

Corredores o canales endémicos y su elaboración usando planillas de cálculo

Marcelo Bortman

Probablemente la más corta y simple definición de epidemia la ha dado Benenson quien la define como “La aparición de casos de una enfermedad en evidente exceso de lo esperado”^[1]. Last, en “A Dictionary of Epidemiology”^[2] agrega a la definición el concepto de espacio, “en una comunidad”, y la extiende a otros eventos relacionados con la salud, manteniendo la idea de “un evidente exceso”.

En ambas definiciones está implícito una comparación entre el número de casos detectados y una cifra “normal de casos”. Esa cifra “normal” dependerá obviamente de la población, del área geográfica y del período de tiempo que se considere. Este último punto es especialmente importante para aquellas enfermedades agudas en las que la estacionalidad juega un rol importante.

La palabra epidemia tiene un “aura pestilente” que hace que muchos prefieran usar en su lugar el término “brote” para referirse a los excesos de casos de una enfermedad o daño.

Si bien originalmente la definición de epidemia fue pensada en las enfermedades transmisibles, su uso es totalmente compatible con las no transmisibles y es así como por ejemplo durante los fines de semana largos (con viernes o lunes feriados) se producen verdaderos “brotes” o epidemias de accidentes de tránsito.

Johan Giesecke destaca que no sólo el número de casos, (tiempo y lugar), debe ser considerado, y que una distribución poblacional particular, (tiempo, lugar y persona) puede estar evidenciando un número de casos mayor a lo esperado para ese grupo poblacional específico, (mujeres, trabajadores de una particular industria, con un característico estilo de vida, etc)^[3]

Teniendo en cuenta lo hasta aquí expuesto, a la hora de hacer vigilancia epidemiológica en una población, será fundamental conocer "cuanto es endémico". Esto es, cual es la incidencia histórica de la enfermedad para dicha población o grupo y para ese "momento" del año.

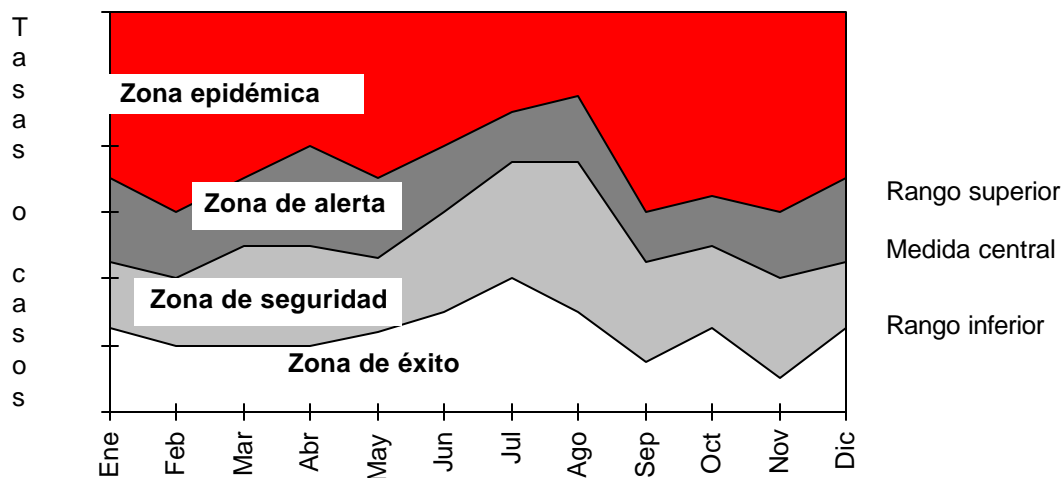
Es aquí entonces donde se hace necesario la preparación de corredores o canales endémicos en los que la incidencia actual pueda ser graficada sobre la incidencia histórica y de esta forma detectar precozmente cifras anormalmente altas (o bajas) de casos de la enfermedad en estudio.

Ya en 1970 los Dres Héctor Boffi Borggero y Carlos Alvarez Herrera describen varios métodos para la realización de corredores endémicos ^[4]. En todos ellos a partir de una serie de casos de un período de 5 a 7 años, se trata de obtener una medida central y un rango de fluctuación normal de la incidencia para cada uno de los meses.

