

Aplicación del estimador intrínseco a tasas de mortalidad por cáncer de mama

Nora Arnesi¹ y Leticia Hachuel¹

Forma de citar

Arnesi N, Hachuel L. Aplicación del estimador intrínseco a tasas de mortalidad por cáncer de mama. Rev Panam Salud Publica. 2011;30(3):225-30.

RESUMEN

Objetivo. Valorar la utilización de la técnica del estimador intrínseco (EI) en el campo de la epidemiología.

Métodos. Se aplicó el enfoque EI al análisis de datos sobre cáncer de mama en la Argentina a fin de observar las tendencias asociadas a "edad, período y cohorte" (EPC). Esta metodología recurre a la aplicación de una regresión por componentes principales para obtener un único conjunto de tendencias estimadas. Se compararon sus resultados con los obtenidos por el método convencional "modelos lineales generalizados restringidos" (MLGR) que incluye una restricción adicional a las tradicionales en el modelo estadístico.

Resultados. Ambos métodos proporcionaron resultados compatibles en las tendencias asociadas a EPC, aunque difirieron en los intervalos de confianza, con mayor eficiencia por parte del EI. La curva asociada a edad mostró el patrón de cambio esperado según el transcurso de la vida: a mayor edad, mayor riesgo. En relación a cohortes, se evidenció un decrecimiento de los efectos asociados a las cohortes más recientes, mientras que para período los efectos estimados presentaron muy poca variación.

Conclusiones. La comparación entre los resultados obtenidos por el método EI y el de MLGR reveló los alcances de la solución genérica provista por el EI al problema de estimación en un modelo EPC. El método EI se basa en una transformación de los datos observados utilizando una matriz de ponderaciones de sencilla aplicación y proporciona estimaciones con propiedades estadísticas deseables.

Palabras clave

Modelos estadísticos; epidemiología y bioestadística; cáncer mamario; salud de la mujer; Argentina.

Los eventos demográficos y epidemiológicos, tales como incidencia o mortalidad por cáncer de mama, pulmón o por alguna otra causa específica, están influenciados por la *edad* de los individuos que presentan el evento, el *período* (o momento) en que ocurre y la *cohorte* a la cual pertenecen dichos individuos. La tarea de desentrañar la influencia compleja e interactiva de estas tres variables

estrechamente relacionadas se suele denominar en la literatura como análisis "edad-período-cohorte" (EPC). En este contexto, cuando se recurre a la utilización de modelos estadísticos con el fin de estimar por separado los efectos de estos tres factores sobre el evento en estudio, surgen ciertas dificultades provocadas precisamente por la relación lineal exacta que los caracteriza.

Dado que no hay un consenso generalizado sobre cómo resolver este problema, a lo largo de los años se han propuesto diferentes alternativas, entre las cuales las más sencillas limitan el análisis

a solo dos de esas tres dimensiones, evitando de este modo el problema de la dependencia lineal planteado. Sin embargo, como en el estudio de la mayoría de los eventos demográficos o epidemiológicos resulta importante la consideración simultánea de las dimensiones EPC, se desarrollaron diferentes opciones orientadas a resolver el problema de la identificación al momento de ajustar modelos (1-10).

En general, las opciones más utilizadas se basan fundamentalmente en dos criterios. Uno de ellos se sustenta en imponer a los parámetros del modelo una

¹ Instituto de Investigaciones Teóricas y Aplicadas, Escuela de Estadística, Universidad Nacional de Rosario, Argentina. La correspondencia se debe dirigir a Nora Arnesi. Correo electrónico: narnesi@fcecon.unr.edu.ar

restricción adicional a las comúnmente utilizadas. Este enfoque, conocido como “modelos lineales generalizados restringidos” (MLGR), presenta el inconveniente de que las estimaciones de cada uno de los efectos y sus tendencias varían de acuerdo a dicha restricción adicional. El otro criterio más usado se basa en la definición de funciones estimables, que tienen la ventaja de ser invariantes a las restricciones impuestas pero resultan difíciles de interpretar (11–15).

El denominado “estimador intrínseco” (EI), método de más reciente aparición, integra de alguna manera estos dos criterios, dado que por un lado parte de la definición de funciones estimables y por otro utiliza la descomposición de la matriz del modelo en valores singulares con el fin de restringir el espacio paramétrico de una forma única (16). De este modo el EI proporciona estimaciones robustas de las tendencias de los efectos asociados a edad, período y cohorte, determinadas unívocamente por los coeficientes estimados (17).

