

# Imputación del instante de seroconversión al VIH en cohortes de hemofílicos

S. Pérez-Hoyos<sup>a</sup> / I. Ferreros<sup>a,b</sup> / J. del Amo<sup>b</sup> / M. Quintana<sup>c</sup> / I. Ruiz<sup>d</sup> / J.M. Cisneros<sup>e</sup> / R. Muga<sup>f</sup> / M. García de la Hera<sup>b</sup> / J. del Romero<sup>g</sup> / P. García de Olalla<sup>h</sup> / R. Guerrero<sup>i</sup> / I. Hernández-Aguado<sup>b</sup> y GEMES\*

<sup>a</sup>Unidad de Epidemiología y Estadística. Escuela Valenciana de Estudios para la Salud (EVES). Valencia. España. <sup>b</sup>Departamento de Salud Pública. Universidad Miguel Hernández. San Juan de Alicante. Alicante. España. <sup>c</sup>Unidad de Hemofilia. Hospital La Paz. Madrid. <sup>d</sup>Unidad de Hemofilia. Hospital Vall d'Hebron. Barcelona. <sup>e</sup>Unidad de Hemofilia. Hospital Virgen del Rocío. Sevilla. España.

<sup>f</sup>Hospital Universitario Germans Trias i Pujol. Badalona. Barcelona. <sup>g</sup>Centro Sanitario Sandoval. Madrid.

<sup>h</sup>Agència de Salut Pública. Barcelona. <sup>i</sup>Prisiones de Cataluña. Barcelona. España.

\*La relación de miembros de GEMES aparece recogida al final del artículo.

Correspondencia: S. Pérez Hoyos. Escuela Valenciana de Estudios para la Salud. Joan de Garay, 21. 46017 Valencia. España.

Correo electrónico: perez\_san@gva.es

Recibido: 14 de mayo de 2003.

Aceptado: 16 de septiembre de 2003.

(Imputation of the date of HIV seroconversion in cohorts of haemophiliacs)

## Resumen

**Objetivos:** Describir la técnica utilizada para imputar la fecha de seroconversión al virus de la inmunodeficiencia humana (VIH) en cohortes de hemofílicos y validar su utilización en sujetos con fecha de seroconversión datada dentro de un intervalo de las cohortes del proyecto GEMES (Grupo Español Multicéntrico para el Estudio de Seroconvertidores al VIH).

**Método:** Se dispone de 632 hemofílicos procedentes de tres unidades de hemofilia identificados como VIH+ entre 1983 y 1985 (fechas de las primeras pruebas VIH+) y 1.092 individuos procedentes de 5 cohortes del proyecto GEMES con una ventana de seroconversión (tiempo entre pruebas VIH- y VIH+) menor de 3 años en los que se asume el punto medio (PM) como fecha de seroconversión. Se imputa la fecha de seroconversión para los dos grupos a partir de la estimación de la distribución de probabilidad de seroconversión que se obtiene mediante el algoritmo EM. Se utilizan dos métodos de imputación: uno a partir del valor esperado de la distribución estimada y otro de la media geométrica de 5 muestras aleatorias de la distribución estimada. Las imputaciones se han validado en las cohortes de no hemofílicos al comparar con el PM y el tiempo libre de sida y la supervivencia con las diferentes estimaciones de seroconversión.

**Resultados:** La mediana de la seroconversión se sitúa en mayo de 1993 para los no hemofílicos y en 1982 para los hemofílicos. En las cohortes de no hemofílicos no se observan grandes diferencias entre las fechas de la seroconversión estimada como el punto medio y las medianas estimadas a partir de los dos métodos de imputación. Lo mismo ocurre en el caso de las cohortes de hemofílicos. No se observan diferencias en las estimaciones del tiempo libre de sida y supervivencia en ambos grupos de cohortes.

**Conclusiones:** La imputación de la media geométrica de varias muestras aleatorias y la imputación de valor esperado proporcionan una buena estimación de la fecha de seroconversión frente al VIH que puede ser utilizada para estimar el tiempo

libre de sida y la supervivencia en una cohorte de hemofílicos para los que se desconoce la fecha de seroconversión

**Palabras clave:** Seroconversión. VIH. Imputación. Cohortes.

## Abstract

**Objectives:** To describe the methods used to impute HIV seroconversion date in the haemophiliac cohorts from GEMES project and to validate its use.

**Method:** 632 haemophiliacs coming from three hemophilia units identified as HIV+ and 1.092 individuals coming from 5 project GEMES cohorts with a seroconversion window (time among test HIV- and HIV+) less than 3 years where mid point (PM) was assumed as seroconversion date. For both groups, seroconversion date was imputed after estimating the probability distribution of seroconversion by means of the EM algorithm. Two imputation methods are used: one obtained from the expected value and the other from the geometric mean of 5 random samples. from the estimated distribution. Imputations have been validated in the non haemophiliacs cohorts comparing with the PM seroconversion date. Also AIDS free time and survival from the different seroconversion imputed dates were compared.

**Results:** Median seroconversion date is located in May of 1993 for the non haemophiliacs and in 1982 for the haemophiliacs. Not big differences are observed among the imputed seroconversion dates and the mid-point seroconversion date in the non-haemophiliac cohorts. Similar results are found for the haemophiliac cohorts. Also no differences are observed in the estimated AIDS-free time for both groups of cohorts.

**Conclusions:** Geometric mean imputation from several random samples provides a good estimate of the HIV seroconversion date that can be used to estimate AIDS-free time and survival in haemophiliac cohorts where seroconversion date is ignored.