El más sencillo de estos métodos consistía en graficar el número máximo y mínimo de casos denunciados en cada mes, generando así una "banda endémica" con un área inferior de seguridad y una superior o epidémica. Un poco más complejo es el método de la mediana y los cuartiles, que genera cuatro zonas: una debajo del cuartil inferior o zona de éxito, una entre el cuartil inferior y la mediana o zona de seguridad, una entre la mediana y el cuartil superior o zona de alerta y una por arriba del cuartil superior o zona epidémica. Los más complicados modelos eran el de los mínimos cuadrados, (que analizaba las tendencias lineales de cada año), y el de la media aritmética y desvíos estándar que requerían contar con personal con considerables conocimientos en estadística para su realización. Estos últimos modelos también cuentan con las cuatro zonas ya descriptas y que pueden verse el gráfico nº 1.

GRAFICO N° 1.

ESQUEMA CON LAS CUATRO ZONAS DE LOS CORREDORES ENDEMICOS



Todos estos métodos pueden ser considerados modelos matemáticos más o menos complejos en los que se pretende, a partir de 5 a 7 cifras, pronosticar cual en el número de casos que cabría esperar para ese período del año con un rango superior y uno inferior. Obviamente a medida que se quiera obtener un mejor "modelo", se requerirá de un mayor número de cálculos. Con el uso cada vez más corriente de computadoras y de programas que facilitan tanto la realización de cálculos repetitivos como la graficación de los resultados de estos cálculos, estos modelos complejos pueden ser realizados sin la necesidad de contar con personal estadístico especializado. Esto facilitará la realización de corredores endémicos en niveles locales y regionales para la vigilancia de la incidencia de diferentes eventos de salud en estos niveles. De esta forma se incrementará la sensibilidad de los sistemas de vigilancia para detectar cambios del número esperado de casos, que desde la perspectiva de los niveles centrales podrían verse diluidos.

Factores a tener en cuenta antes de hacer un corredor endémico

La selección de la entidad, las poblaciones, la serie de años a incluir y los intervalos de tiempo condicionarán el grado de precisión de los corredores endémicos. En entidades con incidencia baja, en poblaciones pequeñas, o usando intervalos de tiempo corto, el rol del azar se hace más importante. La consiguiente inestabilidad o dispersión de los casos denunciados en los años previos condiciona notablemente la posibilidad de predicción, resultando en corredores de líneas dentadas con anchas áreas de seguridad y alarma. Se podría entonces decir que a mayor nivel de desagregación de la información menor será la precisión de la predicción. Una posibilidad alternativa para patologías endémicas pero de baja incidencia es la realización de corredores acumulativos. Se trata de un corredor endémico en el que en lugar de usar la incidencia semanal, el gráfico se construye usando la incidencia acumulada. Luego de la descripción de los pasos necesarios para realizar un corredor endémico con planillas de cálculos se describirá como realizar un paso intermedio adicional que permitirá graficar corredores acumulativos.

Criterio de selección de entidades o patologías:

Al seleccionar una entidad o patología para realizar un corredor endémico se deberá tener presente que se trate de una entidad endémica, que tenga un período de incubación breve y una evolución aguda. No tendrá sentido realizar corredores endémicos para patologías de muy baja frecuencia en las que la sola presencia de uno o pocos casos debe alertar a los sistemas de vigilancia. De igual modo, en las patologías de evolución crónica, la acumulación de nuevos casos en un breve período de tiempo en general no tendrá mayor significación para la vigilancia epidemiológica y seguramente serán la consecuencia de cambios o intensificación de los métodos diagnósticos (tamizajes) o simplemente del azar. Solo en situaciones muy excepcionales serán la consecuencia de una verdadera epidemia como lo fueron los sarcomas de Kaposi durante la irrupción del SIDA.

Agrupamiento de entidades o patologías:

Cuando la incidencia lo permita será útil realizar corredores independientes para cada patología y sus subgrupos, por ejemplo diarreas por grupos etéreos, o por agentes etiológicos. En el otro extremo, en poblaciones pequeñas, se podría mantener agrupadas las entidades con una similar forma de transmisión vigilando así ya no el comportamiento de un germen o una entidad, sino el de una vía de transmisión.

