

ORIGINAL**IMPUTACIÓN DEL INSTANTE DE INICIO DE SEGUIMIENTO EN ESTUDIOS LONGITUDINALES: APLICACIÓN A LA INFECCIÓN POR VIH****Santiago Pérez-Hoyos (1), Inmaculada Ferreros (1,2), Isabel Hurtado (1), Julia del Amo (2), Ildefonso Hernández-Aguado (2)**

(1) Unidad Epidemiología y Estadística. Escuela Valenciana de Estudios en Salud (EVES)

(2) Departamento de Salud Pública. Universidad Miguel Hernández

RESUMEN

Fundamentos: En los estudios de cohorte o longitudinales los sujetos son reclutados con un tiempo de evolución desde el inicio del problema, como en el caso de la infección por VIH. El objetivo de este trabajo es mostrar diversas técnicas de imputación del inicio del seguimiento y valorar su adecuación en el marco de un estudio sobre la evolución de la infección por VIH.

Sujetos y métodos: Se dispone de tres subcohortes de sujetos VIH+ reclutados en los centros de información y prevención del SIDA de Valencia y Castellón hasta 1996. 244 seroincidentes, 887 prevalentes con mediciones de CD4 y 337 sin mediciones de CD4 en los que se estima la fecha de seroconversión. Para los seroincidentes se considera el punto medio entre la última visita VIH- y la primera VIH+. Para los prevalentes con CD4 se imputa la fecha de seroconversión a partir de 5 extracciones aleatorias de un modelo de progresión desde la infección hasta un nivel determinado de CD4. Para los sujetos prevalentes sin CD4 se imputa la fecha de seroconversión a partir de 5 muestras aleatorias de la densidad de incidencia de VIH obtenida de las otras dos cohortes. Se replicó 500 veces la imputación, asignando como seroconversión la mediana de las imputaciones y obteniendo límites de confianza a partir de los percentiles 5 y 95. La validez de las imputaciones se contrastó comparando los tiempos a SIDA y Muerte para cada uno de los tres grupos.

Resultados: Se observaron 443 casos de SIDA y 405 defunciones en los 1.468 sujetos. La mediana de infección fue enero de 1993 para los incidentes, enero de 1991 para los prevalentes con CD4 y noviembre de 1988 para los prevalentes sin CD4. Este último grupo muestra una peor supervivencia y tiempo libre de sida respecto a las otras dos cohortes.

Conclusiones: Los instrumentos de imputación utilizados han mostrado su utilidad para disminuir el sesgo de supervivencia en estudios observacionales. Su generalización depende de la viabilidad de poder obtener cohortes incidentes, de disponer de un marcador de progresión o un punto de origen.

Palabras clave: Imputación. Seroconversión. VIH. Síndrome de inmunodeficiencia adquirida.

ABSTRACT**Imputation of the beginning of follow up in Longitudinal Studies Applied to the HIV Infection**

Background: In cohort or longitudinal studies, subjects are recruited some time after the beginning of the problem, as in HIV infection. The aim of this paper is to show several imputation techniques of the beginning of follow up and evaluate its use in the framework of a study of HIV progression.

Methods: Three subcohorts of HIV+ subjects recruited in Valencia and Castellón CIPS up to 1996 are available. Seroconversion date was estimated for 244 Seroincidentes, 887 seroprevalents with CD4 measurements and 337 without CD4 measurements. For seroincidentes midpoint between last HIV- and first HIV+ visits was considered. For prevalent with CD4 seroconversion date was imputed from 5 random samples of a progression model of infection to a CD4 level. For prevalent without CD4 seroconversion date was imputed from 5 random samples of HIV incidence density obtained from the other subcohorts. The imputation was repeated 500 times, assigning the seroconversion date as the median of imputations and obtaining confidence limits from 5 and 95 percentiles. Imputation validity was tested comparing time to AIDS and death for each one of the 3 groups.

Results: 443 and 405 deaths were observed among the 1468 subjects. Median of seroconversion was January 1993 for incidents, January 1991 for prevalents with CD4 and November 1988 for prevalents without CD4. The latest group showed a worse survival and AIDS free time compared to the other two cohorts.

Conclusions: The imputation tools used showed their usefulness to reduce the survival bias in observational studies. Their generalization depends on the viability of incident cohorts, the availability of a progression marker or a origin on time.

Key words: Imputation. Seroconversion. HIV. Immunodeficiency acquired syndrome.

Correspondencia:
Santiago Pérez Hoyos
Unidad de Epidemiología y Estadística
Escuela Valenciana de Estudios para la Salud (EVES)
C/ Joan de Garay 21 - 46017 Valencia
Correo electrónico: perez_san@gva.es

INTRODUCCIÓN

Los estudios de cohorte o longitudinales se caracterizan por analizar el tiempo que transcurre entre el inicio de entrada de los participantes en el estudio hasta que se produce un evento de interés, como la muerte, una enfermedad, el cambio en alguna medida biológica, etcétera. Una de las características de los estudios observacionales con seguimiento es que los datos que se recogen reflejan la realidad de la progresión y evolución de los problemas que se analizan¹ aproximándose a la historia natural de la enfermedad. Así, cuando se analiza la efectividad de medidas terapéuticas, éstas han sido asignadas a los diferentes sujetos según sus características o necesidades, no según la aleatorización de un ensayo clínico, aunque ello pueda conllevar el llamado sesgo por indicación² generado por el hecho de que aquellos que tienen mayor necesidad o acceso a los tratamientos van a ser susceptibles de recibirlos.

