$ : número de celdas espaciales,  $z_j$ : número de celdas geográficas adyacentes a la celda  $j$ ,  $j = 1, 2, \dots, e$ ;  $t$ : número de celdas temporales,  $y_{j,k}$ : número de celdas tridimensionales adyacentes a la celda  $(j, k)$ ,  $j = 1, 2, \dots, e$ ;  $k = 1, 2, \dots, t$ . Si  $\bar{z}$  y  $V(z)$  son la media y la varianza de las celdas adyacentes geográficas, entonces la media  $\bar{y}$  y la varianza  $V(y)$  de las celdas en tres dimensiones viene dada por las expresiones:

$$\begin{aligned}\bar{y} &= 3\bar{z} + 2 - \frac{2}{t}(\bar{z} + 1) \\ V(y) &= 9V(z) - \frac{10}{t}V(z) + \frac{2(t-2)}{t^2}(\bar{z} + 1)^2\end{aligned}$$

DEMOSTRACIÓN: La demostración es una consecuencia inmediata de que la cantidad de celdas adyacentes a la celda  $(j, k)$  es:

$$y_{i,k} = \begin{cases} 2z_j + 1 & k = 1 \text{ o } k = t \\ 3z_j + 2 & k = 2, \dots, t-1 \end{cases} \quad (3)$$

■

#### 4.1 Vía particular de inclusión de factores de riesgo al método de Grimson

Supongase que se quiere extender el método de Grimson geográfico a un espacio de la forma  $(X, Y, F)$ , donde  $F$  representa un factor de riesgo dicotómico ( $F = 1$ : presencia del riesgo,  $F = 0$ : ausencia).

Se había llamado  $e$  al número de celdas geográficas en las que se ha dividido la zona en análisis, sea  $y_{0,j}$  la cantidad de celdas adyacentes a la celda  $j$ . En el espacio tridimensional se tienen un total de  $c = 2e$  celdas cúbicas,  $e$  de ellas en el nivel  $F = 0$  y otras  $e$  en el nivel  $F = 1$ . Para ubicar un caso en ese nuevo espacio hay primero que situarlo en el plano y luego colocarlo en el nivel  $F = 1$  si el riesgo está presente o en  $F = 0$  si ocurre lo contrario. En este sentido dos celdas se consideran adyacentes si lo son desde el punto de vista geográfico y en ambas el número de pacientes en los que está presente el factor de riesgo excede significativamente el valor esperado de Poisson. La media  $\bar{y}_1$  y la varianza

$V(y_1)$  variarán acorde con las fórmulas:

$$\begin{aligned}\bar{y}_1 &= \frac{\bar{y}_0}{2} \\ V(y_1) &= \frac{V(y_0)}{2} + \left(\frac{\bar{y}_0}{2}\right)^2\end{aligned}$$

donde  $\bar{y}_0$  y  $V(y_0)$  son la media y la varianza del método de Grimson en el espacio.

La siguiente proposición constituye una generalización del método a un espacio de  $f+2$  dimensiones:  $(X, Y, F_1, F_2, \dots, F_f)$ .

**Proposición 2** Sean:  $e$ : número de celdas geográficas,  $y_{0,j}$ : número de celdas geográficas adyacentes a la celda  $j$ ,  $j = 1, 2, \dots, e$ ; La media  $\bar{y}_0$  y la varianza  $V(y_0)$  se calculan en la forma usual. A cada celda geográfica  $j$  se le asocian las  $2^m$  celdas de la forma:  $(j, F_1, F_2, \dots, F_f)$  donde cada  $F_k$  es 0 ó 1. El criterio de adyacencia es el mismo anterior considerando ahora  $f$  factores en vez de uno. La media de las celdas adyacentes  $y_f$  y su varianza  $V(y_f)$  se calculan por:

$$\begin{aligned}\bar{y}_f &= \frac{\bar{y}_0}{2^f} \\ V(y_f) &= \frac{V(y_0)}{2^f} + (2^f - 1)\left(\frac{\bar{y}_0}{2^f}\right)^2\end{aligned}$$

DEMOSTRACIÓN: Se obtiene por inducción sobre  $f$ . ■

Supongamos ahora que el test de Grimson se aplica sucesivamente en un espacio de  $f$  dimensiones acorde a lo planteado en la proposición 2. Sea  $G_f$  el estadístico de Grimson y  $E(G_f)$  y  $V(G_f)$  su esperanza y varianza respectivamente. El valor estandarizado  $Z_f$  se calcula para cada caso en la forma usual.