El presente trabajo se propone precisamente evaluar la técnica del “estimador intrínseco” en el análisis de datos sobre mortalidad por tumor maligno de mama en la mujer en Argentina y comparar sus resultados con los obtenidos mediante el ajuste de un MLGR con el objetivo de valorar su utilización en el área epidemiológica.

MATERIALES Y MÉTODOS

Tradicionalmente los datos para el análisis EPC se disponen en una tabla denominada “de cohortes” que clasifica, por ejemplo, las tasas de mortalidad por alguna causa de interés según edad y período de registro. Dichas tasas se obtienen a partir del cociente entre el número de muertes ocurridas y_{ij} y el número de personas-año expuestas al riesgo N_{ij} , donde i indica el grupo de edad de los individuos al momento de ocurrencia del evento ($i = 1, \dots, E$) y j el período de registro ($j = 1, \dots, P$) (cuadro 1).

De esta manera, el cuadro 1 revela las tasas de mortalidad por cáncer de mama en Argentina durante el período 1979–2008. Su construcción se basó en datos sobre mortalidad según causas de muerte obtenidos de los “Registros de estadísticas vitales” que publica el Ministerio de Salud, los que fueron relacionados con el correspondiente número de personas-año expuestas al riesgo to-

CUADRO 1. Tabla de cohortes con tasas de mortalidad (por 100 000 personas-año) por tumor maligno de mama, según grupos de edad y períodos, Argentina, 1979–2008 (tabla de cohortes)

Grupo de edad (años)	Período						Cohorte
	1979–1983	1984–1988	1989–1993	1994–1998	1999–2003	2004–2008	
25–29	1,2	1,0	1,2	1,2	1,2	0,8	1979
30–34	5,7	4,3	4,4	4,6	3,8	3,5	1974
35–39	13,1	11,3	11,6	10,2	9,2	8,8	1969
40–44	22,3	22,4	23,5	20,8	19,1	17,3	1964
45–49	33,3	34,2	37,3	36,6	32,6	28,9	1959
50–54	51,5	48,2	45,3	52,5	50,4	43,6	1954
55–59	66,3	65,4	61,6	62,6	61,6	61,3	1949
60–64	71,5	78,0	78,5	74,5	73,1	72,6	1944
65–69	83,0	86,8	91,6	88,0	85,6	84,4	1939
70–74	108,0	109,3	115,5	112,2	107,1	97,0	1934
75–79	119,8	131,6	139,4	140,7	136,9	129,3	1929
	1904	1909	1914	1919	1924		
	Cohorte						

Fuente: elaboración de las autoras con base en las referencias 18 y 19.

mado de datos censales publicados por el Instituto Nacional de Estadísticas y Censos (INDEC) de ese país (18, 19).

En este trabajo las tasas abarcan solamente a los grupos de edad de 25 a 79 años, en razón de que la mortalidad por tumor maligno de mama en menores de 25 años es poco frecuente y la certificación de muerte por esta causa en mayores de 79 años es limitada. Las defunciones y la población se agrupan en $E = 11$ clases quinquenales de edad y $P = 6$ períodos quinquenales.

En el cuadro 1 las filas se refieren a las categorías de edad y sus comparaciones informan acerca de las diferencias en edad de las personas respecto al evento bajo estudio. De igual manera, la comparación entre columnas revela diferencias entre los distintos períodos en los cuales se registró el evento. No obstante, lo que caracteriza a una tabla de cohortes es que incorpora un tercer camino de cotejo entre los valores contenidos en las celdas: la comparación diagonal. En este sentido, las diagonales del cuadro representan cohortes de nacimiento sucesivas siempre y cuando los grupos de edad y los períodos de diagnóstico tengan la misma amplitud. Considerando que cada cohorte se identifica según $K = E - i + j$, en el cuadro se pueden distinguir $K = 16$ cohortes de nacimiento que corresponden a las $K = E + P - 1$ diagonales del cuadro de clasificación cruzada ExP . Dichas cohortes, denominadas sintéticas, se identifican por el año medio del intervalo de cohortes de nacimientos que comprenden.