Muchas de las cohortes que analizan períodos de incubación o evolución de enfermedades hasta un desenlace, están basadas en sujetos prevalentes para los que se desconoce el tiempo de evolución de la enfermedad. Es el caso de la infección por el virus de la inmunodeficiencia humana (VIH) y el tiempo hasta sida o muerte. En el momento de reclutar a los sujetos muchos llevan ya tiempo infectados, incluso algunos de ellos son identificados como VIH+ cuando ya tienen sida, lo que puede acarrear importantes sesgos en las estimaciones de supervivencia a lo largo del tiempo³.

Una forma de obviar estos problemas es utilizar estudios de seroconvertidores, donde la fecha de infección está basada en la aparición de síntomas de primoinfección, o en la existencia de una ventana estrecha entre dos mediciones VIH negativa y VIH positiva, que permita asignar como fecha de seroconversión a un punto intermedio⁴⁻⁶. Sin embargo los sujetos candidatos a formar parte de

estos estudios no son tan numerosos como en el caso de las cohortes prevalentes. El desarrollo en los años 90 de métodos de análisis que permiten la entrada retrasada (*late o staggered entry*), es decir que tienen en cuenta el tiempo transcurrido desde la infección hasta el reclutamiento, han permitido la posibilidad de yuxtaponer la información proveniente de cohortes de sujetos seroprevalentes y seroincidentes. Numerosos avances en el conocimiento de la infección por VIH y en la evaluación de nuevas terapias son el resultado de estudios en los que se ha mezclado esta información^{7,8}.

Estos métodos pretenden resolver el problema de los datos doblemente censurados⁹ es decir, donde se conoce que el inicio está en un intervalo (censura a izquierda), por ejemplo desde que un usuario de drogas empieza a inyectarse hasta que se le realiza una prueba VIH+ y donde al final del seguimiento unos sujetos han desarrollado el evento y otros no (censura a la derecha). Existen formas de estimar directamente el período de incubación utilizando datos doblemente censurados, aunque conlleva ciertas complejidades de cálculo y no pueden utilizarse programas de análisis estándar¹⁰⁻¹².

El otro tipo de análisis utilizado consiste en estimar el instante de la seroconversión al VIH, y utilizar un método tradicional de análisis del tiempo de incubación con censuras a la derecha. En el caso de los sujetos seroincidentes, para los que se dispone de una ventana estrecha de seroconversión, se ha comprobado que asignar como instante de seroconversión el punto medio de dicho intervalo es una buena opción^{13,14}.

En el caso de los sujetos prevalentes, en los que las ventanas de seroconversión son más amplias, se han utilizado otras alternativas. Una es imputar, para cada individuo, la fecha de la seroconversión a partir de la estimación de la distribución de la incidencia de infección condicionada a sus fechas de

VIH+ y VIH-^{10-15,16}. La otra alternativa es utilizar imputaciones basadas en la correlación con marcadores de la progresión como los CD4 o la carga viral. Estas imputaciones se basan en modelos generados con cohortes de seroconvertidores en los que se modela el tiempo transcurrido desde la seroconversión hasta el nivel del marcador tanto de forma paramétrica^{17,18} como no paramétrica¹⁹.

El objetivo del presente trabajo es mostrar diversas técnicas de imputación del inicio del seguimiento y valorar su adecuación en el marco de un estudio sobre la evolución de la infección por VIH

SUJETOS Y MÉTODOS

Sujetos

Se dispone de datos procedentes de un estudio longitudinal de la infección por VIH a partir de sujetos usuarios de drogas intravenosas (UDIs) identificados en los centros de información y prevención del SIDA (CIPS) de Valencia y Castellón, reclutados desde finales de 1987 hasta finales de junio de 1996 (figura 1). En estos centros se ofrece información y consejo y se practica de forma voluntaria y anónima la prueba serológica frente al VIH, sugiriendo su retorno al cabo de 6 meses o su derivación a un centro asistencial²⁰.

Del total de 4.860 UDIs que acudieron antes de julio de 1996 2.198(45,2%) fueron identificados como VIH+ (seroprevalentes), de los cuales 1.224 (55,7%) disponían de seguimiento. Del 54,7% que resultaron VIH-, 216 resultaron VIH+ en una posterior medición y fueron identificados como seroincidentes. Posteriormente se añadieron 28 sujetos que fueron identificados como seroincidentes tras ser reclutados después de la fecha indicada.

Del total de 1.224 seroprevalentes con seguimiento, 887 (72,5%) disponían de una

medición de un marcador de progresión, en concreto del porcentaje de CD4 y fueron clasificados como la subcohorte de seroprevalentes con CD4. Los 337 sujetos restantes no disponían de esta medición y fueron clasificados como subcohorte de seroprevalentes sin CD4.

Métodos

Como se indicó, el objetivo del estudio era valorar la adecuación de la imputación de la fecha de seroconversión al VIH para cada uno de los sujetos de las subcohortes mencionadas para su posterior uso en la estimación del tiempo libre de sida y Supervivencia.

