

# La distribución binomial negativa frente a la de Poisson en el análisis de fenómenos recurrentes

A. Navarro<sup>a</sup> / F. Utzet<sup>b</sup> / P. Puig<sup>b</sup> / J. Caminal<sup>a</sup> / M. Martín<sup>a</sup>

<sup>a</sup>Laboratori de Bioestadística i Epidemiologia. Facultat de Medicina. Universitat Autònoma de Barcelona.

<sup>b</sup>Servei d'Estadística de la Universitat Autònoma de Barcelona.

*Correspondencia:* A. Navarro. Laboratori de Bioestadística i Epidemiologia. Facultat de Medicina. UAB. 08193 Cerdanyola del Vallès. Barcelona. Correo electrónico: Albert.Navarro@Uab.es

*Recibido:* 19 de enero de 2001.

*Aceptado:* 20 de julio de 2001.

(Negative binomial distribution versus Poisson in the analysis of recurrent phenomena)

## Resumen

*Objetivos:* Exponer la posible problemática en el cálculo de riesgos en bases de datos agregadas cuando el fenómeno estudiado es recurrente y presentar la distribución binomial negativa como una alternativa válida y sencilla para analizar este tipo de fenómeno.

*Métodos:* En el contexto de los fenómenos recurrentes, el análisis mediante la regresión de Poisson puede provocar sobredispersión o variancia extra-Poisson. Esto conduce a la subestimación de los errores estándares de los coeficientes, pudiendo derivar en la significación estadística de factores que realmente no estén asociados con el fenómeno. La binomial negativa puede captar parte de la variancia que no identifica la regresión de Poisson. Para comprobarlo se comparó ambas distribuciones sobre el número de hospitalizaciones que presentaron individuos, entre 65 y 69 años de edad, durante el año 1996. Esta comparación fue realizada en dos bases de datos agregadas distintas: por individuo y según las variables de interés.

*Resultados:* El ajuste mediante ambas distribuciones presenta diferencias en las dos bases de datos. Según el estudio de los residuos, en la base por individuo la binomial negativa ajusta correctamente el 67,9% de las observaciones mal ajustadas por la regresión de Poisson. Este porcentaje es del 50% en la base agregada según las variables. Además, en ambos casos, la regresión de Poisson estima significativas cuatro de las seis variables estudiadas. Para la binomial negativa son dos en la base por individuo y una en la base por variables.

*Conclusiones:* La existencia de sobredispersión es frecuente en fenómenos recurrentes. Cuando esto sucede, el uso de la binomial negativa es más apropiado que el de la regresión de Poisson.

**Palabras clave:** Binomial negativa. Sobredispersión. Extra-Poisson. Episodios recurrentes. Poisson.

## Summary

*Objective:* The aim is to unfold the difficulties likely to arise in risk calculations through aggregated database when the studied phenomenon is recurrent and to display the negative binomial distribution as a valid and simple alternative to analyse this kind of phenomenon.

*Methods:* When the studied phenomenon is recurrent, the analysis by means of the Poisson regression can provoke overdispersion or extra-poisson variance, what leads to underestimating the standard errors in coefficients and may divert into the statistical significance of factors which as a matter of fact are not associated with the phenomenon beforehand. The negative binomial can grasp part of the variance which the Poisson is unable to identify. In order to check this out, the fit of both distributions were compared, based on the number of hospitalizations of individuals aged between 65 and 69, during 1996. This comparison was carried out by means of two different aggregated databases: by individuals and by variables.

*Results:* There were differences in the fitted models by means of both distributions in both databases. By the analysis of the residuals, when using the base by individuals, the negative binomial fits correctly 67.9% of the observations badly fitted by the Poisson. Using the aggregated variables database, the percentage is 50%. In both cases, Poisson estimates four out of the six studied variables as significant. As to the negative binomial, there are two significant based on individuals and one in the variable database.

*Conclusion:* The existence of overdispersion is frequent in recurrent-type phenomena. When this occurs, the negative binomial distribution is more appropriate than the Poisson.