Modelos edad-período-cohorte

El interés principal de un análisis EPC radica en identificar tendencias o patrones de comportamiento de un evento en estudio según precisamente la edad, el período y la cohorte, sin centrarse en realizar predicciones sobre las tasas asociadas a dicho evento. Tradicionalmente el método que se utiliza para lograr este objetivo es el ajuste de modelos estadísticos a datos sobre el evento de interés presentados en una tabla de cohortes. Sin embargo, debe tenerse en cuenta que existe una combinación lineal perfecta entre los tres variables intervinientes, ya que se verifica que “cohorte = período – edad”.

En principio, un modelo apropiado para la clasificación cruzada de tasas según grupos de edad y período de registro —como la presentada en el cuadro 1— pertenece a la clase de “modelos lineales generalizados” (MLG) (20) y se define según los siguientes supuestos:

- i) El número de ocurrencias del evento en el grupo de edad i , en el período de registro j , y_{ij} , es una realización de una variable aleatoria Poisson con media $\mu_{ij} \ i = 1, \dots, E; j = 1, \dots, P$.
- ii) El número de personas expuestas al riesgo en el grupo edad i en el período j , N_{ij} , es un valor fijo conocido.
- iii) Las variables aleatorias y_{ij} son independientes.
- iv) El logaritmo de la tasa esperada se supone función lineal de efectos de grupos de edad, períodos de registro y cohortes de nacimiento, es decir:

$$\ln\left(E\left(\frac{y_{ij}}{N_{ij}}\right)\right) = \ln\left(\frac{\mu_{ij}}{N_{ij}}\right) = \lambda + \alpha_i + \beta_j + \gamma_k \quad [1]$$

donde:

- λ : media general
- α_i : efecto principal del i -ésimo grupo de edad; $i = 1, \dots, E$
- β_j : efecto principal del j -ésimo período; $j = 1, \dots, P$
- γ_k : efecto principal de la k -ésima cohorte; $k = 1, \dots, E + P - 1$

El valor correspondiente al subíndice k queda entonces determinado al asignar al individuo un grupo de edad y un período. De ahí el uso de solo dos subíndices en y, N y μ .

Las restricciones habituales que se imponen a los parámetros son:

$$\sum_i \alpha_i = \sum_j \beta_j = \sum_k \gamma_k = 0 \quad [2]$$

o alternativamente identificar uno de los niveles de edad, período y cohorte como categoría de referencia, es decir:

$$\alpha_i = 0; \beta_j = 0; \gamma_k = 0 \quad [3]$$

para i, j y k elegidos convenientemente.

El número de personas-año en riesgo N_{ij} surge de una estimación basada en datos censales y su variación se considera no aleatoria. Por lo tanto su coeficiente no debe ser estimado, con lo cual el modelo se puede re-exresar de la siguiente manera:

$$\ln(\mu_{ij}) = \ln(N_{ij}) + \lambda + \alpha_i + \beta_j + \gamma_k \quad [4]$$

La dependencia lineal exacta entre los factores dada por: cohorte = período - edad determina, en términos de los índices de los parámetros, que el subíndice que identifica a la cohorte se relaciona con los subíndices que acompañan a los grupos de edad y períodos a través de la siguiente ecuación:

$$k = E - i + j, \quad [5]$$

Esta es la razón por la cual se hace necesario incluir una restricción adicional a las habituales, como las especificadas en [2] y [3], a fin de estimar los parámetros y que resulten identificables —por ejemplo, imponer la igualdad entre dos coeficientes de regresión asociados a dos niveles diferentes de la misma variable da lugar a un particular MLGR.

El principal logro del EI es que permite despegarse de la arbitrariedad de la elección de esta restricción adicional y resulta ser una estrategia útil para obtener estimaciones robustas de las tendencias temporales asociadas a cada uno de los factores *edad, período y cohorte*.

El estimador intrínseco

Para presentar el EI es conveniente expresar el modelo completo [4] en forma matricial de la siguiente manera:

$$\ln(\mu) = \ln(N) + Xb \quad [6]$$

y considerar la restricción [2].