Poblaciones:

De forma similar que en el caso anterior mantener la información y los corredores endémicos desagregados por pequeñas áreas geográficas favorecerá la posibilidad de detectar pequeños brotes locales que resultarían diluidos en cifras acumuladas de regiones más amplias. Pero este nivel de desagregación deberá equilibrarse con el incremento de la imprecisión de los resultados. Cabe destacar aquí que en ciertas circunstancias es justamente la vigilancia sobre poblaciones más grandes y su consiguiente aumento de precisión lo que permite la detección de un brote. Así, podrá ser que en varias áreas se genere un número de casos mayor que la media pero sin ingresar en la zona de alarma, mientras que sobre el corredor endémico de la región este "ligero" aumento en los niveles locales resulte en un "notable" incremento en la región.

Otro punto a tener en cuenta, son los cambios que sufren las poblaciones en el número de habitantes a través de los años. Por esta razón no será adecuado trabajar los datos como casos. Siempre que se pueda, se deberá contar con cifras de población que posibiliten la realización de los corredores usando tasas de incidencia. De igual manera, si se consideran patologías según ciertos grupos etéreos, se deberá trabajar con tasas de incidencia específicas para esos grupos, debiéndose contar entonces con la información de habitantes por grupos etéreos y para cada año de las series incluidas.

Longitud de las series e intervalos de tiempo:

Usualmente los corredores se construyen con series de casos de 5 a 7 años. Se puede presuponer que al incrementarse el número de años, se mejorará el modelo de predicción, pero deberá tenerse en cuenta que si se toman series muy largas es probable que tanto las condiciones que mantienen la endemia como los criterios diagnóstico y los mecanismo de denuncia y registro hayan cambiado. Esto no es igual para todas las entidades, pero en líneas generales si se toma una serie de 15 años, difícilmente las condiciones de hace 15 años puedan compararse con las actuales. Este tipo de cambios deberán ser también tenidos en cuenta aún cuando se tomen pocos años. Por ejemplo, si recientemente se han realizado substanciales mejoras en la red cloacal de una localidad, la incidencia de enfermedades de transmisión fecal-oral debería ser menor a la "pronosticada" por el corredor endémico.

En cuanto a los intervalos de tiempo, sería siempre deseable realizar corredores por semanas epidemiológicas. Pero ante bajas incidencias, se deberá utilizar períodos mayores, (2 o 4 semanas) los que tenderá a estabilizar las fluctuaciones debidas al azar. La desventaja de usar períodos de 2 a 4 semanas es que disminuyen la posibilidad de detección precoz de los brotes y de las consiguientes acciones de control.

Selección del programa de computación para hacer el corredor

Si bien existe la posibilidad de desarrollar programas específicos para la realización de corredores endémicos, éstos trabajarían a manera de "caja negra". Los datos serían ingresados en un extremo de la "caja", y los corredores "saldrían" por el otro extremo distanciando a quien hace vigilancia epidemiológica de todo el proceso de su elaboración. Por esta razón y teniendo en cuenta el extenso uso de las planillas de cálculo y las facilidades que éstos programas tienen para realizar cálculos y gráficos, este tipo de utilitarios resultan los más aptos para esta tarea.

Las planillas de cálculo a las que aquí se hará referencia son: QPro versión 4.0 o 5.0 y Microsoft Excel versión 5.0 y 97 y obviamente será conveniente tener algún grado de experiencia en el uso de alguno de estos programas.

Método de las medias geométricas de las tasas

El método de la media geométrica de las tasas y sus intervalos de confianza es desde el punto de estadístico uno de los más sólidos. El detalle de este método podrá resultar un poco técnico, pero será útil su lectura.

Como ya se ha mencionado, el uso de tasas en lugar de casos, ajusta las distorsiones que generan los cambios en los tamaños de las poblaciones a través de los años.

Media geométrica

Teniendo en cuenta que las 5 a 7 tasas con las que hay que calcular un valor central, difícilmente tengan una distribución normal, el uso de la media aritmética no sería lo más adecuado. En cambio la media geométrica es una medida de tendencia central apropiada para distribuciones de valores asimétricas o no normales y esta especialmente indicada para distribuciones con aislados valores altos o muy altos. Esto último hace que eventuales cifras muy altas producto de alguna epidemia ocurrida durante la serie de tiempo incluida en la construcción del corredor se diluyan y no distorsionen la incidencia histórica.

Para el cálculo de la media geométrica se procede a realizar una transformación logarítmica de los valores.

$\mu = \log x$ Donde x es el valor inicial, y μ el valor transformado.

