Aplicación de los modelos de regresión tobit en la modelización de variables epidemiológicas censuradas

M.J. Bleda Hernández[a,b] / **A. Tobías Garcés**[b]
[a]Unidad de Investigación en Tuberculosis. Centro Nacional de Epidemiología. Subdirección General de Epidemiología e Información Sanitaria. Instituto de Salud Carlos III. Madrid.
[b]Departamento de Estadística y Econometría. Universidad Carlos III de Madrid. Getafe.

Correspondencia: María José Bleda Hernández. Unidad de Investigación en Tuberculosis. Centro Nacional de Epidemiología. Instituto de Salud Carlos III. C/ Sinesio Delgado, 6. 28028 Madrid.
Correo electrónico: mjbleda@isciii.es

Recibido: 23 de mayo de 2001.
Aceptado: 7 de enero de 2002.

(Application of tobit regression models in modelling censored epidemiological variables)

Resumen

Muchas variables en estudios epidemiológicos corresponden a medidas continuas obtenidas mediante aparatos de medición con determinados límites de detección, produciendo distribuciones censuradas. La censura, a diferencia del truncamiento, se produce por un defecto de los datos de la muestra. La distribución de una variable censurada es una mezcla entre una distribución continua y otra discreta. En este caso, no es adecuado utilizar el modelo de regresión lineal estimado para mínimos cuadrados ordinarios, ya que proporciona estimaciones sesgadas. Con un único punto de censura debe utilizarse el modelo de regresión censurado (modelo tobit), mientras que cuando hay varios puntos de censura se utiliza la generalización de este modelo. La ilustración de estos modelos se presenta a través del análisis de las concentraciones de mercurio medidas en orina, correspondientes al estudio sobre los efectos para la salud de las emisiones de la incineradora de residuos sólidos de Mataró.
Palabra clave: Regresión. Truncamiento. Censura. Máxima verosimilitud. Modelo tobit.

Summary

Many variables in epidemiological studies are continuous measures obtained by means of measurement equipments with detection limits, generating censored distributions. The censorship, opposite to the trucation, takes place for a defect of the data of the sample. The distribution of a censored variable is a mixture between a continuous and a categorical distributions. In this case, results from lineal regression models, by means of ordinary least squares, will provide biased estimates. With one only censorhip point the tobit model must be used, while with several censorship points this model's generalization should also be used. The illustration of these models is presented through the analysis of the levels of mercury measured in urine in the study about health effects of a municipal solid-waste incinerator in the county of Mataró (Spain).
Key words: Regression. Truncation. Censorship. Maximum likelihood. Tobit model.

Introducción

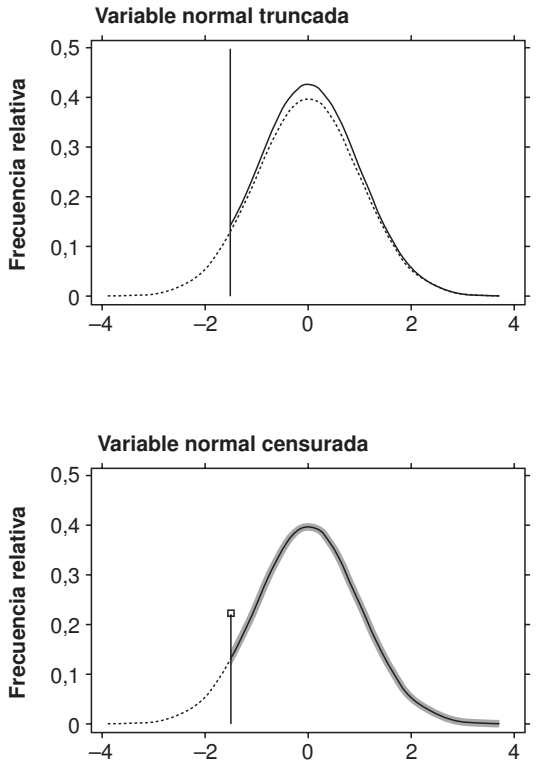
En muchas ocasiones, variables objeto de investigación en estudios epidemiológicos se corresponden a medidas continuas obtenidas mediante aparatos de medición debidamente ajustados y calibrados. Es habitual que dichos aparatos tengan determinados límites de detección, tanto inferiores como superiores. Estos límites pueden hacer que, a pesar de que la variable que nos interesa estudiar tenga una distribución determinada, los valores que realmente se observen en la muestra no sean representativos. Algunos ejemplos los podemos encontrar en la distribución de los niveles de inmunoglobulina E en sangre[1] o los niveles de metales medidos en sangre u orina[2].