En esta expresión, μ es el vector del número esperado de eventos cuyos elementos son: $\{\mu_{ij} = E(y_{ij})\}$; $\ln(N)$, con $N = \{N_{ij}\}$, es el término considerado como "offset" o variable de exposición en el modelo loglineal; X es la matriz del modelo formada por vectores columnas de variables indicadoras y b , de dimensión $m = 1 + (E - 1) + (P - 1) + (E + P - 2)$, es el vector de parámetros linealmente independientes del modelo formado por:

$$b = (\lambda, \alpha_1, \dots, \alpha_{E-1}, \beta_1, \dots, \beta_{P-1}, \gamma_1, \dots, \gamma_{E+P-2})^T \quad [7]$$

El problema de la identificación, inducido por la relación lineal exacta entre edad, período y cohorte, puede ser visto técnicamente como un caso especial de regresores colineales que conducen a que la matriz del modelo X sea singular con un rango menor en una unidad al rango completo. Este hecho produce múltiples estimadores de los efectos, lo que dificulta estimar en forma separada cada uno de ellos y sus tendencias.

Dado que la matriz del modelo tiene rango columna menor que el completo, el espacio paramétrico se puede descomponer como la suma directa de dos espacios lineales que son perpendiculares entre sí: uno de ellos es el espacio nulo de X , unidimensional, y el otro es el subespacio complemento ortogonal del espacio nulo.

La descomposición ortogonal del espacio paramétrico permite que cada una de las infinitas estimaciones \hat{b} del modelo se pueda a su vez descomponer de la forma:

$$\hat{b} = B + tB_0 \quad [8]$$

El vector B , denominado estimador intrínseco, es único y resulta ortogonal a

B_0 , autovector normalizado asociado al autovalor nulo de $(X^T X)$, es decir, el EI no tiene componentes en la dirección de B_0 , mientras que tB_0 representa tendencias lineales arbitrarias.

Para obtener B se realiza la proyección de cualquier estimación \hat{b} sobre el espacio ortogonal al nulo, esto es:

$$B = (I - B_0 B_0^T) \hat{b} \quad [9]$$

siendo I la matriz identidad. La matriz de ponderaciones $(I - B_0 B_0^T)$ es única para un conjunto fijo de edades y períodos, mientras que \hat{b} puede variar de acuerdo a la restricción adicional de igualdad que se imponga a los coeficientes.

Resumiendo, la idea básica del EI es eliminar la influencia de la matriz del modelo X sobre los coeficientes estimados, dado que dicha matriz está determinada únicamente por el número de grupos de edad y períodos sin tener en cuenta el recuento del evento bajo estudio.

Otra forma de calcular el EI se basa en el método de regresión por componentes principales, el cual permite obtener fácilmente los desvíos estándares de los coeficientes estimados (16). Desde este enfoque, el EI resulta un tipo especial de estimador obtenido a partir del método de componentes principales, ya que conlleva un paso adicional a fin de referir las estimaciones al espacio original de las coordenadas edad, período y cohorte, proporcionando así coeficientes comparables con los que arroja la aplicación de un MLGR. Además, el EI tiene propiedades estadísticas deseables, como la de ser insesgado para muestras finitas, eficiente y consistente o —dicho de otro modo— asintóticamente insesgado para muestras grandes (17).

En el presente trabajo la implementación del EI se lleva a cabo a partir del desarrollo de una rutina en el entorno y ambiente computacional R,² que garantiza la comprensión exhaustiva de su cálculo (21). No obstante, esta metodología se está incorporando en paquetes estadísticos de uso frecuente³ y también está disponible una página en internet de uso gratuito que efectúa el cálculo online.⁴

² Software de distribución libre (www.r-project.org.ar).

³ Macro apc del software STATA.

⁴ Disponible en: http://apcsoft.epi.msu.edu/cgi-bin/sign_up.pl Acceso el 02 de junio de 2011.

RESULTADOS

Se analizaron los datos dispuestos en una tabla de cohortes (cuadro 1) utilizando modelos EPC y se aplicó la metodología del EI con objeto de obtener estimaciones de los efectos asociados a edad, período (influencia secular) y cohorte (factor generacional).

Dado que el interés se centró en las tendencias que arrojaron las estimaciones de los efectos asociados a las tres variables y no en el valor de los coeficientes en sí mismos, se consideró que la representación gráfica era el mejor modo de presentarlas. Asimismo, con el propósito de valorar los resultados se los comparó con los obtenidos con un MLGR, en el cual se impuso como restricción adicional la igualdad de los coeficientes de dos períodos consecutivos.