Esta transformación “estira” los valores bajos y “comprime” los valores altos. Luego de esto, todos los procedimientos, (cálculo de medias y desvíos estándar e intervalos de confianza), se realizan con éstos valores transformados. Finalmente los datos obtenidos se pasan a las unidades originales calculando el antilogaritmo, o lo que es lo mismo, calculando 10 elevado a μ , ($x=10^\mu$). Otros tipos de logaritmos, por ejemplos los de base n (naturales), pueden ser usados con exactamente los mismos resultados:

$$\mu = \ln(x) \quad \text{y} \quad x = \exp(\mu).$$

La tabla N° 1 muestra las tasas de sarampión de los años 1983 a 1989 de la provincia del Neuquén, Argentina. En esta tabla se han calculado la media aritmética y la media geométrica de las tasas y sus respectivos desvíos estándar. Como puede verse, el “brote” de 1984 (tasa de 37,92/100000) y la “epidemia” de 1989 (136,43) afecta mucho más a la media aritmética (27,64) y su desvío estándar (49,7), que a la media geométrica (7,1) y su respectivo DE (5,88).

Tabla N° 1. Tasas de sarampión por 10000 habitantes, media aritmética, media geométrica y desvíos estándar. Provincia del Neuquén, Argentina. período 1983 a 1989.

Año	Tasa /100000 hab.	Escala logarítmica Tasa /100000 hab.	Transformados a unidades originales
1983	2,13	0,76	
1984	37,92	3,64	
1985	4,84	1,58	
1986	1,54	0,43	
1987	1,17	0,16	
1988	9,46	2,25	
1989	136,43	4,92	
Media	27,64	1,96	7,10
DE	49,70	1,77	5,88

Una limitación del cálculo de las medias geométricas es que los valores deben ser siempre superior a 0, ya que es imposible calcular el logaritmo de 0 o de números negativos. Esto obviamente constituye una dificultad ya que es probable tener para algunos períodos tasas

igual a 0. Para sortear esta dificultad Betty Kirkwood ^[5] recomienda sumarle 1 a todos los valores originales, realizar todos los cálculos y finalmente restar 1 a los valores finales.

Rangos superior e inferior

Para obtener los rangos superior e inferior, se puede usar el desvío estándar de las tasas, o el intervalo de confianza de la media de las tasas.

El desvío estándar tiene en cuenta la dispersión de los datos que dieron origen a la media (los de la serie). Así, al considerar la media ± 1 DE se estarán incluyendo al 68.26% de los valores que dieron origen a la media, y si se consideran 2 DE se estarán incluyendo al 95.45% de estos valores. Por lo tanto al usar los desvíos estándar en la construcción de los rangos superior e inferior estaremos comparando las tasas del corriente año con las tasas de los años anteriores y no con su media. Una tasa que supere el rango superior e ingrese en la zona epidémica será entonces similar o mayor a los valores más altos de la serie de los últimos años.

El concepto de intervalo de confianza (IC) de la media es distinto, ya que determina el rango en el cual se encontraría la "verdadera media del universo", bajo el supuesto que las 5 o 7 tasas usadas para el cálculo constituyen una muestra del mismo. Un IC del 95% será por lo tanto un rango en el cual existirá un 95% de probabilidades de que esté incluida la media del universo. Al usar el intervalo de confianza en la construcción de los corredores, se estará comparando las tasas del año en curso con la "verdadera media" de las tasas del período anterior. A su vez, la tasa que supere el límite superior del corredor e ingrese a la zona epidémica tendrá un 95% de probabilidades de ser realmente superior a la "verdadera media histórica". El uso de intervalos de confianza dará lugar a corredores con zonas de seguridad y alerta más estrechas y por lo tanto más sensibles a los cambios de incidencias que los corredores formulados con base en desvíos estándar.

Si bien las planillas de cálculo tienen incluida las fórmulas del desvío estándar como funciones, no ocurre lo mismo para el cálculo del intervalo de confianza de la media.

Por esta razón será necesario escribir éstas fórmulas en las celdas correspondientes.

$$\text{IC 95\%} = \text{media} \pm t \cdot \text{de} / \sqrt{n}$$

Donde **n** es el número de años considerados en el cálculo, **de** el desvío estándar y **t** es un valor que sale de la tabla de distribución de t y que varía en función de **n**.