Imputación para la subcohorte seroincidentes

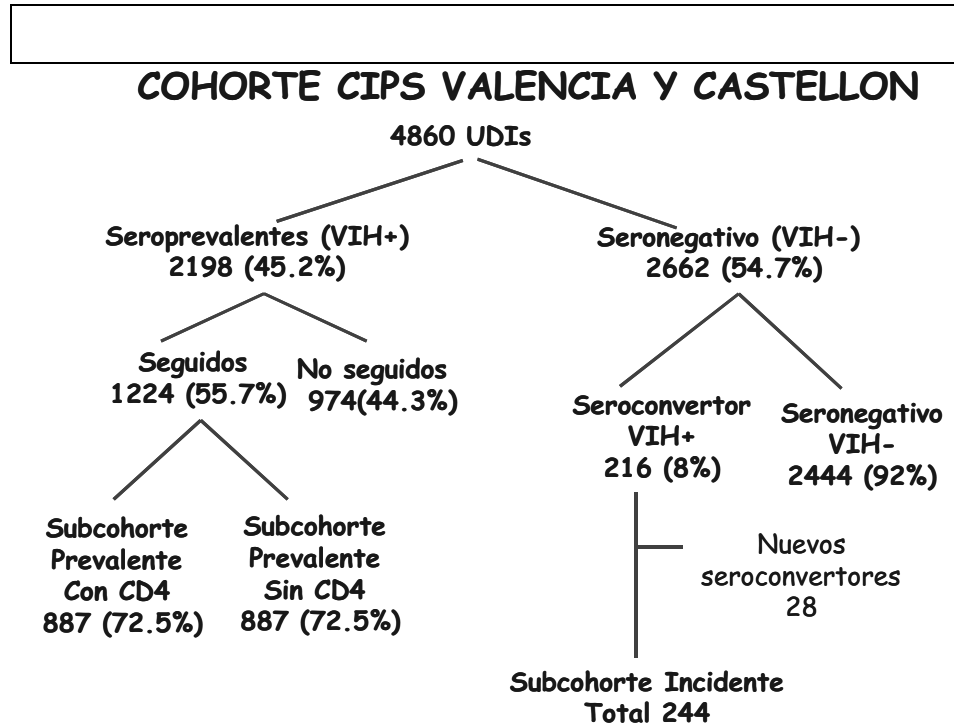
Para la cohorte seroincidente se utilizó como instante de seroconversión el punto medio entre la última medición VIH negativa disponible y la primera medición VIH+ que se ha considerado como una buena aproximación al instante de la seroconversión¹³.

Imputación para la subcohorte de seroprevalentes con CD4

Para la subcohorte de sujetos seroprevalentes se imputó la fecha de seroconversión a partir del modelo de progresión de la infección descrito por Muñoz et al¹⁷. En este trabajo se modeló, para una cohorte de personas seroincidentes, el tiempo T de progresión entre el instante de la infección (S_c) y el instante de una medición del CD4 (F_{cd4}), dado el porcentaje de CD4 observado, utilizando un modelo de supervivencia Weibull²¹. Así la probabilidad de que el tiempo transcurrido desde la seroconversión sea mayor que t dado un nivel x de porcentaje de CD4 se calcula como:

Figura 1

Diagrama de flujo de los individuos usuarios de drogas intravenosas (UDIs) que acuden a los centros de información y prevención del SIDA de Valencia y Castellón



$P(T>t) = \exp(-\alpha t^{1/\sigma})$ donde $\alpha = \exp(-\beta x/\sigma)$, con β el vector de regresión asociado al nivel x del porcentaje de CD4 y σ es el factor de escala o dispersión de dicho modelo. En el trabajo de Muñoz se describen los valores de β y σ para distintos niveles del % de CD4 (tabla 1).

En el caso de la cohorte de sujetos prevalentes estamos interesados en conocer el tiempo w entre la fecha de seroconversión (Sc) y la primera visita VIH+ (Fp). Por una parte conocemos el tiempo d_1 transcurrido

entre la fecha en la primera visita VIH+ y la fecha de la medición del % de CD4 ($Fcd4$) y por otra el tiempo d_2 transcurrido entre la primera vez que se inyectaron o 1980, considerados como límites del inicio de la posible infección y la fecha de la medición del CD4 (figura 2).

Aplicando el modelo Weibull anterior, la probabilidad de que el tiempo entre la seroconversión y la primera visita VIH+ sea $w+d1$ condicionado a que este tiempo no puede ser mayor que d_2 se puede calcular como:

$$w = -d_1 - \left(\ln \left\{ \frac{\exp(-\alpha d_1^{1/\sigma}) - \exp(-\alpha d_2^{1/\sigma})}{\exp(-\alpha d_2^{1/\sigma})} \right\} \right) \frac{1}{\alpha} \sigma$$

Tabla 1

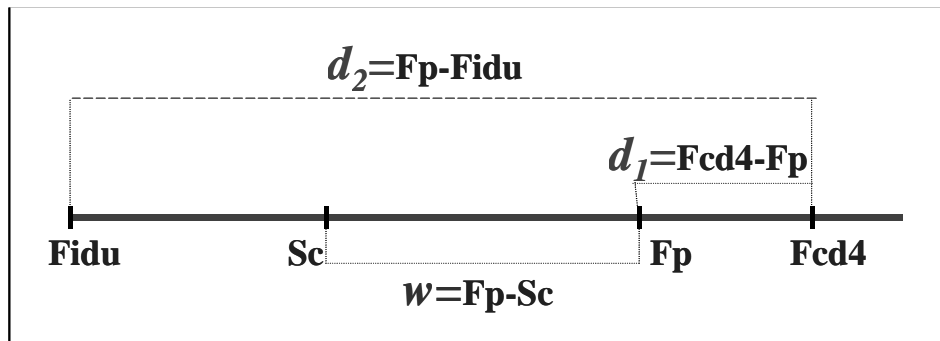
Parámetros del modelo de regresión Weibull para el tiempo transcurrido desde la seroconversión utilizado para imputar dicho tiempo en la cohorte de prevalentes con CD4¹⁷

	% CD4				
	(0-17,5)	(17,5-22,5)	(22,5-27,5%)	(27,5-37,5)	(37,5-100)
β	4,29	4,61	3,38	3,23	3,06
σ	0,41	0,60	0,60	0,67	0,74

Los parámetros de la tabla permiten obtener la probabilidad de tener un determinado porcentaje de CD4 en función de la ecuación $P(\text{Tiempo desde seroconversión} > t) = \exp(-\alpha t^\sigma)$ donde $\alpha = \exp(-\beta/\sigma)$ con β y σ dependiendo del nivel de % de CD4

Figura 2

Esquema de fechas disponibles y a imputar en la cohorte de seroprevalentes



Fidu= Máximo entre fecha inicio inyección de drogas o 1980
 Sc= Fecha de seroconversión
 Fp= Fecha de primera visita VIH positiva
 Fcd4= Fecha de medición de % CD4

En la ecuación anterior se puede despejar el tiempo de interés w en función del resto de parámetros y se puede obtener un valor si se conoce la probabilidad del tiempo desde la seroconversión a la primera visita VIH+.