**Key words:** Seroconversion. HIV. Imputation. Cohorts.

## Introducción

**A** finales de la década de los setenta y principios de los ochenta, una proporción de personas con hemofilia se infectó con el virus de la inmunodeficiencia humana (VIH) a través de productos con sangre contaminada. En España, las técnicas de inactivación viral se empezaron a aplicar de forma obligatoria desde 1985<sup>1</sup>, aunque en algunos centros, como el Hospital Universitario de La Paz de Madrid, se venían utilizando desde 1983. Se estima, por tanto, que los hemofílicos infectados por el VIH en España se contagiaron entre finales de 1970 y 1985. Esta ventana es demasiado imprecisa para el estudio del período de incubación del sida y la valoración de los cambios en la supervivencia desde la seroconversión.

Al analizar el período de incubación del sida, y los tiempos desde la seroconversión al VIH hasta un evento como la muerte, se dispone de lo que se llaman «datos doblemente censurados». Es decir, donde al final del seguimiento el sujeto puede o no haber desarrollado el evento de interés (censura a derecha) y donde se desconoce el inicio del seguimiento (infección en nuestro caso), aunque se sabe que ocurrió en una ventana de tiempo más o menos amplia (censura a izquierda). Para manejar este tipo de información, doblemente censurada, se pueden utilizar diversas técnicas con el objeto de obtener una estimación del período de incubación. En primer lugar, se puede estimar directamente este período, lo que acarrea ciertas complicaciones de cálculo<sup>2</sup>. Otra opción es imputar la fecha de seroconversión y utilizar un método de análisis de supervivencia habitual con únicamente censuras a derecha<sup>3-8</sup>. Algunos de los métodos de imputación están basados en la correlación entre biomarcadores de progresión como los CD4 con el tiempo de incubación<sup>4,5,8,7</sup>.

En España, se han utilizado dos alternativas para estimar la probable fecha de infección. Montoro et al asignan a todos ellos una fecha común de seroconversión, 1982, considerando que sería la fecha más probable de infección<sup>9</sup>. En un trabajo previo del Grupo Español Multicéntrico para el Estudio de Seroconvertidores al VIH (GEMES) se utilizó como fecha de seroconversión el punto medio entre un instante del calendario que se considera inicio de la exposición al virus y la fecha en la cual los sujetos fueron identificados como seropositivos<sup>10</sup>. Ninguna de estas dos estrategias parece la más adecuada para estimar la fecha de seroconversión.

En el marco del proyecto GEMES se han identificado 3 cohortes de hemofílicos infectados por el VIH de tres de las unidades de hemofilia más grandes de España (Hospital la Paz de Madrid, Hospital Vall d'Hebron de Barcelona y Hospital Virgen del Rocío de Sevilla), de las que se dispone de la fecha de la pri-

mera prueba VIH+ y un seguimiento que permite analizar la evolución del período de incubación del sida y otros eventos. Por otra parte, el mismo proyecto GEMES dispone de 5 cohortes de seroconvertidores al VIH no hemofílicos, con una ventana de seroconversión suficientemente estrecha como para poder asignar el momento de la seroconversión con precisión.

El objetivo del presente trabajo es describir las técnicas utilizadas para imputar la fecha de seroconversión en las cohortes de hemofílicos de GEMES y validar su utilización en el resto de cohortes del proyecto para las que se dispone de una estimación precisa de dicha fecha.

## Sujetos y métodos

### *Sujetos*

Se dispone de 632 hemofílicos, procedentes de tres unidades de hemofilia de los hospitales Hospital la Paz de Madrid, Hospital Vall d'Hebron de Barcelona y Hospital Virgen del Rocío de Sevilla, que han sido identificados como seropositivos al VIH (mediana de fecha de diagnóstico en noviembre de 1985). Al no disponer de pruebas VIH negativas, se ha considerado que la seroconversión no pudo ocurrir antes de 1979 ni después de 1985, fecha en la que todos los concentrados de plasma fueron tratados con calor para inactivar el VIH, según la ley en España. Para los nacidos después de 1979 se considera el nacimiento como el límite inferior de la ventana de seroconversión.

Se dispone de 1.092 seroconvertidores, con una ventana de seroconversión inferior a los 3 años, provenientes de 5 cohortes de seroconvertidores del proyecto GEMES: la cohorte de usuarios de drogas por vía intravenosa (UDI) de los Centros de prevención del sida de Valencia, Alicante y Castellón (CIPS-Valencia), la cohorte del Centro de enfermedades de transmisión sexual Sandoval de Madrid (Sandoval-Madrid), la cohorte de la unidad de desintoxicación de drogas del Hospital Universitario Germans Trias i Pujol de Badalona (HUG-TIP-Badalona), la cohorte reclutada por el Institut Municipal de Salut de Barcelona de los centros de tratamiento y vigilancia de drogadictos (IMSP-Barcelona) y la cohorte reclutada de las prisiones de Cataluña (Prisiones-Cataluña). Más detalle de estas cohortes se puede encontrar en otros trabajos previos<sup>10-11</sup>.

### *Métodos*

Para los sujetos de las cohortes de seroconvertidores no hemofílicos, se consideró la fecha de seroconversión como el punto medio entre la última visita con

un resultado negativo de la prueba para el VIH y la primera visita con un resultado positivo de la prueba para el VIH, que es una buena aproximación al instante de la seroconversión, sobre todo para su utilización en el análisis del tiempo de progresión de la infección<sup>12</sup>. Este modelo presume que todos los instantes de tiempo dentro de la ventana de seroconversión tienen la misma probabilidad de haber sido la fecha de seroconversión (fig. 1).