En las figuras 1A y 1B se muestran las estimaciones puntuales y por intervalos obtenidas con el ajuste del MLGR y mediante la aplicación del EI. Como se puede observar, las tendencias obtenidas a partir de los dos estimadores para los efectos asociados a edad fueron similares, aunque el rango de las estimaciones resultó diferente debido a que usaron distintas restricciones para identificar los parámetros. En cuanto a las estimaciones asociadas a período, las obtenidas a partir del MLGR mostraron un comportamiento prácticamente constante, mientras que las extraídas del EI revelaron un leve crecimiento. La tendencia general decreciente en los efectos asociados a cohorte fue similar, pero en este caso las pendientes y el rango de los coeficientes estimados difirieron.

De la comparación de ambas figuras se desprenden dos comentarios adicionales concernientes a las estimaciones por intervalos de confianza. En primer lugar, para las estimaciones obtenidas mediante el MLGR los intervalos crecieron en amplitud —se ensancharon— a través de las categorías de edad, período y cohorte; no ocurrió lo mismo con las estimaciones obtenidas mediante el EI, en las cuales los intervalos conservaron casi la misma amplitud. El ancho creciente de los intervalos para las estimaciones obtenidas con el MLGR se debió a la particular elección de las categorías de referencia. Por ejemplo, si en lugar de elegir la primera categoría de edad, período y cohorte como categorías de referencia se eligen las últimas, las bandas de confianza serán más anchas para los

FIGURA 1A. Coeficientes estimados a partir de un modelo lineal generalizado restringido

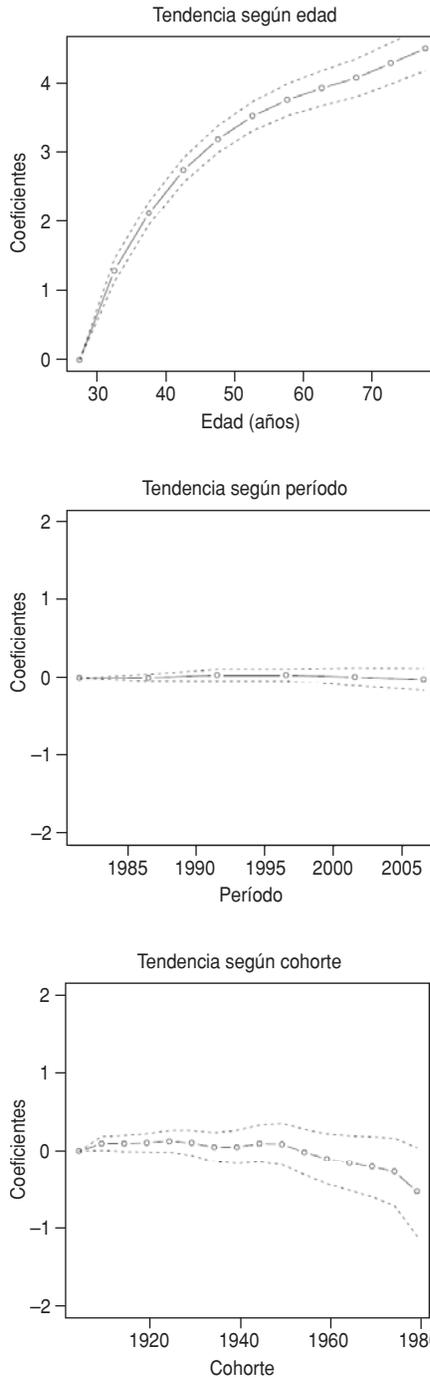
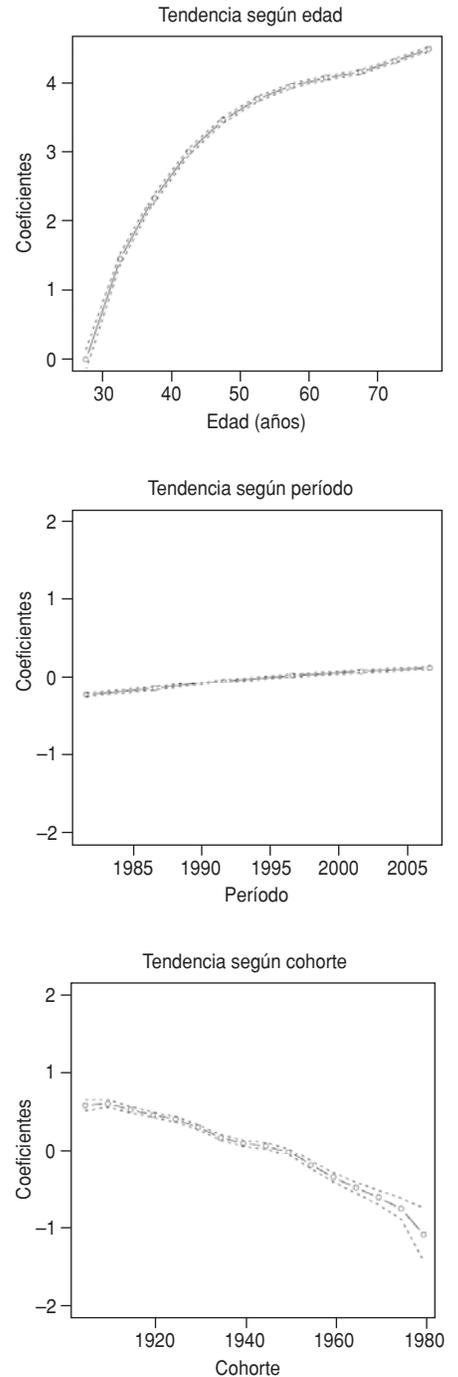