Valores de t para IC 95% según años incluidos en las series

Número de años	Valor de t para IC 95%	Número de años	Valor de t para IC 95%
3	4.30	8	2.36
4	3.18	9	2.31
5	2.78	10	2.26
6	2.57	11	2.23
7	2.45	12	2.20

Por ejemplo, para calcular el intervalo de confianza del 95% de la media geométrica de las tasas de incidencia de sarampión del período 1983 - 1989, se deberá realizar el siguiente cálculo:

$$\text{IC 95\%} = 1,96 \pm 2,45 \times 1,77 / \sqrt{7} = 1.96 \pm 1.639 = 0,32 \text{ a } 3.6$$

Donde 1.96 y 1.77 son la media y el DE en escala logarítmica, 7 el **n** (7 años) y 2.45 es el valor de **t** para un n de 7. Al transformar los resultados a la escala original:

Media geométrica de la tasa de incidencia = 7,10 por 100000 IC 95% de 1,38 a 36,6

Nótese que los intervalos de confianza superior e inferior no son simétricos, sin embargo la razón entre el límite superior y la media y entre la media y el límite inferior si lo son:

$$(36,6 / 7,10 = 7,10 / 1,38)$$

Elaboración en etapas

Para la elaboración del corredor endémico en una planilla de cálculo, será necesario realizar una serie de tablas y cálculos en varias etapas.

1^{ra} ETAPA. Ingreso de los datos.

El primer paso será establecer un bloque en donde se ingresarán los casos detectados por años y períodos. Por ejemplo el bloque 1 [A1 a BB10] con 52 semanas en las columnas y los años en las filas ha sido preparado para ingresar los casos de 1990 a 1998. Nótese que se ha preparado la última columna (BB) para la colocar en ella las poblaciones en cada año.

Bloque 1: Casos por semana y años, y población por años

	A	B	C	D	E			AX	AY	AZ	BA	BB
1	Año \ Sem	1	2	3	4			49	50	51	52	Población
2	1990											
3	1991											
4	1992											
5	1993											
6	1994											
7	1995											
8	1996											
9	1997											
10												
11												

2^{da} ETAPA. Cálculo de las tasas.

El segundo paso será calcular las tasas de incidencia semanal y sumarle 1 a cada una de las tasas para permitir calcular la media geométrica en caso de existir tasas igual a 0. Para esto se preparará un bloque similar debajo del anterior (Bloque 2 [A15 a BA24]) y en cada celda del bloque se ingresará la fórmula correspondiente. La forma más rápida de hacer esto será ingresar la fórmula en la casilla B16 y copiar esta fórmula en las restantes casillas. La fórmula en la casilla B16 será: $((+B2 / \$BB2 * 100000) + 1)$. Donde B2 es el

4^{ta} ETAPA. Cálculo de las medias, desvíos estándar e intervalos de confianza del 95% en escala logarítmica.

Otras cuatros filas serán necesarias para calcular las medias, los desvíos estándar y los IC 95%, (Bloque 4). Dado que hay algunas diferencias en los nombres de las funciones entre QPro y Excel, en las fórmulas usadas en una u otra serán ligeramente distintas. En los ejemplos se usarán los datos de los años 1990 a 1996, (filas 28 a 34), por lo que en los cálculos de los intervalos de confianza, n será igual a 7 y $t = 2.45$

Bloque 4: Cálculo de medias, DE e IC 95% en QPro y Microsoft Excel.

QPro						
	A	B	C	D	E	
37	Ind. \ Sem	1	2	3	4	
38	Ln Media	@Media(b28..b34)				
39	Ln DE	@Stdmod(b28..b34)				
40	Ln IC Sup	+b38 + (2.45 * b39 / @rcuad(7))				
41	Ln IC Inf	+b38 - (2.45 * b39 / @rcuad(7))				
42						

Microsoft Excel						
	A	B	C	D	E	
37	Ind. \ Sem	1	2	3	4	
38	Ln Media	=Promedio(b28:b34)				
39	Ln DE	=Desvest(b28:b34)				
40	Ln IC Inf	=b38 - (2.45 * b39 / raiz(7))				
41	Ln IC Sup	=b38 + (2.45 * b39 / raiz(7))				
42						

5^{ta} ETAPA. Transformación a unidades originales restando 1, y conversión de tasas a casos esperados.