**Key words:** Negative binomial. Overdispersion. Extra-Poisson. Recurrent events. Poisson.

## Introducción

El conjunto de indicadores más utilizados en epidemiología es el constituido por las tasas de densidad de incidencia y de incidencia acumulada, tanto crudas como estandarizadas. A partir de las mismas pueden estimarse los riesgos relativos y atribuibles a la presencia o ausencia de diversos factores asociados a la aparición o no de un fenómeno concreto.

Si el fenómeno que se estudia permite que un individuo presente el mismo episodio más de una vez (número de hospitalizaciones, consultas ambulatorias, ataques de asma, etc.), éste es recurrente. Cronológicamente, los métodos de comparación y evaluación de los efectos asociados a los factores estudiados han sido la estandarización, para comparar tasas de distintas poblaciones, y el análisis multivariante de estas mediante la regresión de Poisson (POI). En este último caso, pueden plantearse serios problemas si se incumplen las condiciones de aplicabilidad inherentes al método. En efecto, en la mayor parte de supuestos reales, los fenómenos que se estudian presentan elevados niveles de recurrencia, por lo que utilizar POI puede llevar a conclusiones, como mínimo, inadecuadas, cuando no equivocadas. La metodología para analizar estas situaciones hasta hace pocos años no ha sido muy utilizada en la práctica, debido fundamentalmente a la carencia de herramientas informáticas adecuadas. Hoy día, se puede acceder fácilmente a diversas alternativas a las existentes para el análisis de este tipo de datos<sup>1,2</sup>.

Tampoco puede ignorarse la estructura de la base de datos disponible, en términos de niveles de agregación, ya que condiciona el uso del tipo de técnica alternativa de análisis. En el contexto de las bases de datos agregadas, que son las de acceso más frecuente por la facilidad de su manejo, la distribución binomial negativa (BN) parece ser adecuada para modelar tasas de fenómenos recurrentes que incumplan las suposiciones de la POI.

El objetivo de este estudio consiste en exponer la posible problemática que generan los fenómenos recurrentes en bases de datos agregadas y presentar la BN como una alternativa válida y sencilla para analizar este tipo de fenómeno.

## Sujetos y métodos

Las técnicas clásicas de modelización de tasas obviaban la problemática que puede presentarse al analizar fenómenos recurrentes. Éstos son aquellos en que el episodio de interés puede producirse en más de una ocasión en un mismo individuo. Ante un fenómeno de

este tipo debe verificarse si éste presenta alguna de las siguientes características<sup>3</sup>:

1. La probabilidad de la aparición del episodio es la misma para todos los individuos, pero depende de episodios previos producidos en cada uno. Es decir, tener un episodio modifica la probabilidad de presentar uno nuevo (episodios dependientes). En este caso existe una dependencia temporal intrapaciente llamada «modelo de contagio».

2. La probabilidad del evento permanece constante a lo largo del tiempo pero no necesariamente es el mismo para cada individuo. Existen individuos más propensos a padecer el evento, pero el hecho de haberlo presentado anteriormente no modifica la probabilidad de volver a tenerlo (eventos independientes). Por tanto se produce una heterogeneidad interpaciente, conocida como «modelo de fragilidad» o «modelo de propensión».

Si no se observa ninguna de estas características, las condiciones de aplicabilidad de la POI (es decir, episodios independientes y aleatorios entre los individuos) se cumplen y, por tanto, su uso puede estar justificado. En cualquiera de las posibles situaciones restantes, contagio y/o propensión, se incumple claramente alguna de estas condiciones. Si esto sucede, modelar el número de episodios en estudio mediante la POI provoca el fenómeno conocido como sobredispersión o variancia extra-Poisson. Ésta, especialmente en el caso de propensión, puede reflejar la influencia de variables no observadas y posiblemente esté indicando que el fenómeno analizado se distribuye mediante una mezcla de distribuciones Poisson, lo que es, matemáticamente, de difícil manejo. La sobredispersión puede identificarse cuando la variancia observada supera a la media, es decir, cuando la variancia observada es superior a la teórica, ya que en la POI la media ( $\mu$ ) y la variancia teórica son idénticas. Aunque la sobredispersión no invalida las estimaciones puntuales (es decir, la estimación de las tasas y sus razones) ya que éstas son robustas si el modelo está correctamente especificado, produce una subestimación de la variancia real<sup>4,5,6</sup>. En la práctica, este hecho conduce a una estimación del intervalo de confianza de la tasa extremadamente estrecho<sup>7</sup>, dando una falsa exactitud de éste.