El truncamiento es una característica intrínseca de la distribución de la variable objeto de estudio, de la cual se extraen los datos de la muestra. Se produce cuando sólo la parte de la distribución de la variable que se encuentra por encima (o por debajo) del denominado punto de truncamiento contiene la información relevante que se desea estudiar. Un ejemplo de variable truncada sería el valor de hemoglobina cuando el interés reside en estudiar a aquellos pacientes con valores inferiores a 8 g/dl en la población. El punto de truncamiento es 8 g/dl y la variable se dice que está truncada. A nivel teórico, para que la función de densidad de una varia-

ble aleatoria truncada integre la unidad, se divide su función de densidad entre la probabilidad de que una observación no pertenezca al área truncada. En la figura 1 se representa gráficamente cómo afecta el truncamiento a la función de densidad de una distribución normal estándar, con punto de truncamiento inferior *a* = –1,5. Otros ejemplos de variables truncadas se pueden encontrar en los modelos usados en el análisis del gasto sanitario[3].

La censura, por el contrario, no es una característica intrínseca de la distribución de la variable objeto de estudio, sino un defecto de los datos de la muestra, que si no estuvieran censurados constituirían una muestra representativa de la población de interés no censurada. Un ejemplo habitual de censura es el que se produce cuando la variable objeto de estudio es el tiempo de supervivencia desde el diagnóstico de una enfermedad hasta la fecha de muerte (evento). En la práctica el estudio tendrá definida una fecha de finalización (punto de censura) en la que ocurrirá que no todos los sujetos de la muestra escogida habrán muerto (algunos seguirán vivos). A pesar de que el objetivo sería estudiar el tiempo de supervivencia en la población de enfermos diagnosticados de dicha enfermedad, no es posible disponer en la muestra de los tiempos de supervivencia de todos los enfermos. La variable tiempo de supervivencia se dice entonces que está censurada superiormente. Cuando la variable está censurada, la distribución que siguen los datos de la muestra es una mezcla (mixtura) entre una distribución continua y otra discreta, existiendo una acumulación de probabilidad en el punto de censura. También en la figura 1, se presenta la función de densidad de una distribución normal estándar, censurada, con un único punto de censura inferior a *a* = –1,5.

Figura 1. Distribución normal estándar N(0,1), con un punto de truncamiento inferior y con un único punto de censura inferior.



La línea discontinua muestra la función de densidad de una distribución N(0,1) y en continuo la función de densidad de una N(0,1) truncada inferiormente en el punto *a* = –1,5, donde el área (probabilidad) de la cola de la N(0,1) que queda a la izquierda del punto de truncamiento se reparte entre el conjunto de puntos no truncados, haciendo que la función de densidad de la N(0,1) truncada integre la unidad. La línea discontinua representa la función de densidad de una distribución N(0,1) y en negrita la función de densidad de una N(0,1) censurada inferiormente en *a* = –1,5, donde el área (probabilidad) de la cola inferior que queda a la izquierda del punto de censura se acumula en dicho punto de censura *a*. Así, la altura de la línea vertical en el punto de censura representa el valor de esta área inferior. Por tanto, la función de densidad de una variable censurada es una mixtura entre una variable discreta, por la acumulación de probabilidad en el punto de censura *a*, y una variable continua ya que los valores no censurados siguen una N(0,1).

Si la variable objeto de estudio es una medición continua que se distribuye según una ley normal, en la que existen uno o varios puntos de truncamiento y/o censura, no es posible utilizar los habituales modelos de regresión lineal estimados por mínimos cuadrados ordinarios (MCO), porque proporcionan estimaciones incorrectas del efecto y de su variabilidad[4,5]. Cuando la variable de interés tiene un punto de truncamiento se debe utilizar el denominado modelo de regresión truncado[4,5]. Análogamente si tiene un único punto de censura tiene que utilizarse el llamado modelo de regresión censurado o modelo tobit[6]. Cuando existen varios puntos de truncamiento o censura, o cuando coexisten al mismo tiempo censura y truncamiento, se utilizan las respectivas generalizaciones de estos modelos que, desarrollados originalmente en el campo de la econometría, se han aplicado con frecuencia en el campo de la economía de la salud[7-10].

La necesidad de utilizar modelos alternativos a los modelos de regresión lineal, estimados por MCO, surgió en el estudio sobre los efectos potenciales para la salud de las emisiones de una incineradora de residuos sólidos urbanos en la población de Mataró (Barcelona)[2,11], al analizar los datos correspondientes a valores de metales medidos en orina. Las concentraciones de mercurio presentaban la particularidad de tener varios puntos de censura inferior en la cola izquierda de la distribución, debidos al límite de detección inferior del aparato de medición. El objetivo de este trabajo es describir las potenciales aplicaciones de la famila de modelos de regresión censurada en la modelización de variables epidemiológicas censuradas.