A partir de las probabilidades p_u de una distribución uniforme (0,1) se extrajeron 5 muestras aleatorias para cada individuo. Con los 5 tiempos w calculados sustituyendo las extracciones en la ecuación anterior, se obtuvo la media geométrica (gmw). Para cada individuo se imputa la fecha de seroconver-

sión como la diferencia entre la fecha de la primera visita menos el tiempo gmw estimado como transcurrido entre la seroconversión y la primera visita VIH+.

Para tener en cuenta la variabilidad de la imputación se repitió el proceso 500 veces, como un procedimiento *bootstrap*²², asignando la mediana de las 500 imputaciones como la fecha de seroconversión y los percentiles 5 y 95 de los valores obtenidos en las 500 imputaciones como los límites de confianza de dicha fecha.

Imputación para la subcohorte de seroprevalentes sin CD4

Para el caso de la cohorte de seroprevalentes sin CD4 no se puede utilizar el método anterior. En este caso se ha utilizado un método de imputación basado en la estimación de la distribución de probabilidad de seroconversión a lo largo del tiempo que se puede encontrar detallado en otras publicaciones^{10;15-16}. Para cada individuo de la cohorte de seroprevalentes sin CD4, se dispone de una ventana de seroconversión entre la fecha de inicio de inyección (*Fidu*) y la primera fecha positiva (*Fp*). Conociendo la densidad de incidencia del periodo, podemos asignar a cada fecha de la ventana de seroconversión una probabilidad de que el individuo seroconvierta en ella.

$$P_{ij} = \text{Prob}(\text{seroconvertir sujeto } i \text{ en instante } y_{ij}) = \frac{\text{prob}(\text{seroconvertir en } y_{ij})}{\text{Suma prob}(\text{seroconvertir en los instantes de la ventana de seroconversion})}$$

Una vez obtenidas estas probabilidades para cada individuo se extrajeron, a partir de ellas, 5 valores aleatorios del instante de seroconversión para cada sujeto. Por la posible asimetría de la distribución de probabilidad de seroconversión se calculó la media geométrica y se asignó como la fecha de seroconversión.

Análogamente al apartado anterior, se ha repetido el proceso 500 veces. Se ha tomado como fecha de seroconversión la mediana de las 500 imputaciones y como límite de confianza de dichas fechas los percentiles 5 y 95 de las imputaciones.

Análisis

Se presenta un análisis descriptivo de las características de los sujetos (fecha y edad en la primera visita VIH+, género, fecha de sida, muerte) según el grupo de imputación al que pertenecen. Se presentan gráficos con la ventana individual de seroconversión, las

Para las cohortes de sujetos seroincidentes y seroprevalentes con CD4 se puede estimar la densidad de incidencia h_t para cada instante de calendario, a partir de las fechas de seroconversión imputadas. Esta densidad no es más que la probabilidad de contraer la infección en cada instante. Para el cálculo de esta incidencia se ha utilizado como ventana de seroconversión, en el caso de la cohorte de prevalentes con CD4, los percentiles 5 y 95 de las fechas imputadas en el apartado anterior.

Asumiendo que para los grupos en los que no se dispone de CD4 la distribución de la incidencia de VIH es la misma que para los otros grupos, se puede obtener la probabilidad de seroconvertir en cada uno de los instantes de la ventana de seroconversión como.

fechas imputadas y los percentiles 5 y 95 de las imputaciones.

Para validar la imputación se estimó el tiempo libre de sida y el de la supervivencia desde la seroconversión utilizando el estimador de Kaplan-Meier con la opción de entrada retrasada hasta la primera visita con seroconversión positiva, con lo cual cada individuo es incluido a riesgo en el momento del inicio del seguimiento, corrigiendo el sesgo de supervivencia de aquellos sujetos que no fueron seguidos²³. Se han comparado estos tiempos para las tres subcohortes utilizando la prueba de log-rank o Mantel-Cox. Además se presentan las estimaciones de los tiempos libres de sida y de muerte utilizando los percentiles 5 y 95 de las fechas de seroconversión imputadas.

Las imputaciones y los gráficos han sido elaborados con programas y macros elaborados *ad hoc* en S-Plus 2000²⁴. Para el análisis de supervivencia se ha utilizado también el programa Stata 8.0²⁵.

RESULTADOS

Como se ha mencionado anteriormente, de los 1.468 sujetos estudiados 244 formaban parte de la cohorte de personas seroincidentes, 337 de la cohorte de personas prevalentes sin CD4 y 887 de la cohorte de personas prevalentes con CD4. En la tabla 2 se muestran algunos análisis descriptivos de estas cohortes. El 25,3% son mujeres, no observándose diferencias dependiendo de las cohortes.