El uso de la BN puede estimar parte de la variancia que no consigue identificar la POI. Así, la variancia de la BN<sup>4</sup> es  $\mu + \phi\mu^2$ , siendo  $\phi$  un parámetro de forma de la distribución que será mayor que 0 para la situación de sobredispersión<sup>4</sup>. De esta manera, la BN mantiene algunos aspectos interesantes en la modelización de variables que registran el número de veces que se produce un episodio, como la asimetría, pero es mucho más flexible en la forma de su distribución<sup>8</sup> y, por tanto, más adaptable a diferentes tipos de datos.

Con el único objetivo de ilustrar estas cuestiones, se presentan diversos modelos POI y BN formulados

como modelos lineales generalizados<sup>9</sup> y ajustados mediante el programa S-Plus 4.5<sup>10</sup>. Se calculó el riesgo de recurrencia de hospitalización durante todo el año 1996, en aquellas personas de edades comprendidas entre 65 y 69 años, dados de alta de su primera hospitalización, durante el citado año, en el mes de enero. Los individuos analizados pertenecían a una muestra de población de Cataluña seleccionada según criterios de factibilidad. El criterio de selección se basó en la unidad sanitaria territorial básica de procedencia, es decir, el área básica de salud (ABS). Así, se escogieron aquellos municipios a los que podía asignarse el ABS, es decir, se excluyeron aquellos para los cuales la ordenación territorial sanitaria asigna dos o más ABS.

Las variables explicativas ajustadas fueron: a) demográficas (sexo e índice de envejecimiento); b) estructurales (habitantes mayores de 14 años por médico y distancia del ABS al que pertenecía el individuo hasta el hospital de referencia), y c) socioeconómicas (tipo de localidad de residencia, rural o urbana, y número de motocicletas por cada mil habitantes). Las variables número de habitantes mayores de 14 años por médico e índice de envejecimiento fueron recodificadas a partir de la mediana del conjunto de ABS estudiadas sin ponderar por el número de individuos de cada una. Todas estas variables fueron obtenidas del proyecto FIS 97/1112<sup>11</sup>, excepto el número de motocicletas por mil habitantes que se obtuvo a través del Institut d'Estadística de Catalunya<sup>12</sup>. Esta variable fue escogida como indicador socioeconómico a pesar de su dudosa relevancia, para comprobar cómo a veces, por diferentes motivos (uno de ellos la subestimación de la variancia real), un modelo puede considerar importantes ciertos factores que tienen una más que discutible significación sustantiva.

Asimismo, para valorar la influencia de la estructura de la base de datos en los resultados obtenidos se utilizaron dos organizaciones distintas:

1. Agregada por individuo. Cada registro de la base representa a un individuo concreto en el cual se especifican sus características, así como el número de recurrencias de hospitalización y el tiempo total de seguimiento de éste.

2. Agregada según las variables de interés. En esta situación se dispone del número total de recurrencias de hospitalización y la suma de personas-días para cada una de las situaciones concretas definidas por las categorías de las variables o factores que interesa contrastar.

## Resultados

Durante el mes de enero de 1996, 965 individuos entre 65 y 69 años de edad recibieron el alta de su primera hospitalización, procedentes de las ABS estu-

diadas. Éstos generaron un total de 1.307 hospitalizaciones a lo largo de todo el año 1996.

En la tabla 1 puede observarse la distribución de las variables utilizadas en el análisis. Las categorías con mayor frecuencia de hospitalización fueron los varones, residentes en localidades rurales –con hasta 1.400 habitantes mayores de 14 años por médico, siendo el índice de envejecimiento de hasta 0,16–, pertenecientes a ABS situados a más de 5 km de distancia del hospital de referencia, y en localidades cuyo número de motocicletas por mil habitantes era inferior a 50.