Material y métodos

Diseño del estudio

En el estudio sobre los efectos potenciales para la salud de las emisiones de una incineradora de residuos sólidos urbanos en la población de Mataró (Barcelona), se seleccionó una muestra de 201 sujetos voluntarios (100 varones y 101 mujeres) de edades comprendidas entre los 18 y los 68 años del padrón municipal durante el período marzo-junio de 1995[2,11]. Para todos los sujetos se recogió información sobre tabaquismo, dieta y alcohol, así como muestras de sangre y orina. Se midieron los valores de mercurio en orina. Los límites de detección inferiores para los métodos analíticos utilizados, definidos como la concentración dado un 1% de absorción, fueron de 0,2 µg/l. Las concentraciones de mercurio se corrigieron posteriormente en función de la concentración de creatinina en la orina, por lo que las unidades en que finalmente se expresaron fueron µg/g de creatinina (µg/g CR).

Modelo de regresión censurado con un único punto de censura (modelo tobit)

El modelo tobit fue propuesto por Tobin[6] en 1958 y es en su honor por lo que se denomina de este modo. Para definir la distribución de la variable censurada, que se denominará *y,* con un único punto de censura inferior *a*, es necesaria la utilización de la variable aleatoria original subyacente (latente) y^*. Entonces, la variable censurada *y* tomará los valores:

$y = a_y$ cuando la variable subyacente $y^* \leq a$
$y = y^*$ cuando la variable subyacente $y^* > a$

Cabe notar la diferencia entre los valores a_y y *a*. El punto de censura *a* determina si y^* está censurada, mientras que a_y es el valor asignado a la variable *y* si y^* está censurada. Usualmente el valor a_y es igual al valor del punto de censura *a*, pero podría no serlo. Por simplicidad se supondrán iguales de aquí en adelante[4,5].

Si además se realiza la asunción de que la distribución de la variable subyacente es $y^* \sim N(\mu,\sigma^2)$ la probabilidad de que una observación esté censurada o no lo esté será:

$$\Pr(\text{censurada}) = \Pr(y^* \leq a) = \Pr(N(\mu,\sigma^2) \leq a) =$$

$$\Pr\left(N(0,1) \leq \left(\frac{a-\mu}{\sigma}\right)\right) = \Phi\left(\frac{a-\mu}{\sigma}\right)$$

$$\Pr(\text{no censurada}) = \Pr(y^* > a) =$$

$$= 1-\Pr(y^* \leq a) = 1-\Phi\left(\frac{a-\mu}{\sigma}\right) = \Phi\left(\frac{\mu-a}{\sigma}\right)$$

donde Φ (.) representa la función de distribución de *a* N(0,1) evaluada en el punto en cuestión.

La función de densidad de la variable censurada será entonces:

$$\Pr(y = a) = \Pr(y^* \leq a) = \Phi\left(\frac{a-\mu}{\sigma}\right) \text{ cuando } y^* \leq a$$

La misma densidad de y^* cuando $y^* > a$

Esta distribución es una mixtura entre una distribución continua y otra discreta, donde se asigna toda la probabilidad contenida en el área censurada al punto de censura *a*. Por esta razón, se habla de un punto de acumulación de probabilidad en el punto de censura (fig. 1).

El interés en un modelo tobit reside habitualmente en estudiar la variable latente y^*. La formulación general del modelo es que el valor medio de esta variable y^* es una función lineal de las variables explicativas $E[y_i^* \mid x_i] = X'_i\beta$. Dado que los valores de y^* son desconocidos, y tan sólo se conocen los valores de la variable censurada *y*, se modelizará la $E[y_i \mid x_i]$ expresándola en función de $E[y_i^* \mid x_i]$ como:

$$E[y_i \mid x_i] = E[y_i^* \mid x_i, y_i^* > a] \cdot \Pr[y_i^* > a \mid x_i] + a^* \Pr[y_i^* \leq a \mid x_i]$$

La estimación de este modelo utilizando el método de MCO proporciona estimaciones sesgadas de los coeficientes. Sin embargo, las estimaciones por el método de máxima verosimilitud facilitan estimaciones de los coeficientes eficientes y consistentes[4,5], ya que la función de verosimilitud que se maximiza integra información tanto de las observaciones censuradas como de las no censuradas:

$$I(\beta,\sigma^2) = \ln L(\beta,\sigma^2) = \sum_{y_i > a} -\frac{1}{2}$$

$$\left[\ln(2\pi) + \ln(\sigma^2) + \frac{(y_i - x_i'\beta)^2}{\sigma^2}\right] + \sum_{y_i \leq a} \ln\left[\Phi\left(\frac{a - x_i'\beta}{\sigma}\right)\right]$$

En esta función se observa cómo se podrán identificar las estimaciones de los efectos sobre la variable latente y^* ($\hat{\beta}$) utilizando únicamente la variable censurada *y*.