Se observó un total de 443 casos de sida, lo que supone el 39,2% de los individuos. Se

observa un menor número de casos de sida en la cohorte de sujetos incidentes. También el porcentaje de muertes es menor en el caso de la cohorte de incidentes, donde sólo un 15,2 % fallecen, frente a un 40,1% en la cohorte de sujetos prevalentes sin CD4 y un 26,7% en la cohorte de sujetos prevalentes con CD4. En total se observaron 405 defunciones.

Si se tiene en cuenta la fecha de la primera visita VIH+ se observa que la cohorte de prevalentes sin CD4 es la que más tiempo lleva en seguimiento (mediana enero 1990) mientras que la cohorte de incidentes tiene

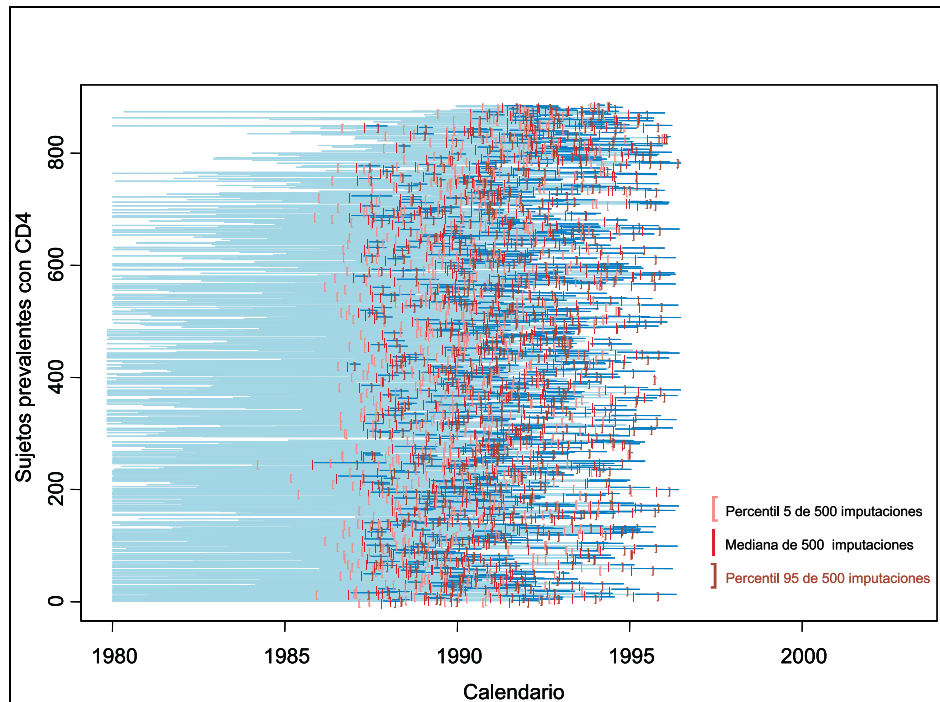
Tabla 2

Estadísticos descriptivos según cohorte de las variables iniciales y de las fechas imputadas

	Total	Incidente	Prevalente sin CD4	Prevalente con CD4
N	1.468	244	337	887
Nº casos (%)				
Mujeres	372 (25,3%)	67 (27,58%)	89 (26,4%)	216 (24,6%)
Hombres	1096 (74,6%)	177 (72,5%)	248 (73,6%)	671 (75,6%)
Sida	443 (39,2%)	59 (24,2%)	112 (33,2%)	272 (30,7%)
Muerte	405 (27,6%)	37 (15,2%)	135 (40,1%)	233 (26,7%)
Medianas intervalo Inter cuartílico)				
Fecha 1ª VIH+	Julio 1991 (Enero 90-Agosto 93)	Febrero 1994 (Enero 92- Julio 96)	Enero 1990 (Agosto 88- Marzo 91)	Octubre 1991 (Abril 90-Agosto 93)
Fecha Sida	Diciembre 1995 (Septiemb. 93- Septiemb. 97)	Noviembre 1997 (Enero 96- Diciemb. 99)	Febrero 1994 (Agosto 92- Marzo 96)	Enero 1996 (Febrero 94-Octubre 97)
Tiempo de seguimiento a sida (años)	5,33	5,32	4,07	6,03
Fecha Muerte	Junio 1996 (Agosto 94-Mayo 98)	Febrero 1998 (Marzo 96- Octubre 99)	Octubre 1994 (Abril 93- Septiemb.96)	Enero 1997 (Septiembre 95- Diciemb. 98)
Tiempo de seguimiento a muerte (años)	6,13	6,03	4,40	7,01
Mediana Imputaciones				
Fecha Seroconversión		Enero 1993	Noviembre 1988	Enero 1991
Percentil 5% Fecha seroconversión			Febrero 1988	Abril 1990
Percentil 95% Fecha seroconversión			Junio 1989	Julio 1991

Figura 3

Imputación de la Fecha de seroconversión para la cohorte de prevalentes con CD4 2



Cada línea representa el tiempo transcurrido entre el inicio de la drogadicción ó 1980 y la primera fecha positiva. Se representan la mediana de las 500 imputaciones y los percentiles 5 y 95.

una mediana cuatro años posterior (mediana febrero 1994). De la misma forma se observa que las fechas de sida y las fechas de muerte también son posteriores en la cohorte de incidentes.

En cuanto al tiempo de seguimiento desde la seroconversión imputada hasta el sida o la muerte, también es inferior para la cohorte de prevalentes sin CD4 que para las otras dos cohortes.

Para la cohorte de incidentes la mediana de la fecha de seroconversión, calculada como el punto medio entre la última visita VIH- y la primera VIH+, fue enero de 1993.

