

# ANÁLISIS EDAD-COHORTE-PERIODO DE INCIDENCIA Y MORTALIDAD POR LINFOMA DE HODGKIN EN LA PROVINCIA DE ZARAGOZA

J.J. Abellán<sup>1,3</sup>, C. Feja<sup>2</sup>, A. López-Quílez<sup>3</sup> y C. Martos<sup>2</sup>

<sup>1</sup> Institut Valencià d'Estadística

<sup>2</sup> Departamento de Salud y Consumo. Gobierno de Aragón

<sup>3</sup> Departament d'Estadística i Investigació Operativa. Universitat de València.

## Introducción

El estudio de la tendencia de la incidencia y mortalidad se ha utilizado en salud pública para realizar predicciones futuras del comportamiento de una patología determinada, así como para sugerir hipótesis etiológicas. El estudio de la evolución de las tasas ajustadas por el método directo, tiene la ventaja de la simplicidad, pero tiene limitaciones, ya que no proporciona información relevante como el efecto de la edad, la cohorte de nacimiento y periodo del diagnóstico o defunción sobre la tendencia. Así, información detallada sobre las tasas específicas por edad proporcionan información sobre si los patrones de la enfermedad es consistente con la biología de la misma, diferentes cohortes de nacimiento han podido tener diferentes niveles de exposición y por último, cambios en los procedimientos diagnósticos y terapéuticos pueden influir en la incidencia y mortalidad de la enfermedad.

Hasta el momento, para estudiar el efecto edad, cohorte y periodo se han utilizado modelos de Poisson log-lineal. Mientras que su utilización para enfermedades frecuentes puede ser adecuado, cuando se trata de patologías raras o áreas geográficas pequeñas estos modelos no pueden ser utilizados por la inestabilidad de las estimaciones causada por la gran cantidad de grupos de edad-cohorte-periodo sin observaciones (Holford, 1991).

Este trabajo aborda el análisis de los efectos de la edad, la cohorte de nacimiento y el periodo en la incidencia y en la mortalidad por enfermedades raras. Proponemos varios modelos jerárquicos bayesianos alternativos para ajustar los datos, estudiar los efectos, y realizar proyecciones para los próximos periodos.

Se han ensayado cinco modelos con formulaciones alternativas basadas en una modelización jerárquica bayesiana. El primer modelo se corresponde con la versión bayesiana equivalente al modelo frecuentista habitual de una regresión de Poisson donde la variable respuesta es el número de óbitos, y las covariables son los factores edad, cohorte y periodo, con el número de personas-año como "offset" (McCullagh y Nelder, 1989). Los restantes se han configurado partiendo de la estructura autoregresiva de los efectos de cada factor propuesta en la metodología de Bray et al. (2001). Las

variantes sobre este modelo son planteadas con el ánimo de recoger posibles mejoras conceptuales del mismo. Todos ellos se han ensayado sobre los mismos datos.

## Datos

Los datos sobre los que se han aplicado los modelos corresponden a la incidencia y la mortalidad por linfoma de Hodgkin en la provincia de Zaragoza entre 1979 y 1998, agregada por grupos de edad y periodos de mortalidad. Se han considerado 17 grupos quinquenales de edad y 4 periodos también quinquenales.

A partir de la edad y el periodo del fallecimiento se puede calcular la cohorte de nacimiento. Así, en nuestro caso hay 20 cohortes, que lógicamente serán también quinquenales.

El linfoma de Hodgkin es un tumor poco frecuente, considerado raro en mujeres. El número total de casos observados es pequeño, por lo que presentan una elevada proporción de ceros entre los conteos por agrupaciones de edad y periodo.

## Modelos

Para un grupo de edad en un periodo determinado es posible calcular el año aproximado de nacimiento, denominado cohorte. El análisis de edad-cohorte-periodo permite describir las tasas de incidencia o mortalidad en términos de los efectos de la edad, del periodo y de la cohorte. El efecto de la cohorte muestra el riesgo diferenciado relacionado con factores que afectan a un grupo de personas nacidas en una misma época. Los modelos que se han considerado son descritos a continuación.

El primer modelo se puede considerar la versión bayesiana equivalente en cierta manera al modelo frecuentista habitual de una regresión de Poisson donde la variable respuesta es el número de ingresos o de óbitos, y las covariables son los factores edad, cohorte y periodo, con el número de personas-año como "offset". En este modelo los efectos de los factores se consideran independientes, si bien se espera que todos ellos se encuentren alrededor del 0 en escala logarítmica, es decir, que su RR correspondiente esté alrededor de 1, según una distribución normal.