Hay que señalar que en este modelo la no normalidad afecta en mayor medida que en los modelos de regresión lineal habituales y produce que los estimadores $\hat{\beta}$ sean inconsistentes. En la actualidad muchos investigadores están estudiando cómo contrastar la hipótesis de normalidad del modelo[3,4]. Los fundamentos teóricos presentados en el modelo tobit son generalizables a situaciones en las que la variable dependiente pueda tener varios puntos de censura, ya sean todos inferiores, todos superiores o inferiores y superiores[4,5].

Interpretación de los coeficientes

El interés en un modelo tobit puede centrarse en la estimación de diferentes medidas de efecto:

1. Cuando el interés reside en el estudio de las variables *x* asociadas con la variable latente *y**, las estimaciones $\hat{\beta}$ obtenidas en el modelo tobit representan directamente el efecto marginal que cada una de las variables *x* tiene en el valor medio de *y**.

2. Sin embargo, si el interés reside en el estudio de las variables *x* asociadas con la variable censurada *y*, las estimaciones $\hat{\beta}$ obtenidas en el modelo tobit deberán ponderarse por la probabilidad de que una observación no esté censurada:

$$\hat{\beta} \cdot \Phi\left(\frac{x'_i\ \hat{\beta} - a}{\hat{\sigma}}\right)$$

Esta probabilidad de no censura depende de los valores que tome cada uno de los sujetos *i* en cada una de las variables *x*, por lo que habitualmente se evalúa en la media, mínimo y/o máximo de dichas variables.

Aunque este último interés no suele darse en el ámbito de la epidemiología, es frecuente en el campo de la economía de la salud.

Análisis estadístico

Para contrastar si existían diferencias estadísticamente significativas entre los sujetos con censura y sin censura, se utilizó el test de la suma de rangos de Wilcoxon para las variables continuas, que pone a prueba si los datos de ambos grupos de sujetos proceden de poblaciones con la misma distribución. Para las variables categóricas se utilizó el estadístico de contraste de la χ^2 de Pearson, el cual pone a prueba si las filas y las columnas en una tabla de contingencia son independientes.

Seguidamente, para cada uno de los tres modelos de regresión analizados se realizaron los respectivos modelos de regresión univariantes para cada una de las variables explicativas x_i consideradas. En el primer modelo de regresión analizado, se consideraron tan sólo aquellos individuos con valores detectados (la muestra con valores observados) y se estimó un modelo de regresión lineal por MCO. En el segundo modelo, se consideraron todos los individuos, aunque se asumió que todos los sujetos con valores censurados tomaban el mismo valor mínimo de censura (a = 0,1 µg/g CR). Se escogió este valor mínimo porque se consideró que era situarse en el peor caso que se podría haber dado.

Se estimó un modelo de regresión lineal censurado con un único punto de censura o modelo tobit. Por último, en el tercer modelo se consideraron de nuevo todos los individuos, aunque los individuos con valores censurados tomaron sus respectivos valores de censura. Se estimó un modelo de regresión lineal censurado con varios puntos de censura, que es la generalización del modelo tobit anterior.

Para cada uno de los tres análisis se construyeron a continuación los modelos multivariados. Se incluyeron todas aquellas variables cuyo valor de la t de Student para el coeficiente estimado resultó en valor absoluto mayor que 1 en los correspondientes modelos univariados y, posteriormente, se fueron eliminando una a una las variables no significativas[12] hasta configurar los modelos finales.

Software estadístico

El análisis estadístico se ha realizado utilizando el paquete estadístico Stata, versión 6.0[13]. Las instrucciones utilizadas para estimar los diferentes modelos han sido: *regress* para estimar el modelo de regresión lineal múltiple por MCO, *tobit* estima el modelo de regresión censurado con un único punto de censura y *cnreg* estima el modelo de regresión lineal censurado con varios puntos de censura.

Resultados

En tres de los 201 sujetos (1,5%) estudiados no se obtuvo la muestra de orina necesaria para realizar la medición. En 63 de los 198 sujetos (31,8%) no se detectó la concentración de mercurio debido al límite de detección inferior del aparato de medición (tabla 1). Para estos sujetos, el valor de censura se correspondió al límite inferior de detección, que varió de unos sujetos a otros en función de la concentración de creatinina en la orina (tabla 2). Además, para normalizar los valores de mercurio, éstos fueron transformados logarítmicamente debido a la forma asimétrica de la distribución (fig. 2).