FIGURA 1B. Coeficientes estimados a partir del estimador intrínseco



Fuente: elaboración de las autoras.

primeros grupos de edad, período y cohorte y su amplitud decrecerá a medida que se acercan a las últimas categorías.

Además se puede ver que, en general, los intervalos de confianza estimados a partir del MLGR fueron más anchos que los obtenidos mediante el EI, demos-

trando que en este sentido el EI es eficiente en relación con el estimador MLGR (16). Sin embargo, aun con estas diferencias, las tendencias detectadas por ambos métodos fueron similares, como se verá en los párrafos que las reseñan a continuación.

La tendencia asociada con *edad* resultó muy similar en ambos métodos. La curva crece de forma acelerada hasta aproximadamente los 55 años, luego presenta un crecimiento más leve hasta los 70 años, edad a partir de la cual el crecimiento se vuelve más pronunciado. Este fenómeno ha sido observado en diferentes países y se suele interpretar como resultante del solapado de dos curvas que se corresponden con la aparición de tumores pre y post menopáusicos, respectivamente.

En relación con el efecto *período*, los resultados obtenidos por el EI mostraron una leve tendencia creciente en los últimos años, comportamiento que no queda claro ya que se contrapone con los hallados en otros estudios. Este resultado podría deberse a cambios en la calidad de los registros, que compensarían en demasía la disminución que se esperaría produzcan los adelantos en detección temprana y la aplicación de terapia adyuvantes.

Respecto al efecto *cohorte*, ambas metodologías produjeron tendencias decrecientes hacia las cohortes más recientes, aunque en el caso del EI la caída fue mayor. Tales reducciones podrían atribuirse a la progresiva conciencia de las mujeres de la necesidad de chequeos rutinarios que favorezcan la detección temprana, algo que no obstante resulta cuestionable a la vista de indicadores que sugieren un incremento de factores de riesgo —como obesidad y alcoholismo— en las cohortes más jóvenes. Por otro lado, se debe ser cuidadoso en estas interpretaciones, dado que estas últimas cohortes son las que tienen menor número de observaciones y podrían ser menos confiables (22, 23).

DISCUSIÓN

El estudio de las tasas de un evento de interés clasificadas en forma múltiple según la edad de los individuos que presentan dicho evento y el momento o pe-

riodo en que se registra es frecuente en el campo de la epidemiología, la demografía y las ciencias sociales en general. En el caso de los modelos EPC, lo que se busca principalmente es identificar si los cambios temporales están asociados al *período* en que se registra el evento o a la *cohorte* de nacimiento, ya que se considera la influencia de la *edad* de carácter indiscutible. Tal identificación, sin embargo, resulta engorrosa debido a la relación lineal que existe entre estas tres variables.

La solución convencional al problema de la identificación impone una o más restricciones de igualdad sobre los coeficientes del vector de parámetros que se van a estimar. Esta opción tiene como limitación principal la necesidad de recurrir a información previa o externa para determinar cuál es la restricción adicional más adecuada ya que diferentes elecciones pueden producir distintos patrones de cambio.