Si los factores considerados caracterizan la enfermedad, estos deben estar presentes en un alto porcentaje de la población, por tanto la significación del test de Grimson se incrementa en la medida en que  $f \rightarrow \infty$ .

En este proceso iterativo, la media y la varianza de los bordes de las celdas cambian acorde a lo enunciado en la proposición 2. Por otro lado si los riesgo caracterizan al conglomerado, el valor del estadístico de Grimson  $G_f$  no debe variar demasiado para los diferentes valores de  $f$  y la mayoría de las celdas marcadas inicialmente no deben perder esta condición con la introducción de los factores. Si esto sucede, entoces los límites de  $E(G_f)$  y  $V(G_f)$  tenderán a cero.

## 5 Simulación de datos

Para mostrar una aplicación de la teoría anteriormente expuesta se necesita simular datos en espacio - tiempo, o sea valores de la forma:  $(X, Y, T)$ .  $X, Y$  representan las coordenadas espaciales que significan por ejemplo, el lugar de residencia o de trabajo de la persona enferma mientras que  $T$  simboliza la fecha de diagnóstico o de primeros síntomas de la enfermedad. Se le llamará  $pob$  al tamaño de la población enferma, este valor se variará para analizar como ello influye en la capacidad de detección de ambos métodos. Los valores se generan de la siguiente forma:

**Coordenadas espaciales**  $(X, Y)$  : los  $pob/4$  primeros casos, ( / simboliza división entera), se generan siguiendo una distribución uniforme en un cuadrado proporcional al tamaño  $pob$  de la población. Los restantes valores, que son la mayoría, se generan según una distribución uniforme en otro cuadrado contenido dentro del primero, estos últimos puntos quedan mucho más cerca unos de otros. Los valores se simulan de esta forma para asegurar la existencia de al menos un conglomerado.

**Dimensión temporal**  $(T)$  : los  $pob/4$  primeros casos se generan siguiendo una distribución uniforme y el resto según el modelo de una epidemia simple. En este modelo la infección se propaga por contacto entre los miembros de una comunidad; no hay eliminación de circulación de los individuos ya sea porque fallezcan o porque se recuperen o se aislen, al final de la epidemia todos los susceptibles se infectan. El modelo se puede aplicar en aquellas situaciones en las que la enfermedad resulta ser muy infecciosa pero no lo suficientemente seria como para que las personas infectadas mueran o haya que aislarlas, los infectados permanecen inmunes durante el resto de la epidemia. Pudiera ajustarse a los catarros comunes y a otras infecciones del tracto respiratorio por solo mencionar algunos ejemplos, [10].

El modelo es en esencia el siguiente: supóngase que se tiene un grupo homogéneo de tamaño  $n$ . Se asume que la epidemia comienza en el instante  $t = 0$  con  $a$  personas infectadas, los  $n$  individuos del grupo se consideran susceptibles. En un instante de tiempo  $t$  hay  $S$  susceptibles e  $I$  infectados, luego  $S + I = n + a$ . La razón de ocurrencia de una nueva infección  $r$  es proporcional a la cantidad de infectados y de susceptibles, por tanto el número de nuevos infectados en el intervalo de tiempo  $\Delta t$  puede escribirse como  $rSI\Delta t$ . Para simular la epidemia, más que el modelo se necesita conocer la distribución del tiempo que transcurre entre la aparición de dos infectados consecutivos y ella se comporta según una exponencial negativa con función de densidad [10]:

$$f(t) = S(n + a - S)e^{-S(n+a-S)t} \quad (4)$$

**Factores de riesgo** : se generaron según distribuciones de Bernoulli con probabilidades especificadas por el usuario. Los primeros  $pob/4$  casos se generaron siguiendo una distribución de Bernoulli con  $1/3$  de la probabilidad correspondiente, mientras que el resto se simuló con la probabilidad completa. Esto posibilita que hayan más individuos con el riesgo en los  $\frac{3}{4}pob$  últimos casos que son a su vez los que se encuentran más próximos en espacio y en tiempo.