En el análisis descriptivo para los sujetos censurados y no censurados (tabla 3), la comparación de los

Tabla 1. Descripción de las concentraciones de mercurio (en µg/g creatinina) para los individuos con valores censurados y no censurados

		Percentiles						
Muestra	n (%)	Mínimo	P5	P25	Mediana	P75	P95	Máximo
No censurados	135 (68,2)	0,1	0,3	1,3	2,3	4,8	12,8	21,0
Censurados	63 (31,8)	0,1	0,1	0,2	0,3	0,5	0,9	1,2

Tabla 2. Distribución de los puntos de censura en las concentraciones de mercurio (en μg/g creatinina)

Concentraciones de mercurio	n (%)
0,1	15 (23,8)
0,2	16 (25,4)
0,3	10 (15,9)
0,4	5 (7,9)
0,5	8 (12,7)
0,6	2 (3,2)
0,7	3 (4,7)
0,9	2 (3,2)
1,2	2 (3,2)
Total	63

valores de las variables incluidas en el análisis no objetivó diferencias estadísticamente significativas a un nivel de significación $\alpha = 0,05$.

En los modelos de regresión, tanto univariantes como multivariantes, muy pocas variables demostraron estar asociadas con los valores de mercurio (tabla 4). Así, las variables que finalmente se incluyeron en los tres modelos fueron la edad (en años), sexo (0 = varón; 1 = mujer), consumo de verduras crudas y consumo de ahumados (ambos medidos en número de raciones semanales). En el modelo de regresión lineal, estimado por MCO, las variables edad y sexo resultaron estadísticamente significativas (p = 0,008 y p = 0,004, respectivamente) mientras que el consumo de verduras crudas resultó marginalmente significativo (p = 0,077). En el modelo tobit, considerando un mismo punto de censura en 0,1 μg/g CR para los 63 sujetos con valores censurados, la variable edad dejó de ser estadísticamente significativa (p = 0,968), mientras que la variable sexo continuó siéndolo (p = 0,045). Sin embargo, el consumo de verduras crudas resultó significativo (p = 0,022) y el de ahumados se mostró al límite de la significación (p = 0,099). Por último, en el modelo censurado con varios puntos de censura (tabla 2), la variable edad de nuevo dejó de ser estadísticamente significativa (p = 0,523), en tanto que la variable sexo continuó siéndolo (p = 0,016). Tanto el consumo de verdudras crudas como el de ahumados resultaron al límite de la significación estadística (p = 0,079 y p = 0,101, respectivamente).

Figura 2. Distribución de las concentraciones de mercurio (en escalas original y logarítmica) para diferentes puntos de censura y para un único punto de censura.