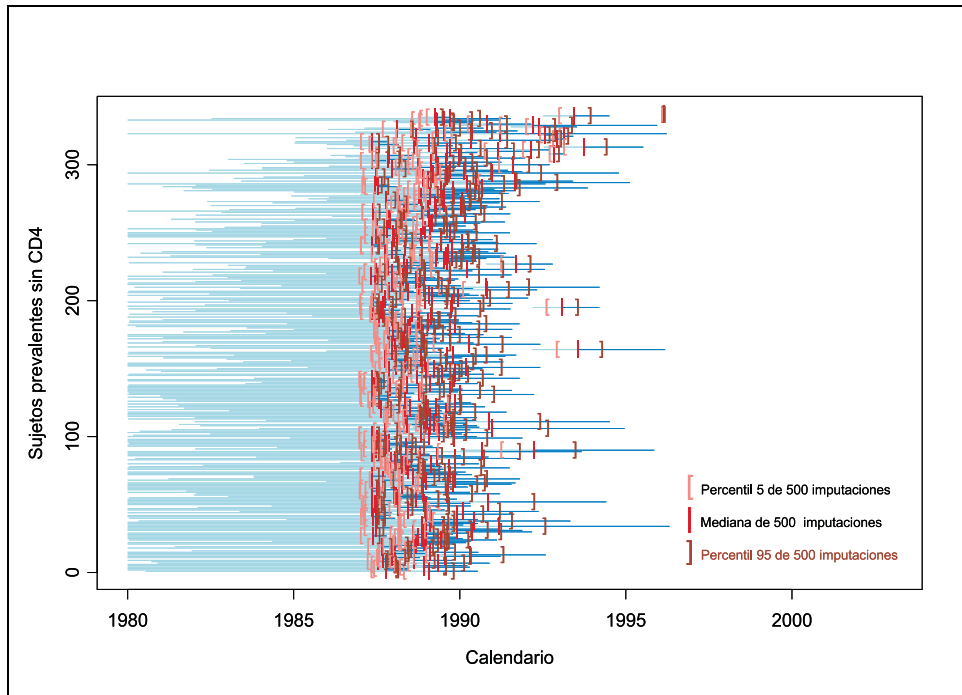
En la figura 3 se muestra la imputación efectuada para el caso de los sujetos preva-

lentes con CD4. Cada línea horizontal representa para cada individuo el tiempo entre la fecha de inicio del uso de drogas o 1980, lo que ocurriera antes. Se muestra la mediana individual de las 500 imputaciones efectuadas así como los percentiles 5% y 95%. La mediana de la fecha de seroconversión imputada se sitúa en enero de 1991, con una mediana de los percentiles 5% y 95% entre abril de 1990 y julio 1991. Así, el tiempo mediano transcurrido entre la fecha de seroconversión imputada y la primera visita VIH es de 0,72 años (intervalo entre percentiles 5 y 95: 0,33-1,22).

En la figura 4 se muestra la imputación efectuada para los sujetos prevalentes sin CD4. La mediana de la fecha de seroconver-

Figura 4

Imputación de la Fecha de seroconversión para la cohorte de prevalentes sin CD4



Cada línea representa el tiempo transcurrido entre el inicio de la drogadicción ó 1980 y la primera fecha positiva. Se representan la mediana de las 500 imputaciones y los percentiles 5 y 95.

sión imputada se sitúa en noviembre de 1988, con una mediana de los percentiles 5% y 95% entre febrero de 1988 y junio de 1989. El tiempo mediana entre la seroconversión imputada y la primera visita VIH+ se sitúa en 1,89 años (intervalo entre percentiles 5 y 95: 0,71-3,56).

Una forma de validar el modelo es comparar el tiempo libre de sida y muerte para las tres cohortes analizadas. En la figura 5 se observa el tiempo libre de sida. Se observa que la cohorte de incidentes y de prevalentes con CD4 tienen un tiempo de incubación semejante (log-rank test $p=0,889$), con una mediana de progresión alrededor de 11 años. En el caso de la cohorte de prevalentes sin CD4 la mediana de progresión se sitúa en 7,5

años, observando diferencias significativas entre las curvas ($p<0,001$). En sombreado, sobre cada curva se sitúan las estimaciones de las curvas de tiempo libre de sida utilizando los percentiles 5% y 95% de la imputación de las fechas de seroconversión de cada individuo.

En la figura 6 se encuentran las curvas de supervivencia desde la seroconversión. Se observan resultados análogos que en el caso del tiempo libre de sida. No hay diferencias entre la cohorte de incidentes y la cohorte de prevalentes con CD4 ($p=0,126$). No así en el caso de la cohorte de prevalentes sin CD4 cuya curva está significativamente por debajo de las otras dos.

Figura 5

Tiempo libre de SIDA desde la seroconversión imputada para cada una de las cohortes