## 6 Aplicación con datos simulados

Para realizar los cálculos explicados anteriormente se programaron diversas funciones en el paquete Mathematica. En el caso del modelo de epidemia simple se utilizaron tres valores para la razón de infección para simular epidemias de rápida propagación ( $r = 0.01$ ) y otras un poco más lentas ( $r = 0.001, 0.0001$ ). En todos los casos se consideró un infectado inicial, ( $a = 1$ ). El tamaño  $pob$  de la población también se varió para analizar su incidencia en la detección de los conglomerados; se consideró una gama de valores que

varió desde poblaciones muy pequeñas ( $pob = 20$ ) hasta poblaciones mayores ( $pob = 100$ ). La simulación de los datos geográficos se realizó en todos los casos siguiendo los principios expuestos anteriormente.

Los parámetros específicos de los métodos no se variaron para que no fueran ellos los que propiciaran la detección o no de las aglomeraciones, así por ejemplo en el caso del método de Knox ambas distancias críticas se tomaron como una unidad.

Para cada combinación de los parámetros  $r$  y  $pob$  se ejecutaron 1000 corridas de los métodos y se contabilizó las veces en las que el resultado fue significativo (probabilidad empírica de significación menor que 0.05). Se considera que el método produce resultados aceptables cuando acierta en la detección de asociaciones significativas de enfermos en el rango comprendido entre el 70% y el 90% de los casos analizados, cuando está entre 90% y 97% los resultados se consideran muy buenos y si supera al 97% se catalogan de excelentes.

A continuación aparecen dos tablas con los resultados obtenidos para 1000 juegos de datos por cada valor diferente de  $r$  y de  $pob$  para ambos métodos.

$pob$	$r = 0.01$	$r = 0.001$	$r = 0.0001$
20	479	107	44
50	1000	925	505
80	1000	1000	876
100	1000	1000	977

Tabla 2: Resultados de la aplicación del método de Knox.

$pob$	$r = 0.01$	$r = 0.001$	$r = 0.0001$
20	838	812	821
50	997	990	995
80	1000	1000	999
100	1000	1000	1000

Tabla 3: Resultados de la aplicación del método de Grimson.

El método de Knox no ofrece buenos resultados cuando el tamaño de la población es pequeño ( $pob = 20$ ), véanse en la tabla 2 valores de 479, 107 y 44 casos de 1000 en los que se ha detectado el conglomerado. Todos esos valores están muy distantes de la categoría de aceptable. Cuando la razón de infección de la epidemia es baja ( $r = 0.0001$ ) y se está en presencia de poblaciones pequeñas ( $pob = 20, 50$ ), Knox tampoco resulta ser un buen método. En todos estos casos es ampliamente recomendable utilizar siempre Grimson, ver tabla 3. Cuando el tamaño de la población aumenta ( $pob = 100$ ) ambos métodos ofrecen resultados excelentes con independencia de la razón de infección de la epidemia simple para los casos aquí considerados, ver tablas 2 y 3.

Para los análisis posteriores sólo se trabajará con aquellos juegos de valores de los parámetros  $r$  y  $pob$  para los cuales los métodos obtuvieron resultados de al menos aceptables.

### 6.1 Vía general de inclusión de factores de riesgo

A continuación se realizó un estudio incorporando factores de riesgo según la vía general. Primeramente se consideró un solo factor de riesgo con una distribución de Bernoulli con probabilidad 0.7, luego se consideraron dos factores con probabilidades 0.7 y 0.6 respectivamente y por último, tres: 0.7, 0.6 y 0.6. Se seleccionaron probabilidades relativamente altas porque se supone que los primeros factores que se tengan en cuenta por los profesionales de la Salud Pública sean aquellos con una mayor frecuencia. Los resultados obtenidos aparecen en las tablas 4 y 5.

$pob$	$r$	{ }	{0.7}	{0.7, 0.6}	{0.7, 0.6, 0.6}
20	0.01	479	-	-	-
	0.001	107	-	-	-
	0.0001	44	-	-	-
50	0.01	1000	939	468	118
	0.001	925	545	284	118
	0.0001	505	-	-	-
80	0.01	1000	995	690	173
	0.001	1000	936	611	334
	0.0001	876	445	170	64
100	0.01	1000	1000	707	165
	0.001	1000	968	794	456
	0.0001	977	621	314	133

Tabla 4: Vía general de inclusión de factores de riesgo al método de Knox.

$pob$	$r$	{ }	{0.7}	{0.7, 0.6}	{0.7, 0.6, 0.6}
20	0.01	838	752	565	356
	0.001	812	710	563	362
	0.0001	821	687	535	319
50	0.01	997	975	967	897
	0.001	990	955	931	833
	0.0001	995	962	911	809
80	0.01	1000	996	991	986
	0.001	1000	979	969	959
	0.0001	999	990	966	953
100	0.01	1000	996	992	988
	0.001	1000	993	988	972
	0.0001	1000	986	973	972

Tabla 5: Vía general de inclusión de factores de riesgo al método de Grimson.

Obsérvese que de manera general en todos los casos se aprecia un decrecimiento de la cantidad de casos en los que se detectan conglomerados, por ejemplo para  $pob = 50$  y  $r =$

0.001 Knox detectó aglomeraciones en 925 de 1000 simulaciones realizadas, al introducir un factor de riesgo con probabilidad 0.7 la cantidad de detecciones se redujo a 545, al incluir el otro con probabilidad 0.6 disminuyó a 284 y con la introducción del tercer factor la cantidad de congregaciones detectadas fue de 118. Por su parte los mismos valores para el método de Grimson fueron de 990 sin incluir ningún factor, 955, 931 y 833 al introducir uno, dos y tres factores respectivamente. Esto se corresponde con lo predicho por la teoría, los demás casos se analizan de la misma manera, ver tablas 4 y 5.

Ambos métodos producen decrecimientos en la cantidad de casos en que se detectan conglomerados cuando se le incorporan los factores de riesgo, pero no en la misma medida. A pesar de que los datos sobre los que se corrieron ambos fueron exactamente los mismos, Knox provocó decrecimientos drásticos, mientras que Grimson resultó ser mucho más conservador. Como se pudo apreciar con anterioridad, los fundamentos matemáticos de estas técnicas son completamente diferentes por lo que no debe esperarse que respondan idénticamente ante la incorporación de factores idénticos.

## 6.2 Vía particular de inclusión de factores de riesgo

Los resultados de la vía particular de inclusión de factores de riesgo se muestran en las Tablas 6 y 7.

$pob$	$r$	{ }	{0.7}	{0.7, 0.6}	{0.7, 0.6, 0.6}
20	0.01	479	-	-	-
	0.001	107	-	-	-
	0.0001	44	-	-	-
50	0.01	1000	985	802	455
	0.001	925	758	366	164
	0.0001	505	-	-	-
80	0.01	1000	1000	953	700
	0.001	1000	981	776	429
	0.0001	876	501	173	67
100	0.01	1000	1000	974	754
	0.001	1000	998	910	555
	0.0001	997	699	317	135

Tabla 6: Vía particular de inclusión de factores de riesgo al método de Knox.

Como que la segunda vía de inclusión de factores de riesgo no es la misma para ambos métodos, las tablas 6 y 7 se analizan de manera diferente. La tabla 6 correspondiente al método de Knox, se interpreta de manera similar a la 4 y la 5 anteriores. Conforme a lo esperado se puede apreciar que hay una disminución de los valores que detectan una aglomeración en la medida en que se introduce un nuevo riesgo. Por ejemplo para  $r = 0.001$  y  $pob = 50$  Knox detectó aglomeraciones de enfermos en 925 de 1000 juegos de datos, al incluir un factor con probabilidad 0.7 este valor disminuyó a 758, al introducir dos se decrementó a 366 y a 164 al considerar los tres. Para los demás casos los resultados son similares.

$pob$	$r$	{ }	{0.7}	{0.7, 0.6}	{0.7, 0.6, 0.6}
20	0.01	838	1000	934*	699*
	0.001	812	997	944*	715*
	0.0001	821	995	938*	675*
50	0.01	997	1000	1000	986*
	0.001	990	1000	1000	991*
	0.0001	995	1000	1000	991*
80	0.01	1000	1000	1000	1000
	0.001	1000	1000	1000	1000
	0.0001	1000	1000	1000	1000
100	0.01	1000	1000	1000	1000
	0.001	1000	1000	1000	1000
	0.0001	1000	1000	1000	1000

Tabla 7: Vía particular de inclusión de factores de riesgo al método de Grimson.

Por su parte la tabla 7, que expone los resultados del método de Grimson, se interpreta de manera inversa: como se explicó con anterioridad, en la medida en que se introducen riesgos, se incrementa la significación de la agrupación, obsérvese por ejemplo que para  $r = 0.001$ ,  $pob = 50$  sin considerar factores, Grimson detecta conglomerados en un 99%, al incorporarle un factor con probabilidad 0.7 el porcentaje de detección llega al 100%, valor este que se mantiene al incorporársele el siguiente riesgo.

En la tabla 7 existen varios datos que no se corresponden con lo planteado en el párrafo anterior, ellos son: poblaciones de tamaño 20 con dos o tres factores y poblaciones de tamaño 50 con tres factores, (marcados con \*). En todos los casos aparecen valores inferiores a 1000 debido a que el estadístico  $\mathbf{G}$  de Grimson se redujo a cero por falta de datos y en entonces la probabilidad empírica de significación fue de 1.0. Obsérvese que esto no ocurrió en poblaciones mayores.

Ninguno de estos métodos tiene la verdad absoluta, ambos tienen el mismo objetivo pero parten de formulaciones matemáticas diferentes, es por ello que se recomienda aplicar los dos y además consultar la opinión del personal médico de más experiencia en el tema.

## 7 Conclusiones

Las investigaciones de apoyo a la epidemiología son muy útiles si se analizan los efectos de prevención o los procedimientos de control que de ellas pueden derivarse, en particular la detección a tiempo de un conglomerado inusitado de enfermos y la información adicional de sus posibles causas, dadas por sus factores de riesgo, puede evitar a tiempo pérdidas humanas. Resulta entonces de vital importancia conocer las limitaciones de las técnicas con las que se cuenta a fin de evitar su aplicación incorrecta.

El método de Knox no puede usarse para detectar focos epidémicos en poblaciones muy pequeñas, ( $pob = 20$ ) pues la probabilidad de error es en todos los casos analizados demasiado grande. Si la infección se propaga con rapidez ( $r = 0.01$ ) Knox puede detectar

satisfactoriamente conglomerados de relativamente pocos individuos, ( $pob = 50$ ). A medida que la razón de infección de una epidemia simple  $r$  disminuye, decrece la capacidad de detección del método con pocos casos. Grimson ofrece muchas más posibilidades, pues para este experimento ofreció resultados favorables con independencia del tamaño de la población y de la razón de infección para los valores considerados.

Por otra parte, la incorporación de factores de riesgo con altas probabilidades siguiendo la vía general, provoca en ambos métodos una pérdida de la significación en un número considerable de juegos de datos, lo que se identifica con lo enunciado en la teoría. Los resultados de la aplicación de la vía particular son también alentadores. Resulta muy difícil para ciertos conglomerados de enfermedades conocer sus verdaderas causas, pero sin dudas un análisis exhaustivo de sus factores de riesgo puede arrojar alguna luz y servir de ayuda al fatigoso trabajo de los profesionales de la Salud Pública.

## Referencias

- [1] Jacquez, G.; Waller, L.; Grimson, R.; Watenberg, D. (1996) “The analysis of Disease clusters, Part I: State of the Art”, *Infection Control and Hospital Epidemiology* **17**(6): 319–327.
- [2] Jacquez, G.; Waller, L.; Grimson, R.; Watenberg, D. (1996) “The analysis of Disease clusters, Part II: Introduction to techniques”, *Infection Control and Hospital Epidemiology* **17**(6): 385–397.
- [3] Casas, G.; Grau, R. (1999) “Un enfoque diferente de las técnicas de clustering para el estudio de epidemias”, *Revista de Matemática: Teoría y Aplicaciones* **6**: 175–187.
- [4] Rose, G. (1985) “Individuos enfermos y poblaciones enfermas”, *International Journal of Epidemiology* **14**: 32–38.
- [5] Knox, E. (1964) “The detection of spece-time interactions”, *Applied Statistics* **13**: 25–30.
- [6] Knox, E. (1964) “Epidemiology of childhood leukemia in Northemberland and Durham”, *Brit. J. Prev. Soc. Med.* **18**: 17–24.
- [7] Pike, M.; Smith, P. (1968) “Disease clustering: a generalization of Knox’s approach to the detection of space-time interactions”, *Biometrics* **24**: 541–554.
- [8] Abe, O. (1973) “A note on the methodology of Knox’s test of time and space interactions”, *Biometrics* **29**: 66–77.
- [9] Grimson, R.; Rose, R. (1991) “A versatile test for clustering and a Proximity Analysis of Neurons”, *Meth. Inform. Med.* **30**: 299–303.
- [10] Bailey, N. (1975) *The Mathematical Theory of Infectious Diseases and it’s Applications*. Charles Griffin Co. Ltd.