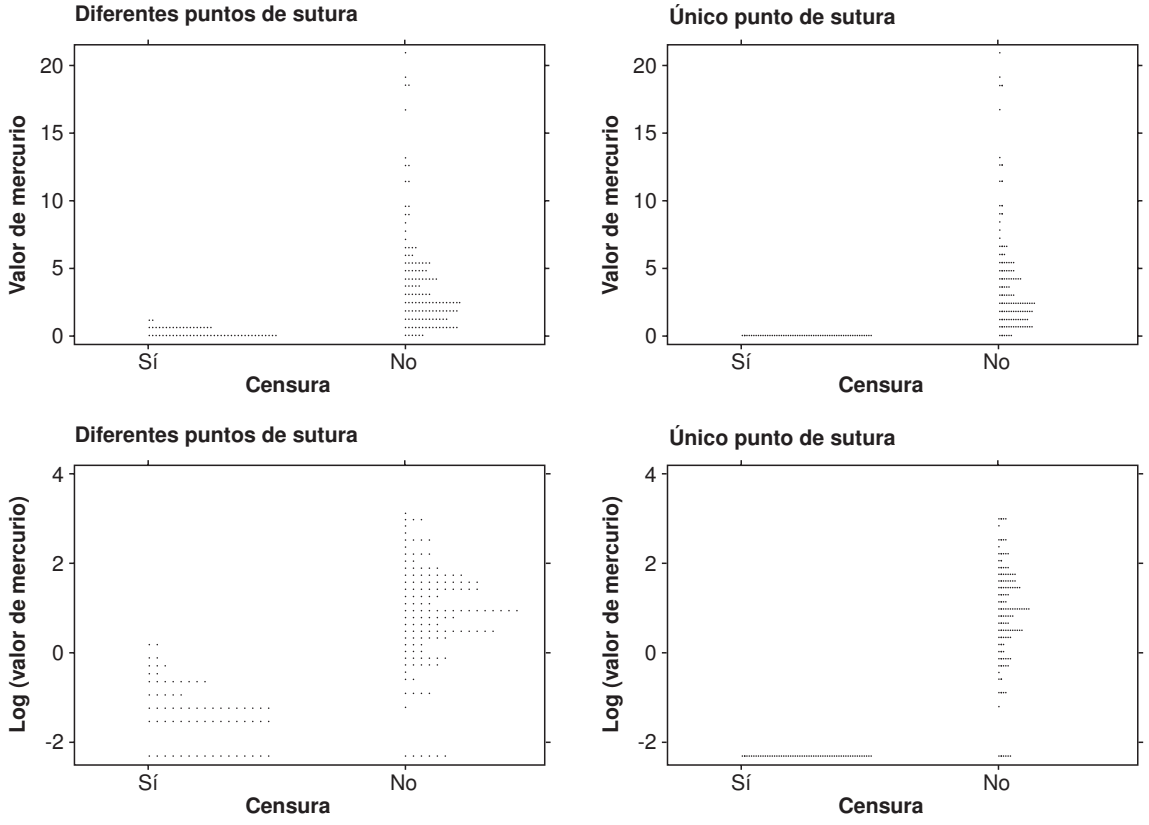


Tabla 3. Análisis descriptivo para la muestra censurada y no censurada[a]

	No censurados (n = 135)	Censurados (n = 63)
Características personales, $\bar{X}$ (DE)		
Edad	39,2 (13,7)	42,1 (4,3)
IMC	25,0 (4,1)	25,4 (4,4)
Sexo, n (%)		
Varón	65 (48,2)	34 (54,0)
Mujer	70 (51,8)	29 (46,0)
Fumador[b]		
Nunca	65 (48,9)	23 (37,1)
Ex	22 (16,5)	12 (19,4)
Sí	46 (34,6)	27 (43,5)
Raciones semanales mediana (P5-P95)		
Leche	7,0 (0-14)	7,0 (0-19,6)
Yogur	3,0 (0-7)	2,0 (0-7)
Queso	2,0 (0-7)	2,0 (0-7)
Verduras		
Crudas	3,0 (0,2-7)	5,0 (0,3-7)
Cocidas	3,0 (0,2-7)	3,0 (0,5-7)
Carnes rojas	3,0 (0,9-7)	3,0 (1-7)
Pescado	2,0 (0,5-7)	2,0 (0-7)
Marisco	0,2 (0-1,2)	0,2 (0-1)
Embutido	3,0 (0-7)	3,0 (0-7)
Ahumados	0,0 (0-0,5)	0,0 (0-0,2)
Café	7,0 (0-28)	7,0 (0-44,8)
Coca-cola	0,2 (0-14)	0,0 (0-7)
Alcohol	2,3 (0-31,5)	2,3 (0-34,7)

[a]No se encontraron diferencias estadísticamente significativas en las variables descritas en la tabla entre no censurados y censurados, a un nivel de significación α = 0,05; [b]dos sujetos con valores no censurados en el mercurio y un individuo con valor censurado no respondieron a la variable fumador. $\bar{X}$ (DE): media (desviación estándar); IMC: índice de masa corporal, (P5-P95): (percentil 5 - percentil 95).

Como el interés residía en investigar los valores de mercurio en la población de Mataró, de la cual se extrajo una muestra representativa, cada estimación $\hat{\beta}$ asociada a una variable *x*, obtenida en el modelo tobit y en el modelo censurado con varios puntos de censura, representa directamente el efecto marginal que cada una de las variables independientes tiene en el valor medio de la variable subyacente *y** cuando varían en una unidad, manteniendo constantes el resto de variables. Si el interés hubiese residido en la variable censurada *y*, la interpretación de los coeficientes en estos dos modelos no hubiese sido directa y se habría tenido que calcular el efecto marginal de las variables incluidas en el modelo, corrigiendo por la probabilidad de no censura.

Comparando los resultados obtenidos en el primer modelo (regresión lineal) con las estimaciones $\hat{\beta}$ obtenidas a través de los modelos tobit, y censurado con varios puntos de censura, se observa cómo las estimaciones de los tres tipos de modelos van en la misma dirección (tienen el mismo signo), aunque difieren bastante en su magnitud. Las estimaciones obtenidas en el modelo tobit y en el modelo censurado con varios puntos de censura son, en general, sustancialmente mayores (en valor absoluto).