Para los sujetos de las cohortes de hemofílicos, al ser la ventana de seroconversión superior a 3 años en la mayoría de los casos, se ha optado por utilizar un método de imputación de la fecha de la seroconversión basado en la estimación de la distribución de probabilidad de seroconversión a lo largo del tiempo<sup>6,3</sup>.

En figura 1 se muestra la ventana de seroconversión del individuo *i*, situada entre las fechas  $y_n$  e  $y_{ip}$ . Parece lógico pensar que el instante de seroconversión  $y_{is}$  tenga más probabilidad de haber ocurrido en la parte de la ventana por la que coinciden más ventanas de seroconversión de otros sujetos que por la parte donde coinciden menos sujetos, es decir, en la parte donde parece que hay más incidencia que en la que parece que hay menos.

Basándose en esta idea, se estima la probabilidad ( $h_j$ ) de que los individuos de la cohorte seroconviertan en un instante del calendario ( $y_j$ ). Posteriormente, a partir de dicha probabilidad, se imputa la fecha de seroconversión. Para el cálculo de esta probabilidad se utiliza una versión del algoritmo autoconsistente propuesto por Turnbull<sup>13</sup>, que no es más que una versión del algoritmo EM (esperanza-maximización)<sup>14</sup>. A continuación, se muestra cómo funciona este algoritmo.

En primer lugar, se considera el calendario dividido en meses y se asume una forma no paramétrica de la distribución de la probabilidad de seroconvertir en un mes.

Se puede definir una variable aleatoria para cada individuo ( $I_{ij}$ ) que sea indicadora de su fecha de seroconversión. Así,  $I_{ij}$  valdrá 1 si el individuo *i* seroconvierte en la fecha  $y_j$  y 0 en otro caso. En la figura 1 esta variable valdrá 1 en el instante  $y_{is}$  (línea gruesa) y 0 en el resto de los puntos.

También se puede definir una variable indicadora auxiliar  $\alpha_{ij}$  que indique para cada sujeto *i* los posibles tiempos  $y_j$  en los que puede seroconvertir. Esta variable vale

1 si el instante  $y_j$  cae dentro de la ventana de seroconversión del sujeto *i* y 0 en otro caso.

Dado que se desconoce el instante de la seroconversión, el valor de la variable  $I_{ij}$  no se puede observar directamente para cada individuo *i*, si bien se sabe que valdrá 0 fuera de la ventana de seroconversión. El valor esperado de esta variable no es más que la probabilidad de que el individuo *i* seroconvierta en  $y_j$ . Esta probabilidad se puede calcular como el cociente entre la probabilidad ( $h_j$ ) de seroconvertir en el instante  $y_j$  y la suma de las probabilidades de seroconvertir en todos los instantes ubicados dentro de la ventana de seroconversión:

$$E[I_{ij}] = \mu_{ij}(h) = \frac{\alpha_{ij}h_j}{\sum_{t=1}^p \alpha_{it}h_t} =$$

Suma P (seroconvertir en los  $y_t$  de la ventana de seroconversión)

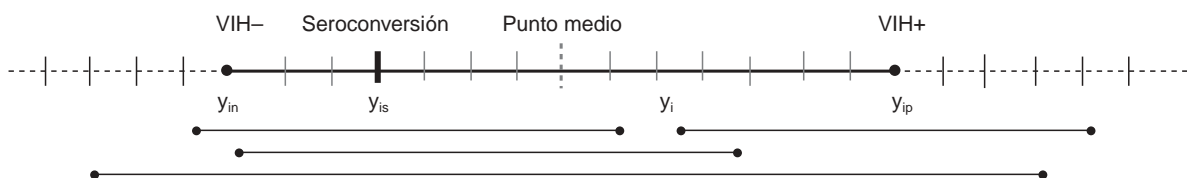
Éste es el paso E del algoritmo EM, donde se observa que la parte derecha de la ecuación incluye el valor de la probabilidad de seroconvertir que se desea calcular.

Por otra parte, esta probabilidad de seroconvertir en un instante  $y_j$  se puede obtener como el promedio de las probabilidades de seroconvertir en ese instante de todos los sujetos de la cohorte, es decir, el promedio de los valores obtenidos en la ecuación 1. Sería la parte M del algoritmo EM:

$$h_j = \frac{\sum_{i=1}^N \mu_{ij}(h)}{N} \text{ donde N es el número de sujetos}$$

Con lo cual, los valores de interés  $h_j$  se encuentran en la parte derecha de la primera ecuación y como resultado de la segunda. Por ello, bastará dar unos valores iniciales de la probabilidad de seroconvertir  $h_j$ , sustituir en la ecuación 1 y calcular el valor de la probabilidad  $E[I_{ij}]$  de que un individuo *i* seroconvierta en el instante  $y_j$ . Sustituyendo en la ecuación 2, se puede obtener un nuevo valor de las probabilidades de seroconvertir  $h_j$  el instante. Sustituyendo los nuevos valores en la ecuación 1 se itera sucesivamente entre las ecuaciones 1 y 2 hasta obtener la convergencia, es decir,

Figura 1. Esquema de fechas en el intervalo de seroconversión.



hasta que las diferencias del estimador de las probabilidades de seroconversión  $h_t$  no sea muy diferente en dos pasos sucesivos. Turnbull demostró que este proceso conduce a obtener un estimador adecuado de la distribución de probabilidad de seroconversión<sup>13</sup>.