El segundo es un modelo jerárquico que corresponde al expuesto por Bray et al. (2001). Como en el modelo básico, establece en su primer capa que la tasa bruta depende del grupo de edad, del periodo y de la cohorte. En la segunda capa se asignan las distribuciones iniciales de los efectos edad, cohorte y periodo, de manera que se asume que los efectos en los dos primeros grupos de edad (el de 0 a 4 años y el de 5 a 9) son normales, de media 0 (en escala logarítmica) y una cierta precisión común a ambos. Para el efecto del resto de grupos de edad se asume una distribución normal, cuya media se hace depender de forma autorregresiva de los efectos de los dos grupos de edad inmediatamente anteriores, y con igual precisión. Y de forma análoga se

definen las distribuciones iniciales de los otros dos factores, periodo y cohorte.

El tercer modelo trata de flexibilizar el modelo anterior. En concreto, los coeficientes autorregresivos pasan a ser parámetros en lugar de dejarlos fijos en 2 y en  $-1$  para todos los factores. Para todos ellos se han considerado hiperdistribuciones iniciales de una distribución uniforme en el intervalo  $(-5, 5)$ .

El cuarto modelo pretende relajar la igualdad de las precisiones para los diferentes grupos del mismo factor. Específicamente, las precisiones de los dos primeros grupos del factor edad, de los dos primeros grupos del factor periodo y los propios del factor cohorte, se han modelizado como parámetros independientes, de manera que puedan ser distintas a las del resto de niveles del mismo factor. Así, en lugar de tener una única precisión para el factor edad, se tienen 3. Y análogamente para los otros dos factores, cohorte y periodo. Esto es así porque los dos primeros efectos, aún tomando un valor cercano a 0, pueden tener un rango más amplio de valores posibles, es decir, pueden tener precisiones menores, mientras que el resto de efectos deberían tener (si su comportamiento es realmente regresivo) un valor muy cercano al término autoregresivo que los liga a los anteriores.

Por último, el quinto modelo considerado es un término intermedio entre el primero y el tercero, tomando como autoregresivos los efectos del factor periodo y manteniendo independientes los relacionados con edad y cohorte.

## Resultados

El análisis bayesiano de estos modelos es posible mediante los métodos de simulación MCMC de la distribución final de los parámetros. Para el proceso de simulación hemos empleado del programa WinBUGS (Spiegelhalter et al., 2002). Este programa de libre distribución implementa técnicas de simulación mediante Gibbs Sampling (Green, 2001) y facilita el análisis bayesiano de modelos jerárquicos como los descritos.

A partir de las simulaciones obtenidas para los diferentes efectos estudiados, obtenemos sus respectivas distribuciones finales simuladas, pudiendo establecer sus valores medios y sus intervalos de credibilidad. El alejamiento del 0 en alguno de los efectos mostraría la importancia del mismo.

Los resultados obtenidos no revelan importantes diferencias entre los grupos considerados. Únicamente algunos grupos de edad (entre 10 y 34 años) presentan intervalos de credibilidad por encima del 0 en los datos de incidencia, tanto en hombres como en mujeres, denotando un riesgo superior.

## Agradecimientos

Este trabajo ha sido parcialmente financiado por la Dirección General de Salud Pública de la Generalitat Valenciana mediante un acuerdo de colaboración con la

Universitat de València.

## Referencias

- Bray, I., Brennan, P. y Boffetta, P. (2001). Recent trends and future projections of lymphoid neoplasms – a bayesian age-period-cohort analysis. *Cancer Causes and Control* 12:813–820.
- Green, P. (2001). *Complex Stochastic Systems*, capítulo A primer of Markov Chain Monte Carlo, 1–62. Chapman and Hall/CRC.
- Holford, T. R. (1991). Understanding the effects of age, period and cohort on incidence and mortality rates. *Annual Review of Public Health* 12:425–457.
- McCullagh, P. y Nelder, J. (1989). *Generalized linear models*. Chapman and Hall, 2.
- Spiegelhalter, D. J., Thomas, A., Best, N. G. y Lunn, D. (2002). Winbugs version 1.4 user manual. T. Rep., MRC Biostatistics Unit.  
URL <http://www.mrc-bsu.cam.ac.uk/bugs>