El último paso será volver a convertir los valores obtenidos a su escala original, restarle 1 para restablecer el corrimiento de las tasas introducidos al sumarle 1 en la 2^{da} etapa y convertir las tasas a casos esperados para facilitar la utilización del corredor. El bloque 5 muestra las fórmulas necesarias para realizar esto. La fila 45 contendrá la media geométrica y las filas 46 y 47 los intervalos de confianza del 95% inferior y superior

respectivamente. En las filas 48 a 50 las tasas son convertidas a casos esperados teniendo en cuenta la población del año 1997, (celda BB9).

Bloque 5: Transformación a unidades originales, menos 1 y cálculo de casos.

	A	B	C	D	E	
44	Ind. \ Sem	1	2	3	4	
45	IC Inf (tasa)	@Exp(b40) - 1				
46	Media (tasa)	@Exp(b38) - 1				
47	IC Sup (tasa)	@Exp(b41) - 1				
48	IC Inf (casos)	+b45 * \$bb\$9 / 100000				
49	Media(casos)	+b46 * \$bb\$9 / 100000				
50	IC Sup (casos)	+b47 * \$bb\$9 / 100000				
51						

6^{ta} ETAPA. Gráfico.

Finalmente con la media geométrica y su intervalo de confianza expresados en casos para la población del corriente año ya se estará en condiciones de realizar un gráfico de áreas. Sobre este gráfico semana a semana se deberá graficar a mano los casos denunciados como parte de la actividad de vigilancia.

Para facilitar la "superposición" de las áreas del gráfico, será conveniente otro bloque para calcular la diferencia entre el intervalo de confianza inferior y la media geométrica y entre esta última y el intervalo de confianza superior.

Bloque 6: Cálculo de las diferencias.

	A	B	C	D	E	
52	IC Inf (casos)	+b48				
53	Media - IC Inf	+b49 - b48				
54	IC Sup -Media	+b50 - b49				
55						

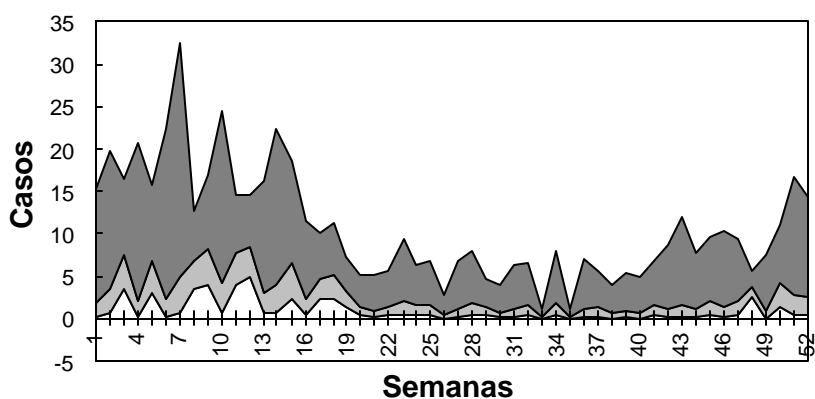
Gráfico con QPro

Desde el menú-gráficos, se creará un gráfico de áreas, se definirán como series 1, 2 y 3 las celdas A52..BA52, A53..BA53 y A54..BA54 respectivamente y pulsando <F10> podrá ser visualizado. Posteriormente se colocarán los títulos adecuados y ya estará listo para la impresión. Se podrán también realizar cambios en los colores, tramas y otros detalles de la presentación gráfica.

Gráfico con Microsoft Excel

Se marca el bloque (A52..BA54) y desde Insertar-Gráfico-en esta hoja, se ingresa al asistente para gráficos. Se selecciona un gráfico de áreas tipo 1 y se continua con el asistente hasta terminar. Como al trabajar con QPro, se podrá luego agregar títulos modificar los colores o tramas de las series, etc. antes de imprimir (Gráfico n° 2).

GRAFICO N° 2: Corredor endémico de diarreas en mayores de 5 años.
Rincón de los Sauces - años 1990 a 1996



CORREDOR ENDEMICO ACUMULATIVO

Tal como se comentó anteriormente, una alternativa para realizar corredores endémicos para entidades endémicas pero de baja incidencia, es utilizar la incidencia acumulada en lugar de la incidencia semanal. De esta forma, se evitará que el corredor presente una imagen muy dentada. Los gráficos 3 y 4 muestran respectivamente el corredor endémico y el corredor endémico acumulativo para meningitis bacteriana en la provincia del Neuquén del período 1990 a 1996. En ambos corredores se graficó también la incidencia del año 1997 hasta la semana 31 como una línea negra. Como puede verse, la vigilancia de la incidencia actual se hace más fácil en el gráfico 4 de incidencia acumulada.