Tabla 4. Resultados de los modelos de regresión lineal, tobit considerando un único punto de censura y con varios puntos de censura

	$\hat{\beta}$	(EE)	t	p
Regresión lineal (MCO) con la muestra no censurada (n = 135)[a]				
Edad	0,019	(0,007)	2,67	0,008
Sexo	0,566	(0,186)	2,93	0,004
Verduras crudas	–0,069	(0,039)	–1,78	0,077
Ahumados	0,492	(0,477)	1,03	0,304
Modelo tobit con un único punto de censura (n = 198)[b]				
Edad	0,0005	(0,013)	0,04	0,968
Sexo	0,699	(0,346)	2,02	0,045
Verduras crudas	–0,152	(0,066)	–2,31	0,022
Ahumados	1,631	(0,984)	1,66	0,099
Modelo con varios puntos de censura (n = 198)[c]				
Edad	0,007	(0,012)	0,64	0,523
Sexo	0,687	(0,282)	2,44	0,016
Verduras crudas	–0,096	(0,054)	–1,77	0,079
Ahumados	1,314	(0,797)	1,65	0,101

$\hat{\beta}$: coeficiente de regresión; EE: error estándar del coeficiente de regresión $\hat{\beta}$; t = $\hat{\beta}$/EE. [a]63 observaciones con valores no detectados; [b]63 observaciones censuradas en el punto 0,1 μg/g creatinina de mercurio; [c]63 observaciones censuradas según los puntos de censura descritos en la tabla 2.

La estimación de los errores estándar asociados a estos coeficientes fueron mayores en el modelo tobit, seguidas de las obtenidas en el modelo censurado con varios puntos de censura. En el modelo de regresión lineal estos errores estándar fueron sustancialmente menores.

Discusión

Muchas variables epidemiológicas que no miden el tiempo transcurrido desde un momento dado hasta que se produce el evento de interés presentan también distribuciones con censura para las cuales los modelos de regresión lineal no deberían utilizarse, porque proporcionan estimaciones sesgadas e inconsistentes[4-6]. En esta situación es aconsejable la utilización de modelos más adecuados a la naturaleza de la variable de estudio que tengan en cuenta la existencia de censura. La familia de modelos de regresión censurada permite tratar este problema, ya sea con un único o con

varios puntos de censura, y con censura inferior, superior o de intervalo[4-6]. En comparación con los resultados que facilita el modelo de regresión lineal, los que se obtienen utilizando los modelos de regresión censurados no cambian la dirección del efecto estimado. Las principales diferencias se encuentran al cuantificar la estimación de los efectos, tal como se ilustra en el análisis de los valores de mercurio, donde los coeficientes estimados pueden variar en gran medida, así como en la estimación de los errores estándar de dichas estimaciones que intervienen en la significación estadística de estos estimadores. Esto debe ser tenido en cuenta, ya que en la mayoría de los estudios epidemiológicos la cuantificación del efecto es de tanto interés como su significación.

Otro punto a destacar, que pone de manifiesto la importancia de tener en cuenta la censura, es el hecho de que ignorar todas las observaciones censuradas y trabajar exclusivamente con observaciones detectadas hace que la variable que se desea estudiar a escala poblacional tenga una distribución diferente de la variable resultante al obtener la muestra. En particular, el valor medio calculado con la muestra resulta mayor que el valor medio poblacional, si los valores no detectados se sitúan en la cola inferior de la distribución, y resulta menor si los valores no detectados se sitúan en la cola superior de la distribución.

Así, si fuese posible realizar un modelo de regresión lineal conocida la variable latente y^* en la población, se obtendrían los valores reales en las estimaciones $\hat{\beta}$. Pero el efecto de eliminar las observaciones censuradas y de estimar un modelo de regresión lineal es que las estimaciones $\hat{\beta}_{MCO}$ que se obtienen a través del modelo de regresión lineal, estimado por MCO, serán menores (en valor absoluto) que las anteriores y menos precisas. El efecto de introducir las observaciones censuradas y de estimar un modelo que tiene en cuenta la censura es que las estimaciones obtenidas son generalmente mayores (en valor absoluto) que las estimaciones $\hat{\boldsymbol{\beta}}_{MCO}$ y más precisas, por lo que serán más próximas a las verdaderas β. En este sentido, el modelo de regresión censurado con distintos puntos de censura es el más adecuado dada la naturaleza de los datos de nuestro estudio, donde existen varios puntos de censura para las concentraciones de mercurio. Sin embargo, por ser modelos muy sensibles a la falta de normalidad, es muy importante tener en cuenta este aspecto antes de realizar cualquier análisis. Conviene señalar que la ausencia de normalidad de los errores del modelo ocasiona que los estimadores obtenidos sean inconsistentes[4,14].