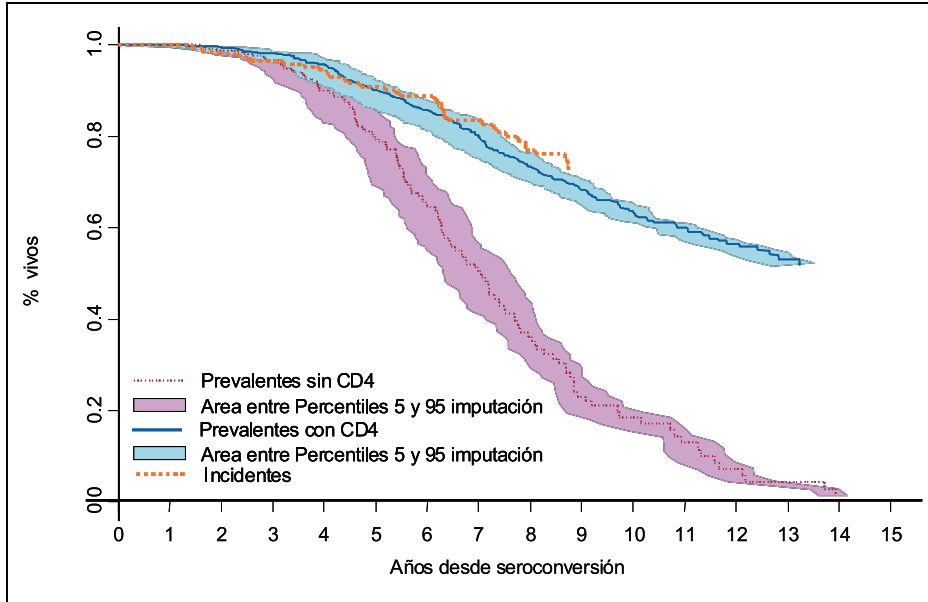
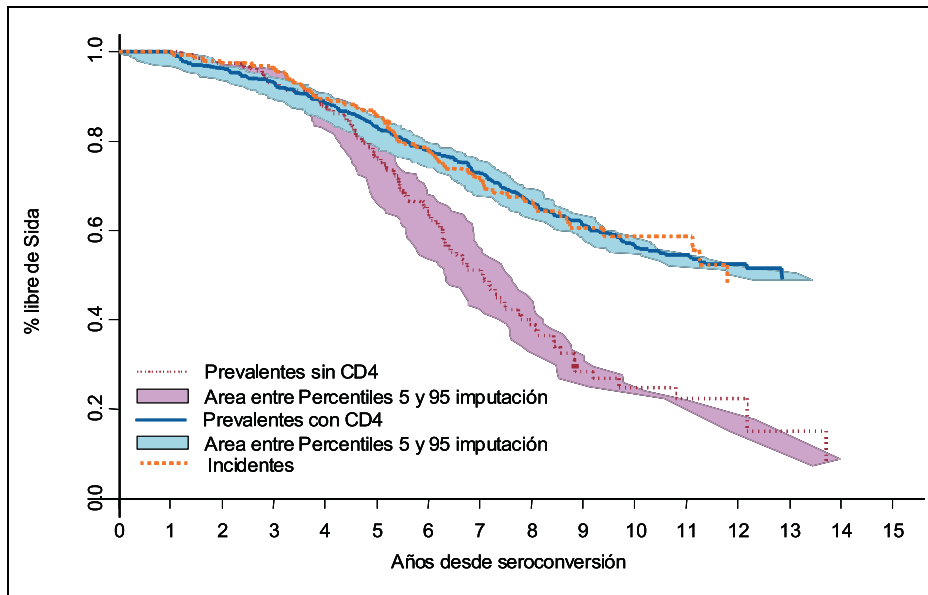


Figura 6

Supervivencia desde la seroconversión imputada para cada una de las cohortes



DISCUSIÓN

El presente trabajo muestra el uso de la fecha de imputación de la seroconversión al VIH como un método para poder incluir datos de sujetos seroprevalentes en el seguimiento de la progresión de la infección, aumentando así el número de sujetos en el estudio. Con ello podemos evitar o disminuir los importantes sesgos que se producen en el análisis de períodos de incubación o evolución de enfermedades basadas en estudios de sujetos prevalentes³.

En el caso de los sujetos incidentes, el uso del punto medio se ha mostrado como un buen estimador del instante de infección, produciendo buenas estimaciones tanto de la incidencia como del tiempo de progresión a sida o muerte¹³⁻²⁰.

En el caso de los sujetos prevalentes la opción del punto medio no es válida por el excesivo tiempo que pasa, en general, entre una medida posible de inicio de la infección, como es el inicio de la drogadicción y la primera visita VIH+.

La existencia de un marcador de la progresión como el CD4 nos ha permitido imputar la fecha de seroconversión y analizar el tiempo libre de sida y la supervivencia. La alternativa de utilizar modelos de doblecensura comporta la necesidad de cálculos algo más complejos y no poder utilizar software estándar^{10,11,12}.

En el caso de disponer de información sobre un marcador de progresión como el CD4, lo que ocurre en una parte importante de los sujetos seroprevalentes, se puede imputar la fecha de seroconversión a partir de un modelo paramétrico, en nuestro estudio el propuesto por Muñoz et al¹⁷. para la cohorte del estudio MACS. Otra opción hubiera sido ajustar un modelo parecido en el grupo de incidentes y obtener una estimación de los parámetros del modelo en nuestro entorno. Brancato et al¹⁸. efectuaron esta

estimación en Italia y observaron que los parámetros eran muy semejantes a los obtenidos en el MACS. En un estudio previo, aplicando el modelo a la cohorte de incidentes, observamos que no habían diferencias sensibles al utilizar la imputación por medio del modelo o por medio del punto medio, aunque existía un pequeño sesgo en el caso de sujetos con un porcentaje bajo de CD4 (<17,5%)²⁶. Otra alternativa es utilizar métodos no paramétricos de imputación basándose en la trayectoria de los CD4 de los individuos incidentes para los que se conoce su fecha de seroconversión¹⁹. El principal problema es que requiere disponer de buenas mediciones de este indicador, que no es el caso de nuestros datos.