### Base de datos por individuo

El análisis previo de los datos indica la posible presencia de sobredispersión. La distribución de recurrencias presenta una media de 0,45 y una variancia de 0,86, con lo cual el coeficiente de dispersión es claramente superior a 1. Este hecho parece confirmar la sospecha de la existencia de sobredispersión. En la tabla 2 puede observarse la distribución de las recurrencias de hospitalizaciones. El 72,2% de individuos únicamente fueron hospitalizados en una ocasión durante el período de estudio. Por tanto, no tuvieron ninguna recurrencia. En cambio, un 27,8% fueron hospitalizados más de una vez. De las 1.307 hospitalizaciones totales, 437 (33,4%) fueron segundas o posteriores. De éstas, 264 (60,4%) fueron generadas por un pequeño contingente (9,8%) del total de individuos. Entre las personas que fueron hospitalizadas en más de una ocasión, este porcentaje de recurrencias fue debido a las segundas y

Tabla 1. Distribución de la frecuencia de hospitalización según variables de estudio

Variable	n	%
<i>Sexo</i>		
Masculino	562	58,2
Femenino	403	41,8
<i>Tipo de localidad</i>		
Rural	680	70,5
Urbana	285	29,5
<i>Habitantes mayores de 14 años por médico</i>		
Hasta 1.400	595	61,7
Más de 1.400	370	38,3
<i>Índice de envejecimiento</i>		
Hasta 0,16	548	56,8
Más de 0,16	417	43,2
<i>Distancia del ABS al hospital</i>		
Hasta 5 km	350	36,3
Más de 5 km	615	63,7
<i>Motocicletas por mil habitantes</i>		
Menos de 50	493	51,5
50 o más	464	48,5

**Tabla 2. Distribución del número de hospitalizaciones y recurrencias de hospitalización**

N.º de hospitalizaciones	Individuos			Recurrencias		
	n	%	% acumulado	n	%	% acumulado
1	697	72,2	72,2	0	0,0	0,0
2	173	17,9	90,2	173	39,6	39,6
3	52	5,4	95,5	104	23,8	63,4
4	25	2,6	98,1	75	17,2	80,5
5	9	0,9	99,1	36	8,2	88,8
6	8	0,8	99,9	40	9,2	97,9
10	1	0,1	100,0	9	2,1	100,0

posteriores hospitalizaciones en el 35,4% de los individuos.

En los modelos ajustados a la base de datos agregada por individuo (tabla 3) puede observarse la distinta significación de las variables según el modelo. En el modelo POI todas las variables resultan significativas, exceptuando el índice de envejecimiento y los kilómetros hasta el hospital de referencia, las cuales se mantuvieron en el modelo debido a que estaban al borde de la significación. Para el modelo BN las únicas variables sig-

nificativas fueron el sexo y el tipo de localidad de residencia. Las tasas estimadas por ambos modelos son muy parecidas, aunque los intervalos de confianza de los riesgos relativos se incrementan en el modelo BN, debido a que los errores estándares de los coeficientes estimados, mediante esta distribución, aumentaron entre un 48 y un 53% respecto a los estimados por la POI. El parámetro que indexa la sobredispersión (tabla 3) es significativo, lo que confirma definitivamente la sospecha de la presencia de sobredispersión.

En la figura 1 puede observarse el estudio de los residuos para los modelos ajustados. Cada punto es la representación del residuo *deviance*<sup>9,13</sup> de la misma observación según el ajuste de POI (eje de abscisas) o BN (eje de coordenadas). Cuando el residuo *deviance* es menor de -2 o mayor de 2, se puede considerar que el valor observado es muy distinto al esperado según el modelo ajustado. Por tanto, el gráfico se divide en 9 rectángulos: las observaciones que se sitúan en el rectángulo número 5 son aquellas ajustadas correctamente tanto por la POI como por la BN (residuo entre -2 y 2 en ambos ajustes). Aquellas colocadas en los rectángulos 2 y 8 son bien ajustadas por la POI y no por la BN, mientras que en 4 y 6 estarían las que POI no ajusta correctamente mientras sí lo hace la BN. Finalmente, en los rectángu-