GRAFICO N° 3: Corredor endémico de Meningitis Bacterianas.
Provincia del Neuquén - años 1990 a 1996 e incidencia 1997 (sem 31)

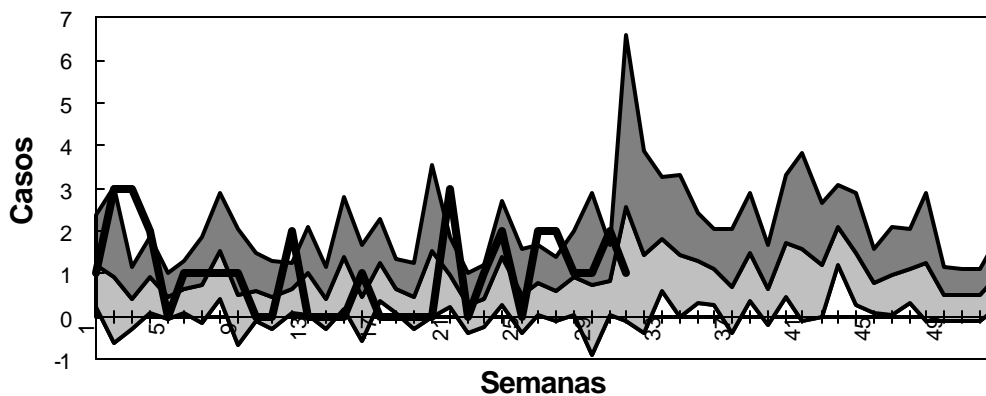
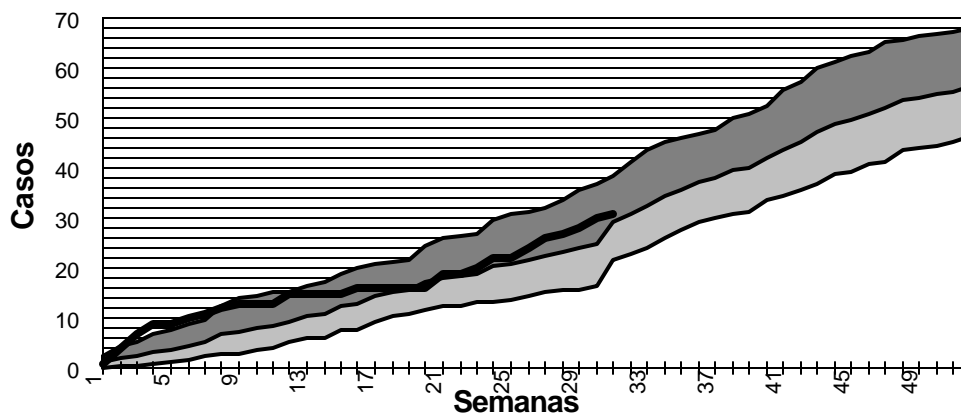


GRAFICO N° 4: Corredor endémico acumulado de Meningitis Bacterianas.
Provincia del Neuquén - años 1990 a 1996 e incidencia 1997 (sem 31)



Para realizar corredores con incidencia acumulada, bastará con introducir la incidencia acumulada semanal en lugar de la incidencia semanal, o incluir un bloque que acumule la incidencia a partir de los datos de incidencia semanal.

REFERENCIAS

1. Benenson AS (de). Control of Communicable Diseases in Man, 15th edition. Washington D C: American Public Health Association, 1990.
2. Last J M. A Dictionary of Epidemiology, 2nd Edition. Oxford; Oxford University Press, 1995.
3. Giesecke, J. Modern Infectious Disease Epidemiology. Oxford University Press, Inc., New York, 1994.
4. Hector Boffi Boggero, Carlos Alvarez Herrera. Contribución al estudio de las enfermedades transmisibles mediante el uso de algunos métodos estadísticos. *Salud Argentina N°1, Ministerio de Bienestar Social, Secretaría de Estado de Salud Pública, 1970.*
5. Kirkwood B. Essentials of Medical Statistics. Blackwell Science 1988.

Para obtener una copia de los archivos con las planillas para realizar los corredores endémicos pueden ser solicitados por correo electrónico a Bortman@altavista.net