Como indica Geskus¹⁰ en el caso de sujetos seroprevalentes, la imputación a partir de la estimación de la distribución de la incidencia como la utilizada en el caso de los prevalentes sin CD4, es una buena aproximación a la fecha de la seroconversión. En su estudio observó que la curva de Kaplan-Meier era aproximadamente la misma que la de la cohorte incidente. En nuestro caso la distribución de la incidencia ha sido la obtenida de forma conjunta para las cohortes seroincidente y de prevalentes con CD4, una vez imputada la fecha de seroconversión. Al ser estas dos cohortes ligeramente posteriores, quizá haya algún tipo de sesgo en la incidencia observada en los primeros instantes del calendario y produzca desviaciones en las estimaciones. En este caso, los instantes de seroconversión se desplazarían hacia la derecha, donde habría una mayor incidencia. Ello podría explicar, en parte, el hecho de que la cohorte de prevalentes con CD4 tenga una peor progresión tanto a sida como a muerte, ya que el tiempo de progresión se ha visto recortado al desplazar a la derecha la fecha de imputación. Sin embargo, como los sujetos de esta cohorte son anteriores en el calendario, es probable que fueran sujetos que llegaran con un mayor tiempo de evolución desde la infección, debido al desconocimiento sobre el VIH que existía y que, efec-

tivamente, tengan una peor evolución a sida y a muerte.

Este método se ha mostrado útil para cohortes de hemofílicos y al aplicarlo a la cohorte de personas seroincidentes se obtienen fechas de seroconversión imputadas semejantes a la fecha obtenida por el punto medio. Además, en este caso el tiempo de progresión a sida y a muerte es el mismo cualquiera que sea el método de imputación utilizado.

Para efectuar las imputaciones se utilizan macros de S-Plus²⁴ por la facilidad de replicar 500 veces el proceso y dibujar los resultados, pero en el caso de las cohortes prevalentes con CD4 el modelo se puede implementar fácilmente en una hoja de cálculo.

El uso de 500 réplicas de la imputación permite obtener un intervalo de variación de las fechas imputadas. Como se observa en el apartado de resultados, el método de imputación es bastante robusto y no se producen grandes variaciones en la estimación de la progresión independientemente de la imputación utilizada.

Sin embargo, si en lugar de imputar la fecha de seroconversión utilizáramos directamente la primera fecha de VIH+ como inicio de la infección, obtendríamos importantes sesgos en las estimaciones de los tiempos de progresión. Así por ejemplo la mediana de progresión a sida en el grupo de prevalentes sin CD4 se reduciría de 7,5 a 5,9 años.

Los resultados de este estudio muestran diversas formas de poder controlar el sesgo de supervivencia. Su generalización a diferentes estudios observacionales de seguimiento depende de la viabilidad de poder obtener una cohorte de incidentes, de disponer de un marcador de progresión o un límite inferior de inicio. Recientemente están desarrollándose métodos que permiten tener en cuenta el tiempo que transcurre hasta el

inicio del estudio y que permiten tener en cuenta aquellos progresores rápidos que no han sido reclutados por el estudio y que permiten replicar en condiciones reales ensayos clínicos²⁷. Por ello la disponibilidad de instrumentos como los mostrados permite aumentar el poder de los estudios observacionales y mejorar la calidad de sus resultados.

AGRADECIMIENTOS

A los compañeros del proyecto GEMES (Grupo Español Multicéntrico para el Estudio de Seroconvertidores) en cuyo seno se pudieron desarrollar algunas de las ideas y métodos del presente estudio. Este trabajo ha sido financiado parcialmente gracias a ayudas para la investigación de FIPSE (Fundación para la investigación y la prevención del sida en España) integrada por el Ministerio de Sanidad, Abbott Laboratories, Boehringer Ingelheim, Bristol Myers Squibb, Glaxo Smith Kline, 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.

BIBLIOGRAFÍA

1. Samet JM, Muñoz A. Evolution of the cohort study. *Epidemiol Review* 1998;20: 1-14.
2. Ahdieh L, Gange SJ, Greenblatt R, Minkoff H, Anastos K, Young M et al. Selection by indication of potent antiretroviral therapy use in a large cohort of women infected with human immunodeficiency virus. *Am J Epidemiol* 2000;152:923-33.
3. Cascade Collaboration. Effect of ignoring the time of HIV seroconversion in estimating changes in survival over calendar time in observational stu-

- dies: results from CASCADE. *AIDS* 2000;14:1899-906.
4. GEMES (Grupo Español Multicéntrico para el Estudio de Seroconvertidores). El periodo de incubación del SIDA en España antes de la terapia HAART. *Med Clin* 2000;115:681-6.
 5. Cascade Collaboration. Time from HIV-1 seroconversion to AIDS and death before the widespread use of highly active anti-retroviral therapy: a collaborative re-analysis. *Lancet* 2000;355:1131-7.
 6. Muñoz A, Wang MC, Bass S, Taylor JM, Kingsley LA, Chmiel JS et al. Acquired immunodeficiency syndrome (AIDS)-free time after human immunodeficiency virus type 1 (HIV-1) seroconversion in homosexual men. Multicenter AIDS Cohort Study Group. *Am J Epidemiol* 1989;130:530-9.
 7. Muñoz A, Sabin A. The incubation period of AIDS. *AIDS* 1997;11(Suppl A):S69-S76.
 8. Detels R, Muñoz A, McFarlane G, Lawrence AK, Margolick JB, Giorgi J et al. Effectiveness of potent antiretroviral therapy on time to AIDS and death in men with known HIV infection duration. *JAMA* 1998;280:1497-503.
 9. De Gruttola V, Lagakos SW. Analysis of doubly-censored survival data, with application to AIDS. *Biometrics* 1989;45:1-11.
 10. Geskus RB. Methods for estimating the AIDS incubation time distribution when date of seroconversion is censored. *Stat Med* 2001;20:795-812.
 11. