

## Medidas utilizadas en la investigación epidemiológica (Segunda Parte)

Juan José García García\*

### Método actuarial

Cuando las duraciones ( $Dt_i$ ) de los periodos de seguimiento individuales para los no casos varían sustancialmente en poblaciones fijas o dinámicas, se utiliza el método actuarial o de tabla de vida para calcular la incidencia acumulada.

El método resulta más apropiado cuando el inicio del periodo de seguimiento de los individuos está caracterizado por un evento discreto, tal como el primer diagnóstico de una enfermedad, el comienzo de un tratamiento, el primer uso de una medida preventiva, o la primera exposición a un factor probablemente de riesgo.

En el método actuarial, la incidencia acumulada para un periodo dado, y, por tanto, el riesgo, se estima por:

$$R_{(t_0, t)} = IA = \frac{1}{N_0 - (p/2)}$$

donde  $p$  es el número de individuos perdidos de la población en estudio durante el periodo  $(t_0, t)$  de duración  $Dt$ .

El denominador puede ser considerado como el número efectivo de personas en riesgo de desarrollar la enfermedad, suponiendo que las pérdidas ocurren a la mitad del periodo de seguimiento. Representa el número de personas sin la enfermedad que produciría los nuevos casos ( $I$ ) si todas las personas pudieran ser estudiadas durante todo el periodo. Desde este punto de vista, el método acumulativo simple es un caso especial del método actuarial.

Si la población en estudio es dinámica, el método actuarial convierte el diseño en una cohorte fija para la cual el tiempo de inicio común para cada sujeto corresponde al momento en que ingresa al estudio.

Para calcular la incidencia acumulada para cualquier periodo de 5.5 años o menos, puede imaginarse que la lí-

nea que representa a cada sujeto es desplazada al margen izquierdo (tiempo =  $t_0$ ) y que el total del periodo de observación es dividido en 5 intervalos consecutivos de 1 año  $(t_{j-1}, t_j)$ , donde  $j = 1, \dots, 5$ . Así, por ejemplo, el sujeto 6 desarrolla la enfermedad en el primer año ( $j = 1$ ), y el sujeto 12 es perdido de vista en el segundo año ( $j = 2$ ). Como resultado de la ocurrencia de la enfermedad y de la pérdida de los individuos, el tamaño ( $N'_{oj}$ ) de la cohorte libre de la enfermedad al inicio del intervalo  $j$ -ésimo disminuye de 12 ( $j = 1$ ), a 10 ( $j = 2$ ), a 7 ( $j = 3$ ), a 2 ( $j = 4$ ), a 1 ( $j = 5$ ). Utilizando la fórmula anteriormente expresada, puede calcularse la incidencia acumulada en un año para el intervalo  $j$ -ésimo.

Por ejemplo, el riesgo estimado  $R_1 = IA_1$  de desarrollar la enfermedad para el primer año después de la exposición es:

$$R_1 = IA_1 = \frac{1}{12 - (1/2)} = \frac{1}{12 - 0.5} = \frac{1}{11.5} = 0.087$$

Los resultados para los cinco intervalos se presentan en el siguiente cuadro:

Estimación del riesgo ( $R$ ) de desarrollar la enfermedad  $x$  para una cohorte hipotética de 12 sujetos, por intervalos de un año  $(t_{j-1}, t_j)$ , usando el método actuarial.

J	$(t_{j-1}, t_j)$	$N'_{oj}$	$I_j$	$P_j$	$R_j$	$R_{(t_0, t_j)}$
1	(0, 1)	12	1	1	0.087	0.087
2	(1, 2)	10	1	2	0.111	0.188
3	(2, 3)	7	2	3	0.364	0.484
4	(3, 4)	2	1	0	0.500	0.742
5	(4, 5)	1	0	1	0.000	0.742
Total	(0, 5)	-	5	7	-	-

Nota:  $R_j$  es el riesgo estimado a un año para el intervalo  $j$ -ésimo  $(t_{j-1}, t_j)$ .  $R_{(t_0, t_j)}$  es el riesgo estimado Daños para el periodo acumulado  $(t_0, t_j)$ , donde  $D = t_j - t_0$ .

\* Coordinador de enseñanza de 3o. y 4o. año. Departamento de Salud Pública. Facultad de Medicina, UNAM.

Para estimar el riesgo para un periodo acumulado ( $t_0, t$ ) de Daños se combina la estimación ( $D \in Dt$ ) del riesgo a 1 año ( $R_j = IA_j$ ) usando la fórmula:

$$R_{(t_0, t_j)} = IA_{(t_0, t_j)} = 1 - \prod_{j'=1}^j (1 - IA_{j'})$$

donde  $IA_j$  es la incidencia acumulada  $Dj$  año para el intervalo de edad (o tiempo)  $j$ -ésimo ( $t_{j-1}, t_j$ ), de acuerdo a:

$$D = \sum_{j'=1}^j Dj' = t_j - t_0, \quad Dj = t_j - t_{j-1}$$

y  $j'$  es un subíndice para el  $j$ -ésimo estrato, de tal manera que puede acumularse el tiempo desde el inicio del primer intervalo de edad, hasta el final del intervalo  $j$ -ésimo.

Es importante destacar que  $Dt$  se refiere a tiempo calendario (número de años en que se sigue a una población), mientras que  $D$  se refiere a tiempo de edad, es decir, un número designado de años en la vida de un individuo, por ejemplo, de 50 a 59 años.

Observe la mecánica de cálculo para estimar el riesgo acumulado para intervalos sucesivos, a partir de los datos de la tabla anterior.

Riesgo a:

2 años	$1 - (1 - 0.087) (1 - 0.111) = 0.188$
3	$1 - (1 - 0.188) (1 - 0.364) = 0.484$
4	$1 - (1 - 0.484) (1 - 0.500) = 0.742$
5	$1 - (1 - 0.742) (1 - 0.000) = 0.742$

Implícitamente al aplicar estos cálculos se está suponiendo que las pérdidas ocurren, en promedio, en el punto medio de los intervalos. Si estos son suficientemente cortos ( $Dj \leq 1$  año), o hay relativamente pocas pérdidas, la violación a este supuesto no afecta seriamente el resultado; sin embargo, aun si todas las pérdidas ocurren en dicho punto medio, el método actuarial conduce a una estimación sesgada del riesgo. Debido a este problema se han propuesto varias otras alternativas, pero todas producen aproximadamente los mismos resultados.

### Método de la densidad de incidencia

Este método para la estimación del riesgo está basado en la estimación de tasas de incidencia, y depende de la relación funcional entre ambos tipos de medidas. Para apreciar esta relación, considérese el caso de una cohorte

te fija de  $N'_0$  sujetos sin la enfermedad (en el tiempo cero), que es seguida por  $t$  años para la detección subsecuente de una enfermedad (primera incidencia).

La letra  $N'_t$  será el número de sujetos sanos que continúan en el estudio en el tiempo  $t$  durante el periodo de seguimiento.

Los cambios de  $N'_t$  en el transcurso del tiempo, suponen que la tasa de incidencia (instantánea) permanece constante durante todo el periodo ( $t_0, t_2$ ). Matemáticamente, dada una tasa constante, el número de sujetos sanos  $N'_t$  en el tiempo  $t$  es una función exponencial con una pendiente negativa. Esto se expresa:

$$N'_t = N'_0 \text{ antilog } [-TI (D)]$$

donde  $D = t - t_0$  es el tiempo transcurrido (o edad) entre el inicio del seguimiento y el tiempo  $t$ , y  $D \leq Dt$  es la duración del periodo de seguimiento observado. Puede estimarse entonces el riesgo en años,  $R_{(t_0, t)}$ , como una función de la tasa promedio estimada (TI), utilizando la siguiente expresión:

$$R_{(t_0, t)} = 1 - (N'_t / N'_0) = 1 - \text{antilog } [-TI (D)]$$

A partir de los datos de la tabla anteriormente analizada, la estimación del riesgo ( $R$ ) de desarrollar la enfermedad  $x$  para una cohorte hipotética de 12 sujetos, por intervalos de un año, utilizando el método de la densidad, se observa a continuación.

$j$	$(t_{j-1}, t_j)$	$t - p_j$	$I_j$	$DI_j$	$R_j$	$R_{(t_0, t_j)}$
1	(0, 1)	11.0	1	0.091	0.087	0.087
2	(1, 2)	8.5	1	0.118	0.111	0.189
3	(2, 3)	4.5	2	0.444	0.359	0.480
4	(3, 4)	1.5	1	0.667	0.487	0.733
5	(4, 5)	0.5	0	0.000	0.000	0.733
Total		26.0	5	0.192	-	-

donde el riesgo puede obtenerse por

$$R_{(t_0, t_j)} = 1 - \text{antilog } \left[ -\sum_{j'=1}^j DI_{j'} (Dj) \right]$$

de tal manera que al sustituirse los valores, se tiene:

Riesgo a:

1. año  $1 - \text{antilog } [-0.091 (1)] = 0.087$
2.  $1 - \text{antilog } [-0.091 (1) - 0.118 (1)] = 0.189$
3.  $1 - \text{antilog } [-0.091 (1) - 0.118 (1) -$

- |    |  |         |
|----|--|---------|
|    | $0.444 (1) =$  | $0.480$ |
| 4. | $1\text{-antilog} [-0.091 (1) - 0.118 (1) -$<br>$0.444 (1) - 0.667 (1)] =$           | $0.733$ |
| 5. | $1\text{-antilog} [-0.091 (1) - 0.118 (1) -$<br>$0.444 (1) - 0.667 (1) - 0.0 (1)] =$ | $0.733$ |

El uso de esta vía, supone que la tasa es constante dentro de cada intervalo ( $t_{j-1}$ ,  $t_j$ ) de duración  $D_j$ . Al dividir  $D$  en suficientes intervalos, razonablemente puede estar seguro que esta condición no se viola para afectar la estimación del riesgo.

Como puede observarse, la cuantificación de la incidencia de una enfermedad puede realizarse de diversas maneras: el método depende de las características particulares del diseño del estudio que ha sido planteado. Esencialmente hay dos decisiones a tomar para estimar la incidencia en un estudio de seguimiento, elegir la medida empírica apropiada (incidencia acumulada vs densidad de incidencia) y, elegir la medida teórica deseada (riesgo vs tasa).

Teóricamente las densidades de incidencia pueden ser utilizadas para estimar tasas o riesgos; las incidencias acumuladas están limitadas para estudiar riesgos. ¿Cuál elegir en la práctica? El criterio principal a considerar es el objetivo del estudio: si el propósito de la investigación es predecir la ocurrencia individual de la enfermedad o probar una hipótesis etiológica.

Para predecir un cambio individual en el estado de salud con base a ciertas características conocidas (como exposiciones o conductas), debe obtenerse el riesgo de desarrollar el desenlace, como probabilidad condicional que es. La tasa de incidencia no tiene una interpretación útil a nivel individual.

La estimación del riesgo es más relevante para valorar el pronóstico de un paciente, para seleccionar una estrategia terapéutica y para tomar decisiones acerca de conductas relacionadas con la salud.

Por otro lado, si se desea probar una hipótesis etiológica específica respecto a los posibles efectos de uno o más factores, la elección de la medida de incidencia depende de la naturaleza de la enfermedad y la forma en que se va a observar la ocurrencia de los casos nuevos.

En teoría, las pruebas de hipótesis requieren la estimación de tasas para enfermedades crónicas con largos periodos de latencia. En tal situación, el periodo de seguimiento de cada sujeto representa sólo una parte del tiempo total que una persona está en riesgo de desarrollar la enfermedad. Dichas enfermedades tendrían un periodo de riesgo extendido, dado que la susceptibilidad va más allá (antes y después) del periodo de seguimiento observado.

Por otra parte, si la enfermedad tiene un periodo de latencia corto y el factor en estudio representa un evento bien definido, ya sea una exposición de corta duración o un atributo fijo (por ejemplo, grupo sanguíneo), el periodo total de riesgo para cada sujeto estará incluido en el periodo de seguimiento del estudio. Por ejemplo, una enfermedad transmitida por una fuente común y puntual, da lugar a la posibilidad de identificar a todos los casos atribuibles a la exposición, por lo que la enfermedad tiene un periodo de riesgo restringido.

En este caso la tasa instantánea de incidencia variaría importantemente durante el curso del brote.

### • Prevalencia

La observación de los casos existentes, es decir prevalentes de una enfermedad en una población es la característica de los estudios de corte transversal. Debido a que los casos prevalentes representan los sobrevivientes de una enfermedad, no son tan apropiados para identificar factores de riesgo, como los casos incidentes, en una población bien definida.

El conocimiento de la prevalencia de una enfermedad es muy importante en la planeación de los servicios de salud y la administración de las facilidades de la atención médica, pues el número de casos determina la demanda de atención.

Las medidas de prevalencia son útiles también para describir la frecuencia de enfermedades remitentes, o sea, de aquellos que cursan con periodos de sintomatología, y periodos de remisión.

Existen dos tipos básicos para cuantificar la prevalencia en una población: la puntual o momentánea, (o simplemente prevalencia), y la de periodo o láptica.

La primera es la más comúnmente empleada, y corresponde a la probabilidad de que un individuo en una población, sea un caso en el tiempo  $t$ .

Su valor se estima como la proporción de personas que tienen la enfermedad en la población en estudio de tamaño  $N$  en el tiempo  $t$ .

$$P_t = \frac{C_t}{N_t}$$

donde  $C_t = N_t - N'_t$ , que es el número de casos prevalentes en el tiempo  $t$ .

Como puede apreciarse, la prevalencia es una proporción y no una tasa, por tanto, es una medida adimensional cuyo valor se encuentra en el rango 0 a 1.

Una segunda medida, menos frecuentemente utilizada, es la prevalencia de periodo, que es la probabilidad de que un individuo en una población sea un caso en cualquier momento durante un periodo dado ( $t_0, t$ ) de duración  $Dt$ . La prevalencia de periodo es a menudo utilizada como un sustituto de la incidencia acumulada (riesgo) cuando se desconoce el tiempo exacto de inicio de los casos individuales.

En esta situación no es posible distinguir un caso incidente de un caso prevalente.

La prevalencia de periodo (PP) para una población dinámica estable es estimada como la razón del número ( $C_{t_0, t}$ ) de personas que fueron observadas con la enfermedad en cualquier momento durante el periodo de seguimiento ( $t_0, t$ ) con respecto al tamaño de la población ( $N$ ).

$$PP = (C_{t_0, t}) / N = (C_0 + I) / N$$

Donde: ( $C_{t_0, t}$ ) incluye a los casos prevalentes ( $C_0$ ) en el tiempo cero ( $t_0$ ) y a los casos incidentes ( $I$ ) detectados durante el periodo.

La prevalencia de periodo es una razón, y no una proporción, debido a que algunos casos tomados en el numerador pueden no haber estado en la población en todo el periodo de observación ( $t_0, t$ ).

Una medida que puede ser usada para describir la ocurrencia global de una enfermedad remitente en una población, es la prevalencia de por vida. Esta es definida como la probabilidad de que un individuo en el tiempo  $t$  haya sido un caso.

Se estima de la misma manera que la prevalencia puntual, pero en la fórmula

$$P_t = C_t / N_t$$

$C_t$  incluye personas que tienen la enfermedad en el tiempo  $t$ , personas que previamente fueron curadas de la enfermedad, y personas que se encuentran en remisión en el tiempo  $t$ .

La prevalencia de por vida es una medida retrospectiva, por lo que debe ser estimada en base al recuerdo del sujeto y/o de registros médicos.

Aunque en realidad es relativamente más complejo, suele afirmarse que la prevalencia ( $P$ ) depende de la tasa de incidencia ( $I$ ) de la enfermedad, por la duración media ( $D$ ) que ésta tiene:

$$P = I \times D$$

Esta condición es más evidente en el caso de una enfermedad estable y crónica. En esta relación, es posible calcular alguno de los elementos, cuando se conocen los otros dos términos, así:

$$I = P / D \quad \text{y} \quad D = P / I$$

Un cambio en la prevalencia puede ser resultado de un cambio en la incidencia, en el desenlace de la enfermedad, o de ambos, por ejemplo:

- Un mejor tratamiento que impide la muerte, da lugar a un aumento de la prevalencia por una mayor duración de la enfermedad, cuando el restablecimiento no se produce.
- La prevalencia disminuye si se reduce la tasa de incidencia al igual que si la duración de la enfermedad se acorta por una rápida recuperación o por una muerte temprana.
- Cuando la duración disminuye suficientemente, la prevalencia también baja aun cuando la incidencia se incremente.

El presente documento está basado en el libro de Kleinbaum DG, Kupper LL, Morgenstem H. Epidemiologic research. Principles and quantitative methods. New York: Lifetime Learning Publications, Van Nostrand Reinhold Co., 1982: 93-123.