Las variables evaluadas en nuestro análisis explicaron sólo una pequeña parte de la variabilidad total de la distribución de los valores de mercurio. Además, dos de las variables examinadas –edad y sexo–, si bien pueden determinar directamente las concentraciones de metales porque influyen en el metabolismo, son esencialmente variables indicadoras *(proxy)* de otras fuentes de exposición. Resultados similares han sido observados en otras poblaciones[15], donde factores sociodemográficos explicaron una gran parte de la variación y exposiciones específicas sólo una pequeña parte. Por otro lado, puede extrañar la relación hallada con el consumo semanal de verduras crudas y ahumados, aunque esta última podría sorprender en menor medida si se considera que el mercurio en orina se ha asociado con el consumo de pescado en otros estudios[16,17]. Cabe señalar también que en estudios previos que evaluaban los determinantes de los valores de mercurio en orina o en sangre[15-17] no se han utilizado modelos de regresión que tengan en cuenta la posible censura de las concentraciones de mercurio debido a los límites de detección del aparato de medida, con lo que se estarían proporcionando estimaciones segadas e imprecisas para las variables que se revelaron asociadas.

Agradecimientos Los autores agradecen a Carlos Alberto González la autorización para utilizar los datos del estudio sobre la incineradora de residuos sólidos urbanos de Mataró. Agradecemos también a Mercedes Díez y Roberto Pastor sus sugerencias a las versiones previas, y a los dos revisores anónimos y al miembro del equipo editorial por sus comentarios y sugerencias.

Bibliografía

1. Soriano JB, Antó JM, Sunyer J, Tobías A, Kogevinas M, Almar E, et al. Risk of asthma in the general Spanish population attributable to specific immunoresponse. Int J Epidemiol 1999;28:728-34.
2. Bleda MJ, González CA, Kogevinas M, Huici A, Gadea E, Ladona M, et al. Niveles séricos basales de dioxinas, furanos, PCB's y metales en una muestra de población general en una ciudad española en que se ha instalado una incineradora de residuos sólidos [abstract]. Gac Sanit 1996;10(Supl):56.
3. Duan N, Manning WG, Morris CN, Newhouse JP. Comparison of alternative models for the demand for medical care. J Business Econom Stat 1983;1:115-26.
4. Greene WH. Análisis econométrico. 3.ª ed. Prentice Hall Iberia, Madrid, 1999.
5. Long JS. Regression models for categorical and limited dependent variables. Thousand Oaks: Sage, 1997.
6. Tobin J. Estimation of relationships for limited dependent variables. Econometrica 1958;26:24-36.
7. Van der Gaag J, Van der Ven W. The demand for primary health care. Med Care 1978;16:299-312.
8. Luoma K, Jarvio ML, Suoniemi I, Hjerppe RT. Financial in-

centives and productive efficiency in Fihish health centres. Health Econ 1996;5:435-45.
9. Grootendorst PV. Health care policy evaluation using longitudinal insurance claim data: a Tobit estimator. Health Econ 1997;6:365-82.
10. Rosko MD. Impact of internal and external environmental pressures on hospital. Health Care Manag Sci 1999;2:63-74.
11. González CA, Kogevinas M, Gadea E, Huici A, Bosch A, Bleda MJ, et al. Biomonitoring study of people living near or working at a municipal solid-waste incinerator before and after two years of operation. Arch Environ Health 2000 Jul-Aug;55(4):259-67.
12. Sáez M, Barceló MA. Un criterio para omitir variables superfluas en modelos de regresión. Gac Sanit 1998;12:281-3.
13. StataCorp. Stata Statistical Software: Release 6.0. College Station, TX: Stata Corporation, 1999.
14. Chesher A, Irish M. Residual analysis in the grouped and censored normal linear model. J Econometrics 1987;34:33-61.
15. Sartor F, Rondia D, Claeys F, Bochet JP, Ducoffre G, Lauwerys, S, et al. Factors influencing the cadmiun body burden in a population study. IARC Sci Publ 1992;118:101-6.
16. Salonen JT, Seppanen K, Nyyssonen K, Korpela H, Kauhanen J, Kantola M. Intake of mercury from fish, lipid peroxidation and the risk of myocardial infarction and coronary, cardiovascular and any death in eastern Finish men. Circulation 1995;91:645-55.
17. Svensson BG, Schutz A, Nilsson A, Akesson I, Akesson B, Skerfving S. Fish as a source of exposure to mercury and selenium. Sci Total Environ 1992;126:61-74.