Una vez obtenida dicha distribución de probabilidad se puede estimar la fecha de seroconversión  $y_{is}$  para cada individuo de dos modos:

1. En el primer método se calcula el valor esperado de la seroconversión para cada individuo condicionada a la ventana de seroconversión a partir de la distribución de probabilidad obtenida por el algoritmo anterior. No es más que el promedio de las fechas de seroconversión ponderadas por la probabilidad asignada a cada una de ellas. Este promedio obtenido para cada individuo se imputa como fecha de seroconversión (ESP):

$$\hat{y}_{is} = \frac{\sum_{t=1}^p y_t \alpha_t h_t}{\sum_{t=1}^r \alpha_t h_t}$$

2. En el segundo método se seleccionan cinco muestras aleatorias de la fecha de seroconversión para cada individuo, a partir de la distribución de la probabilidad de seroconvertir de cada uno de ellos obtenida en la ecuación 1. A partir de los cinco valores muestreados para cada sujeto y por la posible asimetría de la distribución, se calcula la media geométrica que se imputa como fecha de seroconversión (MED):

$$\hat{y}_{is} = \exp \left\{ \frac{\sum_{t=1}^5 \log(y_{muestra t})}{5} \right\}$$

Las características diferentes de cada una de las cohortes analizadas, tanto desde el punto de vista del momento en el que fueron reclutados los sujetos como por la incidencia de la infección en las distintas áreas geográficas y los diferentes grupos de riesgo, nos llevaron a imputar para cada cohorte por separado las fechas de seroconversión.

Para validar todo el proceso se han aplicado los dos métodos de imputación propuestos a cada una de las cohortes de no hemofílicos para los que se disponía de una ventana de seroconversión precisa (menos de 3 años). En estos casos se utiliza como fecha de seroconversión el punto medio de dicha ventana (PM). Se han comparado los resultados de la imputación con la fecha asignada por el punto medio mediante el gráfico de cuantiles (qq-plot)<sup>15</sup> y se ha contrastado la igualdad de la distribución de las fechas mediante el test de Kolmogorov-Smirnov para dos muestras<sup>15</sup>.

Además, se ha obtenido la estimación del tiempo libre de sida mediante el estimador de Kaplan-Meier utilizando como punto de inicio del seguimiento las fechas de seroconversión imputadas por ambos métodos. Los resultados se comparan con las estimaciones obteni-

das a partir de la que utiliza el PM de la ventana de seroconversión como inicio del seguimiento. En todos los casos se ha utilizado la opción de entrada retrasada (*delay, late or staggered entry*) a la primera visita VIH+<sup>16</sup>, con lo cual se incluye a cada individuo en el período con riesgo que empieza a contar desde el momento en el que se inició el seguimiento, corrigiendo el sesgo de supervivencia de los sujetos VIH- de los que no se efectuó seguimiento<sup>17</sup>.

Para el caso de las cohortes de hemofílicos, también se ha obtenido el estimador del tiempo libre de sida por medio del método de Kaplan-Meier, utilizando como punto de inicio de la infección las fechas obtenidas por los dos métodos de imputación. Se han considerado dos estimaciones más, tomando como inicio para todos los sujetos la fecha de 1982 y como inicio de la infección el punto medio entre 1979 y la primera visita VIH+. Para los nacidos después de 1979 se ha considerado como límite inferior la fecha de nacimiento.

Para las imputaciones y los gráficos se ha utilizado el programa S-Plus<sup>15</sup> y para la estimación del tiempo libre de sida se ha utilizado el programa Stata<sup>16</sup>.

## Resultados

De las 8 cohortes consideradas en el estudio, 3 son de hemofílicos, 4 de usuarios de drogas y una está formada mayoritariamente (69,5%) por homosexuales (tabla 1).

La mediana de la amplitud del intervalo de seroconversión en las cohortes de no hemofílicos se sitúa alrededor de un año en todas las cohortes. Las cohortes de Valencia, Badalona y Sandoval se empezaron a reclutar más pronto y, por tanto, son más maduras. La mediana de la seroconversión se sitúa en mayo de 1993 y la de la primera prueba positiva, en enero de 1994. El porcentaje global de casos de sida en las cohortes de no hemofílicos fue del 21% (10,2-34,8%). La mortalidad total fue del 14% (3,1-29,4%).

En las tres cohortes de hemofílicos la mediana de identificación de los casos se sitúa a finales de 1985 para las cohortes de Vall d'Hebron y La Paz y en los inicios de 1987 para la cohorte de Virgen del Rocío. El porcentaje de casos de sida fluctúa entre el 43,6 y el 69,6% y el de muerte, entre el 43,6 y el 69,1%.