**Tabla 3. Estimación de los coeficientes, errores estándares (EE) y riesgos relativos (RR) para los modelos de regresión de Poisson y binomial negativa (base de datos agregada por individuo)**

Coeficiente	Distribución de Poisson				Distribución binomial negativa <sup>a</sup>			
	Valor	EE	RR	IC del 95%	Valor	EE	RR	IC del 95%
(Constante)	-6,606	0,119	1		-6,521	0,176	1	
Mujer	-0,405	0,100	0,67	(0,55-0,81)	-0,526	0,147	0,59	(0,44-0,79)
Localidad urbana	0,280	0,113	1,32	(1,06-1,65)	0,401	0,173	1,49	(1,06-2,09)
> 1.408 habitantes mayores de 14 años por médico	0,221	0,107	1,25	(1,01-1,54)	0,260	0,163	1,30	(0,94-1,79)
Índice de envejecimiento > 0,16	-0,208	0,116	0,81	(0,65-1,02)	-0,214	0,172	0,81	(0,58-1,13)
Más de 5 km al hospital de referencia	0,189	0,106	1,21	(0,98-1,49)	0,162	0,158	1,18	(0,86-1,60)
> 50 motocicletas por 1.000 habitantes	0,234	0,099	1,26	(1,04-1,53)	0,257	0,148	1,29	(0,97-1,73)

<sup>a</sup>Parámetro de sobredispersión:  $\phi = 0,38$ ; IC del 95% = 0,27-0,48.

**Tabla 4. Estimación de los coeficientes, errores estándares (EE) y riesgos relativos (RR) para los modelos de regresión de Poisson y binomial negativa (base de datos agregada según las variables de interés)**

Coeficiente	Distribución de Poisson				Distribución binomial negativa <sup>a</sup>			
	Valor	EE	RR	IC del 95%	Valor	EE	RR	IC del 95%
(Constante)	-6,795	0,132			-6,827	0,209		
Mujer	-0,405	0,102	0,67	(0,55-0,82)	-0,378	0,164	0,69	(0,50-0,94)
Localidad urbana	0,279	0,115	1,32	(1,05-1,66)	0,250	0,181	1,28	(0,90-1,83)
> 1.408 habitantes mayores de 14 años por médico	0,221	0,109	1,25	(1,01-1,55)	0,287	0,173	1,33	(0,95-1,87)
Índice de envejecimiento > 0,16	-0,209	0,119	0,81	(0,64-1,02)	-0,167	0,180	0,85	(0,59-1,20)
Más de 5 km al hospital de referencia	0,189	0,108	1,21	(0,98-1,49)	0,115	0,166	1,12	(0,81-1,55)
> 50 motocicletas por 1.000 habitantes	0,234	0,101	1,26	(1,04-1,54)	0,229	0,167	1,26	(0,91-1,74)

<sup>a</sup>Parámetro de sobredispersión:  $\phi = 6,66$ ; IC del 95% = 0,70-12,62.

Figura 1. Residuos de *deviance* de los modelos Poisson y Binomial Negativa para la base de datos agregada por individuo.

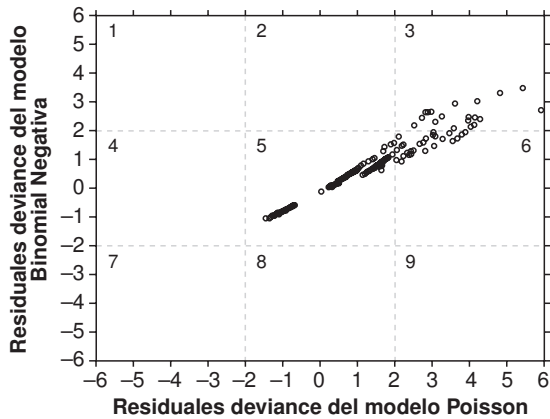
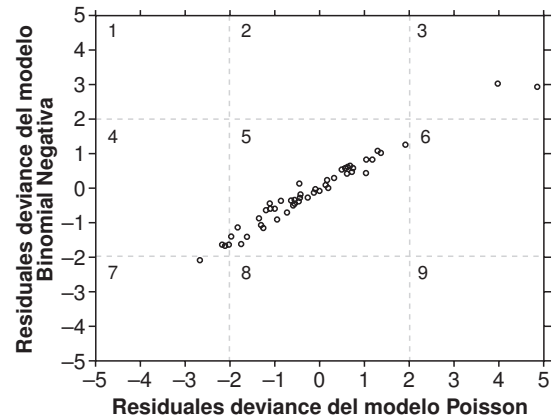