El denominado estimador intrínseco es un método de estimación que conduce a una única solución que, adicionalmente, posee propiedades estadísticas deseables. Construye una serie de variables independientes —no correlacionadas entre sí— mediante la aplicación de una matriz de transformación, permitiendo de este modo tomar en cuenta la variabilidad explicada por *edad* y *período* separándolos de los efectos de *cohorte*.

El presente trabajo tuvo como propósito comprobar empíricamente los alcances de la metodología EI, confrontando sus resultados con los de la metodología convencional. Para el problema particular analizado (mortalidad por cáncer de mama en Argentina), la comparación de los resultados obtenidos con el enfoque EI y con el tradicional no arrojaron mayores discrepancias. Sí se observaron diferencias en los resultados al compararlos con estudios similares realizados en otros países de América Latina, fundamentalmente los referidos a la tendencia asociada a cohorte (22, 24).

Desde el punto de vista epidemiológico, esta discrepancia amerita profundizar en la explicación de las tendencias asociadas a edad, período y cohorte en el problema de salud abordado, teniendo en cuenta no solo los posibles errores de los registros sino también la compleja red de factores biológicos, socioeconómicos y de estilos de vida que contribuyen a provocarlo.

Desde un punto de vista metodológico, el EI resulta satisfactorio en la separación de efectos según edad, período y cohorte sin recurrir a restricciones adicionales que normalmente tienen algún signo de arbitrariedad. Es por lo tanto una herramienta útil para adquirir conocimiento acerca de dichos efectos en diversos estudios de diferentes áreas. Para su cálculo se dispone de programas computacionales accesibles, aunque no es tarea sencilla comprender técnicamente cómo actúa para lograr neutralizar los efectos de edad y período de la estimación de los efectos asociados a cohorte.

Por último, hay que tener en cuenta las limitaciones de todo modelo estadístico y considerar sus resultados con cautela. En principio, la replicación en caso de disponer de datos adicionales es una buena práctica para confirmar comportamientos, pero más allá de la técnica que se utilice para la obtención de los estimadores, el modelo planteado supone una cierta limitación: la aditividad de efectos, es decir el efecto asociado a cada grupo de edad igual para todos los períodos y todas las cohortes. Lo mismo ocurre con los efectos asociados a período y cohorte, que permanecen constantes para los distintos niveles de las restantes variables. Estas consideraciones pueden dar lugar a una pobre interpretación de los cambios en el tiempo, por lo que sigue vigente la necesidad de desarrollar nuevos modelos y técnicas dirigidos a lograr una buena descripción de los fenómenos en estudio.

REFERENCIAS

1. Fienberg SE, Mason WN. Identification and estimation of age-period-cohort in the analysis of discrete archival data. *Sociol Methodol.* 1979;(10):1-67.
2. Hageaars JA, Cobben NP. Age, cohort and period. A general model for the analysis of social change. *Neth J Sociol.* 1978;14:59-92.
3. Hobcraft J, Menken J, Preston S. Age, period and cohort effects in demography: a review. *Popul Index.* 1982;48:4-43.
4. Decarli A, La Vecchia C, Cislighi C, Fasoli M, Mezzanotte G, Negri E. La mortalità per tumori in Italia, 1955-1979: effetti dell'età, della generazione di nascita e del periodo di morte [monografía]. Milano: Lega Italiana per la Lotta contro I Tumori; 1986.
5. Decarli A, La Vecchia C. Age, period and cohort models: A review of knowledge and implementation in GLIM. *Rev Stat Appl.* 1987; 20:397-410.
6. Osmond C, Gardner MJ. Age, period and cohort models applied to cancer mortality rates. *Stat Med.* 1982;1:245-59.
7. Kupper L, Janis J, Karmous A, Greenberg B. Statistical age-period-cohort analysis: a review and critique. *J Chronic Dis.* 1985;38:811-30.
8. Robertson C, Boyle P. Age-period-cohort analysis of chronic disease rates. I: modelling approach. *Stat Med.* 1998;17:1305-23.