En las cohortes de no hemofílicos no se observan grandes diferencias entre las medianas de la fecha de la seroconversión estimada como el punto medio y las medianas estimadas a partir de los dos métodos de imputación, tanto el que utiliza el valor esperado de la distribución de seroconversión (ESP) como el que calcula la media geométrica de 5 muestras tomadas al azar (MED) de la distribución de probabilidad de la fecha de

Tabla 1. Análisis descriptivos de las fechas de interés para las cohortes de seroconvertidores de GEMES

	Cohorte							
	CIPS Valencia	Sandoval- Madrid	HUGTIP- Badalona	IMSP- Barcelona	Prisiones- Cataluña	Virgen del Rocío	Vall d'Hebron	La Paz
N	241	239	201	25	157	94	149	379
Fecha de la última visita VIH- (mediana)	1991,54	1992,96	1987,53	1994,10	1995,00			
Fecha de la primera visita VIH+ (mediana)	1992,75	1994,15	1988,48	1995,35	1996,37	1987,07	1985,86	1985,82
Fecha de sida (mediana)	1996,92	1996,07	1993,50	1996,50	1997,18	1991,42	1991,46	1990,53
Fecha de muerte (mediana)	1996,82	1995,30	1994,59	1996,98	1998,39	1992,85	1993,84	1992,70
Intervalo de seroconversión (mediana)	1,02	0,96	0,99	0,96	1,13			
Fecha de seroconversión (punto medio) (mediana)	1992,14	1993,51	1987,95	1994,75				
Fecha imputada al valor esperado (mediana)	1992,09	1993,65	1988,15	1994,71	1995,80	1982,36	1983,48	1982,68
Fecha imputada (media geométrica) (mediana)	1992,04	1993,67	1988,14	1994,76	1995,83	1982,34	1983,57	1982,68
Categoría de exposición Varones homosexuales (%)		69,5						
Usuarios de drogas por vía parenteral (ADVP) (IDU) Hemofílicos (%)	100,0	24,3	100,0	88,5	84,6			
Sida	27,8	18,8	34,8	17,8	10,2	50,0	43,6	69,6
Muerte (%)	19,9	7,5	29,4	17,8	3,1	47,9	43,6	69,1

seroconversión (tabla 1). En los gráficos de cuantiles de la figura 2 se observa que la distribución de las fechas de seroconversión según los métodos ESP y MED (eje de ordenadas) no son muy diferentes de haber utilizado el PM (eje de abscisas) en cada una de las cohortes, como muestra el hecho de que los puntos se sitúen en torno a la diagonal. En ningún caso se observan diferencias apreciables entre los valores imputados y el PM. Al aplicar la prueba de Kolmogorov-Smirnov para 2 muestras para comparar la igualdad de la distribución de fechas entre el PM y los métodos ESP y MED, no se observaron diferencias significativas entre ambas distribuciones, tanto cuando se comparan todos los valores en el conjunto de todas las cohortes ( $p = 0,99$  para ESP y  $p = 0,87$  para MED) como en cada una de las cohortes por separado.

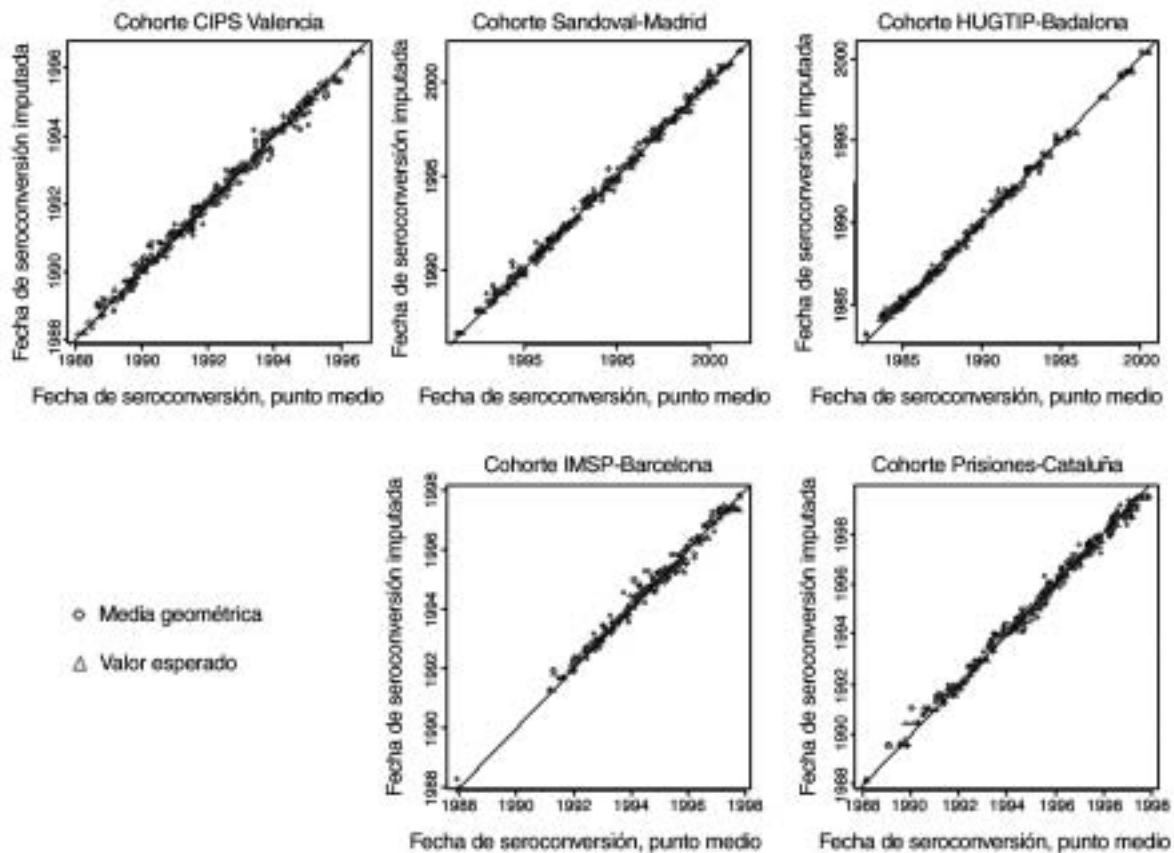
En la figura 3a, se muestra la estimación de Kaplan-Meier del tiempo libre de sida utilizando los tres métodos de estimación de la seroconversión: el PM y los tiempos imputados según ESP y MED. Los resultados de dicha estimación fueron muy similares. En esta figura se ha añadido una zona sombreada que corresponde al área entre las estimaciones del tiempo libre

de sida más extremas, es decir, utilizando como inicio del seguimiento las fechas de la última visita VIH- y la primera visita VIH+. El porcentaje libre de sida a los 10 años fue del 57,7% para el método MED, el 57,8% para el método ESP y el 58,1% para el PM.

Al analizar el tiempo libre de sida para las cohortes de hemofílicos, tampoco se observan grandes diferencias de haber utilizado el PM entre 1979 y la primera prueba VIH+ con respecto a los dos métodos de imputación. Así, el porcentaje de hemofílicos libres de sida a los 10 años fue del 58,6% para el PM, el 59,1% para el MED y el 58,7% para el ESP. Si se hubiera utilizado el año 1982 como fecha de seroconversión para todos los hemofílicos el porcentaje libre de sida a los 10 años sería 59,9%, si bien se observan algunas diferencias al inicio y al final de las curvas de Kaplan Meier (fig. 3b). Lo que es evidente es que no es adecuado utilizar la fecha de la primera visita VIH+, donde el porcentaje libre de sida a los 10 años sería el 44,8%, ni tampoco 1979, donde dicho porcentaje sería el 82,7%.

En la figura 4 se pueden observar los intervalos de seroconversión para las cohortes de hemofílicos y se

Figura 2. Distribución de los cuantiles de la fecha de seroconversión como punto medio frente a la fecha imputada a partir del valor esperado o la media geométrica de 5 muestras de la distribución de probabilidad del instante de seroconversión.



muestran las tres fechas de seroconversión utilizadas, el PM del intervalo de seroconversión, el valor esperado de la distribución de probabilidad de seroconversión para cada cohorte y la media geométrica de 5 muestras extraídas de dicha distribución. Se observa que al utilizar la media geométrica hay una mayor variabilidad en las fechas de seroconversión, por lo que se eligió este último método para ulteriores trabajos. La utilización del PM presume que se produjeron infecciones después de 1985.

### Discusión

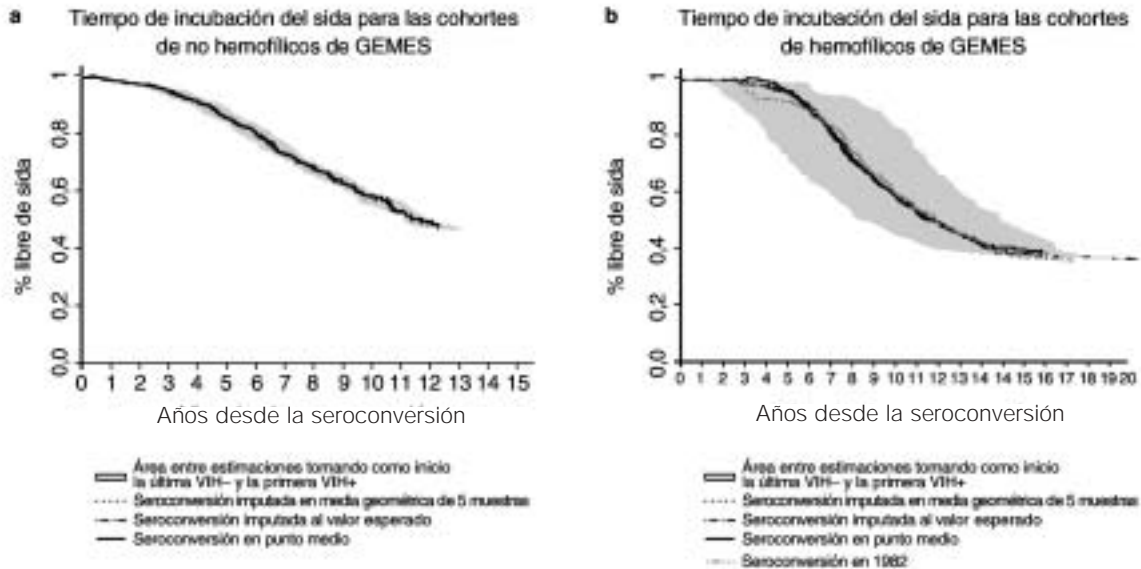
En el presente trabajo se muestra el uso de la imputación de la fecha de seroconversión al VIH como un método para poder solventar la ausencia de pruebas VIH negativas previas a la hora de analizar la historia de su progresión en cohortes de hemofílicos.

La alternativa de intentar imputar la fecha de seroconversión en lugar de utilizar técnicas de análisis de

datos doblemente censurados se ha mostrado como una buena opción para evitar la complejidad de este tipo de análisis<sup>3,6,12</sup>. Geskus<sup>3</sup>, a partir de un estudio de simulación, señala que tanto la imputación del punto medio, como la del valor esperado o la media geométrica de muestras extraídas de la distribución de la infección no presentan diferencias notables con el análisis de los datos doblemente censurados, utilizando la máxima verosimilitud. Si bien este autor indica que el método de la media geométrica ofrece una subestimación de la curva del tiempo libre de sida.

En el caso de aplicación a las cohortes de no hemofílicos, donde la ventana entre dos medidas VIH- y VIH+ era suficientemente pequeña, el PM es una buena estimación de la fecha de seroconversión<sup>12</sup>. Al comparar los resultados del PM con las dos imputaciones obtenidas a partir de la distribución de la probabilidad de infección, hemos observado que no había grandes diferencias, tanto en las fechas de seroconversión como en la estimación del tiempo libre de supervivencia. Tampoco se encuentran grandes dife-

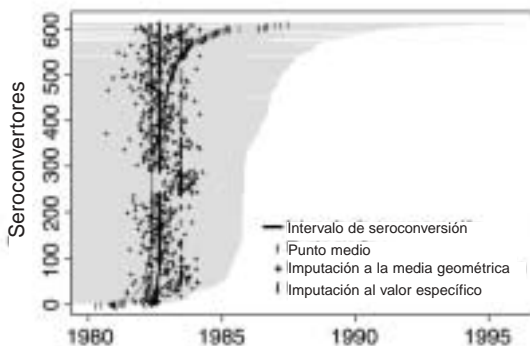
Figura 3. Estimación del tiempo libre de sida para las cohortes de GEMES utilizando diferentes fechas de seroconversión.



rencias en las estimaciones de los coeficientes de modelos de Cox, donde se analizan factores que influyen en la progresión a sida o muerte, como la edad, el sexo o el período de calendario (resultados no expuestos).

En nuestro estudio no encontramos una subestimación del tiempo libre de sida para el caso de la imputación a partir de la media geométrica de 5 muestras como en otros estudios, quizá debido al hecho de que los datos son de seroconvertidores con ventanas de seroconversión no muy grandes; en el caso del trabajo de Geskus se utilizaron datos de casos prevalentes e incidentes<sup>3</sup>.

Figura 4. Intervalos de seroconversión para las cohortes de hemofílicos y fechas de seroconversión imputadas.



Cuando se aplican los métodos a las cohortes de hemofílicos, se observa de nuevo que los tres métodos de imputación (punto medio, valor esperado y media geométrica de la distribución de probabilidad) producen estimaciones sobre el tiempo que no son muy diferentes. Sin embargo, la utilización de un punto de inicio común como 1982<sup>9</sup> produce algunos sesgos en la estimación del período de incubación a pesar de que la mediana de la fecha imputada esté muy próxima.

Tampoco es adecuado utilizar la fecha de la primera visita VIH+ como la fecha de inicio del seguimiento, ya que se produce una infraestimación del tiempo de incubación, y se llega a obtener un porcentaje de tiempo libre de sida a los 10 años inferior al 45%. El hecho de utilizar una entrada retrasada conlleva que en los primeros 3 años de seguimiento no se observen casos de sida, con lo cual es posible que la progresión a la enfermedad mostrada sea para los pacientes que sobrevivieron a este período, y aunque la entrada retrasada corrija el sesgo de supervivencia, no lo elimina del todo<sup>19</sup>.

Law y Brookmeyer<sup>12</sup> indicaron que el punto medio era una manera adecuada de imputar la fecha de seroconversión si la ventana no era demasiado amplia. En el caso de los hemofílicos, tomando como el límite inferior de la seroconversión 1979, hemos observado que el PM no ofrece estimaciones del tiempo libre de sida muy diferentes del caso de los métodos basados en la distribución de probabilidad de seroconversión. Sin embargo, la imputación al PM en las cohortes de hemofílicos presume que se pudieron

infectar después de 1985, lo que no ha sucedido en nuestro medio.

Las diferencias en la incidencia de infección, así como de progresión a sida y muerte que existe entre las cohortes de hemofílicos y no hemofílicos, no deben afectar a la validez del método de estimación utilizado. Más bien lo contrario, ya que el uso del PM cuando la ventana es muy amplia puede ocasionar sesgos en la estimación del período de incubación<sup>6,12</sup>, aunque no parecen relevantes si la ventana de seroconversión es inferior a 3 años<sup>19</sup>.

En este trabajo se han utilizado la media geométrica de 5 muestras aleatorias de la distribución de probabilidad de seroconversión para imputar el instante del inicio de la infección por el VIH. La selección de 5 muestras ha sido arbitraria, pero no se han encontrado diferencias apreciables en la imputación utilizando 10 o 20 muestras (resultados no expuestos).

Otras formas de imputar la seroconversión han utilizado algunos biomarcadores de progresión de la infección, como el CD4. Estos métodos se basan en la modelización paramétrica de la evolución del biomarcador desde la seroconversión<sup>5,8</sup>, o en métodos no paramétricos basados en la evolución individual de los marcadores. En ambos casos, se necesita de una cohorte seroincidente de referencia que permita obtener los modelos de evolución de los biomarcadores de progresión, lo que no sucede en el caso de los hemofílicos.

Los resultados del presente trabajo indican que la imputación de la media geométrica de varias muestras aleatorias proporcionan una buena estimación de la fecha de seroconversión frente al VIH, que puede ser utilizada para estimar el tiempo libre de sida y la supervivencia en una cohorte de hemofílicos para los que se desconoce la fecha de seroconversión.

#### Agradecimientos

Este trabajo ha sido financiado parcialmente gracias a ayudas para la investigación de FIPSE (Fundación para la in-

vestigación y la prevención del sida en España) integrada por el Ministerio de Sanidad Abbott Laboratories, Boehringer Ingelheim, Bristol Myers Squibb, GlaxoSmithKline, Merck Sharp and Dohme and Roche), (FIPSE 3023/99) (FIPSE 36349/02) y FIS (Fondo de Investigación Sanitaria) (FIS 99/0195, 99/1065, 00/294, 02/0639).

Las fuentes de financiación sólo han contribuido en la ayuda para el establecimiento de las cohortes, la recogida de los datos de las cohortes y la financiación de las reuniones de coordinación. El trabajo de Inmaculada Ferreros ha estado financiado con la ayuda otorgada por FIPSE.

#### Miembros de GEMES

*Centro Coordinador. Departamento de Salud Pública. Universidad Miguel Hernández (UMH):* I. Hernández Aguado, J. del Amo.

*Centro de Análisis de datos de GEMES. Escuela Valenciana de Estudios para la Salud (EVES):* S. Pérez-Hoyos, I. Ferreros.

*Cohorte de los CIPS de la Comunidad Valenciana. Departamento de Salud Pública UMH. EVES:* M. García de la Hera, I. Hurtado.

*Cohorte del Hospital Germans Trias i Pujol en Badalona:* R. Muga, A. Sanvicens, B. Clotet, J. Tor.

*Cohorte de Madrid-Sandoval:* J. del Romero, P. Raposo, C. Rodríguez, Soledad García, V. Soriano, y el Grupo de Seroconvertidores de la Comunidad de Madrid.

*Cohorte de los Centros de Atención y Prevención de SIDA de Barcelona. Institut de Medicina i Salut de Barcelona. Cohorte IMSP:* J. Cayla, P. García de Olalla, T. Brugal.

*Cohorte de las Prisiones de Cataluña:* R. Guerrero, A. Marco. *Cohorte de hemofílicos del Hospital La Paz:* M. Quintana, A. Barrasa, J. del Amo.

*Cohortes de hemofílicos del Hospital Vall d'Hebron:* I. Ruiz, J. Tussell.

*Cohortes de hemofílicos del Hospital Virgen del Rocío de Sevilla:* R. Pérez, J.M. Cisneros.

*Registro Nacional de SIDA:* J. Castilla, A. Barrasa.

#### Bibliografía

1. Ministerio de Sanidad. Circular del 30 de septiembre de 1985 de la Dirección General de Farmacia y Productos Sanitarios, por los que se dan normas aclaratorias para el mejor cumplimiento de la resolución del 6 de septiembre de 1985. BOE n.º 235 de martes 1 de octubre de 1985. 2003.
2. De Gruttola V, Lagakos SW. Analysis of doubly-censored data, with application to AIDS. *Biometrics* 1989;45:1-11.
3. Geskus RB. Methods for estimating the AIDS incubation time distribution when date of seroconversion is censored. *Stat Med* 2001;20:795-812.
4. Alcabes P, Muñoz A, Vlahov D, Friedland G. Maturity of human immunodeficiency virus infection and incubation period of acquired immunodeficiency syndrome in injecting drug users. *Ann Epidemiol* 1994;4:17-26.
5. Muñoz A, Carey V, Taylor J, Chmiel JS, Kingsley L, Van Raden M, et al. Estimation of time since exposure for a prevalent cohort. *Stat Med* 1992;11:939-52.
6. Gauvreau K, DeGruttola V, Pagano M, Bellocco R. The effect of covariates on the induction time of aids using improved imputation of exact seroconversion times. *Stat Med* 1994;13:2021-30.
7. Geskus RB. On the inclusion of prevalent cases in HIV/AIDS natural history studies through a marker-based estimate of time since seroconversion. *Stat Med* 2000;19:1756-69.
8. Brancato G, Pezzoti P, Rapiti E, Perucci C, Abeni D, Babbalacchio A, et al. Multiple imputation method for estimating incidence of HIV infection. *Int J Epidemiol* 1997;26:1107-14.
9. Montoro JB, Oliveras J, Lorenzo JI, Tusell JM, Altisent C, Molina R, et al. An association between clotting factor concentrates



- use and mortality in human immunodeficiency virus-infected hemophilic patients. *Blood* 1995;86:2213-9.
10. GEMES. El período de incubación del sida en España antes de la terapia antirretroviral de alta eficacia. *Med Clin (Barc)* 2000;115:681-6.
  11. Pérez-Hoyos S, Del Amo J, Muga R, Del Romero J, García de Olalla P, Guerrero R, et al. Effectiveness of highly active anti-retroviral therapy in Spanish cohorts of HIV seroconverters: differences by transmission category. *AIDS* 2003;17:353-9.
  12. Law CG, Brookmeyer R. Effects of mid-point imputation on the analysis of doubly censored data. *Stat Med* 1992;11:1569-78.
  13. Turnbull BW. The Empirical distribution function with arbitrarily grouped, censored and truncated data. *JRSS* 1976;38:290-5.
  14. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *JRSS* 1977;1-22.
  15. S-Plus 6 for windows. Seattle: Insightful Corporation; 2001.
  16. Pérez-Hoyos S. Análisis de supervivencia. En: *Quaderns de Salut Pública i Administració de Serveis de Salut* n.º 11. Valencia: Institut Valencià d'Estudis en Salut Pública. Conselleria de Sanitat; 1997.
  17. Lamarca R, Alonso J, Gómez G, Muñoz A. Left-truncated data with age as time scale: An alternative for survival analysis in the elderly population. *J Gerontol A Biol Sci Med Sci* 1998; 5:337M-43M.
  18. Stata Corp. *Stata Statistical Software: Release 8.0*. College Station: Stata Corporation, 2003.
  19. Porter K, Johnson AM, Philips AN, Darbyshire JH. The practical significance of potential biases in estimates of the AIDS incubation period distribution in the UK Register of HIV Seroconverters. *AIDS* 1999;13:1943-51.
-