Figura 2. Residuos de *deviance* de los modelos Poisson y Binomial Negativa para la base de datos agregada según las variables de interés.



los de los extremos (1, 3, 7 y 9) se ubicarían aquellas observaciones mal ajustadas por ambas distribuciones.

Se observa el mejor ajuste del modelo BN respecto al POI; 56 de los residuos (un 5,9%) fueron mal ajustados mediante la distribución POI; porcentaje que disminuyó hasta el 1,9% al estimar los valores esperados mediante la BN, lo que significa que ésta consigue estimar correctamente el 67,9% del total de observaciones cuyos residuos superan los límites aceptables bajo la distribución POI. En este caso, éstos son los ubicados en el rectángulo número 6.

*Base de datos agregada según las variables de interés*

En este caso la media es de 8,68 y la variancia de 103,61, siendo la sospecha de sobredispersión muy elevada. Para esta base, puede observarse (tabla 4) que al ajustar mediante BN también desaparece la significación del número de habitantes por médico y el número de motocicletas por cada mil habitantes. En este caso, tampoco el tipo de localidad tiene un efecto relevante. Igual que en la base agregada por individuo, los errores estándares de los coeficientes se incrementan al estimar mediante la BN. La estimación de  $\phi$  (tabla 4) también resulta significativa, aunque mucho más imprecisa debido a que, tal y como está estructurada la base, no se puede conocer cuáles de las hospitalizaciones son recurrentes.

En la figura 2 se incluye el estudio de los residuos para la base de datos agregada según las variables de interés. El gráfico se interpreta de la misma forma que la base de datos individualizada. Así, 6 de los residuos (12%) se ajustaron mal por la distribución POI, reduciéndose hasta 3 (6%) al utilizar la BN.

**Discusión**

Clásicamente, el cálculo de tasas de incidencia se ha analizado mediante la POI, aun teniendo en cuenta las limitaciones que ésta presenta en determinadas situaciones. Cuando el fenómeno de interés es de tipo recurrente, la posibilidad de incumplimiento de los criterios de aplicabilidad de la POI es elevado. En esta situación puede producirse el fenómeno de la sobredispersión o variancia extra-Poisson. Por ello, aunque la estimación de las tasas será robusta si el modelo se ha especificado correctamente, la POI es incapaz de estimar parte de la variancia, estando los intervalos de confianza infraestimados. Este hecho puede provocar que algunos de los coeficientes sean considerados significativos cuando realmente no lo son; por tanto, pueden derivar en la aplicación de intervenciones sobre factores que verdaderamente no sean influyentes.

A menor nivel de agregación de la base de datos, la mejora de estimación de la BN frente a la POI se acrecenta, ya que el fenómeno de la sobredispersión puede ser estimado con más precisión. Así, la estimación de  $\phi$  varía sustancialmente dependiendo de la base de datos empleada. Esto sucede porque el cálculo de  $\phi$  depende de la media observada. Además, las estimaciones que realiza la POI no dependen de la unidad de análisis utilizada. En ambas bases, la POI realiza casi exactamente las mismas estimaciones. Las suposiciones de independencia de las observaciones y la aleatoriedad de éstas entre los individuos implica que la base de datos individualizada puede agregarse por las variables de interés, pues la POI no distinguirá entre la información de cada individuo y la agregada.

De todos modos, especialmente si el período de seguimiento de la cohorte es elevado, deben destacarse algunos posibles inconvenientes que la BN tampoco es capaz de resolver:

1. La imposibilidad de controlar las variables que fluctúan en el tiempo. Esto es, deberemos escoger entre registrar el individuo, suponiendo que sus variables explicativas no cambian, o registrar el cambio de éstas, con lo que el individuo puede aportar más de un registro en la base, pero sin poder identificar que estos registros pertenecen al mismo individuo.

2. La imposibilidad de controlar posibles intervalos de tiempo donde el riesgo es distinto. Al realizar todas las estimaciones sobre una base de datos donde los episodios y el tiempo de seguimiento se acumulan, el riesgo se supone constante a lo largo del período. Pero este tipo de suposición puede no ser así debido a diferentes causas. Algunas posibles correcciones, como dividir el período de seguimiento en otros más pequeños, son de difícil justificación en el caso de que no exis-

ta un criterio claro, y se pueden crear artificios que afecten a los resultados.

Además, la existencia de un elevado contingente de individuos que no presenten el episodio puede implicar correcciones en la BN<sup>5</sup>.

Diversos autores<sup>14-18</sup> han propuesto otras interesantes soluciones para el análisis de fenómenos recurrentes, aunque quizás de interpretación no tan sencilla. Para bases de datos más desagregadas, en las que el registro es el episodio y no el individuo o el grupo, han sido presentadas técnicas estadísticas mucho más depuradas<sup>15,18,19</sup>. El gran inconveniente es la accesibilidad a este tipo de bases de datos.

Por tanto, en un nivel de agregación como el de los datos expuestos, así como por su sencillez de aplicación e interpretación, la BN es adecuada para el análisis de fenómenos recurrentes ya que permite estimar, al menos en parte, la variancia que la POI es incapaz de identificar.

---

## Bibliografía

1. Cumming RG, Kelsey JL, Nevitt MC. Methodologic issues in the study of frequent and recurrent health problems: falls in the elderly. *Ann Epidemiol* 1990; 1: 49-56.
2. Lindsey JK. Counts and times to events. *Statist Med* 1998; 17: 1745-1751.
3. Lindsey JK. Models for repeated measurements. Oxford: Oxford University Press, 1994.
4. Lawless JF. Negative binomial and mixed Poisson regression. *Can J Stat* 1987; 15: 209-225.
5. Miaou SP. The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. *Accid Anal Prev* 1994; 26: 471-482.
6. Venables WN, Ripley BD. Modern applied statistics with S-Plus (2.ª ed.). Nueva York: Springer-Verlag, 1997.
7. Glynn RJ, Stukel TA, Sharp SM, Bubolz TA, Freeman JL, Fisher ES. Estimating the variance of standardized rates of recurrent events, with application to hospitalizations among the elderly in New England. *Am J Epidemiol* 1993; 7: 776-786.
8. Lindsey JK. Introductory statistics. A modelling approach. Oxford: Oxford University Press, 1995.
9. McCullagh P, Nelder JA. Generalized linear models (2.ª ed.). Londres/Nueva York: Chapman and Hill, 1989.
10. S-PLUS 4.5 Professional edition for Windows. Seattle: Data Analysis Products Division, MathSoft, 1998.
11. Caminal J. Las hospitalizaciones por Ambulatory Care Sensitive Conditions: Un indicador de la capacidad de resolución de la atención primaria de salud [Tesis doctoral]. Barcelona: Universitat Autònoma de Barcelona; 1999.
12. Web del Institut d'Estadística de Catalunya. Disponible en: <http://www.idescat.es/>.
13. Pierce DA, Schafer DW. Residuals in Generalized Linear Models. *J Am Stat Assoc* 1986; 81: 977-986.
14. Breslow NE. Extra-Poisson variation in log-linear models. *Appl Statist* 1984; 33: 38-44.
15. Clayton D. Some approaches to the analysis of recurrent event data. *Stat Methods Med Res* 1994; 3: 244-262.
16. Diggle PJ, Liang KY, Zeger SL. Analysis of longitudinal data. Oxford: Oxford University Press, 1994.
17. Benavides FG, Sáez M, Barceló MA, Serra C. Incapacidad temporal: estrategias de análisis. *Gac Sanit* 1999; 13: 185-190.
18. S-Plus 4 Guide to Statistics. Seattle: Data Analysis Products Division, MathSoft, 1998.
19. Kelly P, Lim L. Survival analysis for recurrent event data: an application to childhood infectious diseases. *Statist Med* 2000; 19: 13-33.