9. Carstensen B. Sensible parametrizations of Age-Period-Cohort models for disease incidence data. Gentofte, Denmark: Steno Diabetes Center; 2001.
10. Arbyn M, Van Oyen H, Sartor F, Tibaldi F, Molenberghs G. Description of influence of age, period and cohort effects on cervical cancer mortality by loglinear Poisson models. (Belgium, 1955–94). Arch Public Health. 2002; 60(2):73–100.
11. González JR, Llorca FJ, Moreno V. Algunos aspectos metodológicos sobre los modelos edad-período-cohorte: aplicación a las tendencias de mortalidad por cáncer. Gac Sanit. 2002;16(3):267–73.
12. Clayton D, Schifflers E. Models for temporal variation in cancer rates II: age-period-cohort models. Stat Med. 1987;6:469–81.
13. Holford TR. The estimation of age, period and cohort effects for vital rates. Biometrics. 1983; 39:311–24.
14. Holford TR. Understanding the effects of age, period, and cohort on incidence and mortality rates. Annu Rev Public Health. 1991;12:425–57.
15. Holford TR. Analysing the temporal effects of age, period and cohort. Stat Methods Med Res. 1992;1:317–37.
16. Yang Y, Fu W, Land KC. A Methodological Comparison of Age-Period-Cohort Models: The Intrinsic Estimator and Conventional Generalized Linear Models. Sociol Methodol. 2004;34:75–110.
17. Fu WJ, Hall P, Rohan TE. Age-Period-Cohort Analysis: Structure Estimators, Estimability, Sensitivity, and Asymptotics [technical report]. Department of Epidemiology, Michigan State University; 2004.
18. Ministerio de Salud, Argentina. Anuarios. Disponible en: www.msal.gov.ar. Acceso el 10 de marzo de 2010.
19. Instituto Nacional de Estadísticas y Censos (INDEC), Argentina. Estimaciones y proyecciones de población. Total del país, 1950–2015. Disponible en: www.indec.mecon.gov.ar. Acceso el 10 marzo de 2010.
20. Agresti A. An Introduction to Categorical Data Analysis. Wiley Series in Probability and Statistics. New York: Wiley; 1996.
21. Arnesi N. Modelos edad-período-cohorte para el estudio de eventos demográficos y epidemiológicos [Tesis de Maestría en Estadística Aplicada]. Rosario, Argentina: Universidad Nacional de Rosario; 2007.
22. Franco-Marina F, Lazcano-Ponce E, López-Carrillo L. Breast cancer mortality in Mexico: an age-period-cohort analysis. Salud Publica Mex. 2009;51 supl 2:s157–64.
23. Cayuela A, Rodríguez-Domínguez S, Ruiz-Borrego M, Gili M. Age-period-cohort analysis of breast cancer mortality rates in Andalucía (Spain). Ann Oncol. 2004;15(4):686–8.
24. Hernandez G, Herrán S, Cantor L. Análisis de las tendencias de mortalidad por cáncer de mama en Colombia y Bogotá, 1981–2000. Rev Col Cancerol. 2007;11(1):32–9.

Manuscrito recibido el 1 de octubre de 2010. Aceptado para publicación, tras revisión, el 5 de abril de 2011.

ABSTRACT

Application of the intrinsic estimator to breast cancer mortality rates

Objective. Assess use of the intrinsic estimator (IE) technique in epidemiology.

Methods. The IE approach was applied to the analysis of breast cancer data in Argentina in order to observe the trends associated with “age, period, and cohort” (APC). This method involves the use of a principal components regression to obtain a single set of estimated trends. The results were compared to the findings obtained with the conventional method, which consists of adjusting a generalized linear model that includes the traditional constraints of the statistical model as well as an additional constraint (CGLM).

Results. Both methods yielded compatible results in the trends associated with APC. However, they differed in the confidence intervals, with IE yielding greater efficiency. The curve associated with age showed the expected pattern of change across the life course: the greater the age, the greater the risk. With regard to cohorts, a decrease in the effects associated with the most recent cohorts was evident, whereas there was very little variation in the estimated effects for the period.

Conclusions. A comparison of the results obtained with the IE method and the CGLM method revealed the reach of the generic solution provided by the IE to the problem of estimates in an APC model. The IE method is based on conversion of the data observed using a weighting matrix that is simple to apply and provides estimates with desirable statistical properties.

Key words

Models, statistical; epidemiology and biostatistics; breast neoplasms; women’s health; Argentina.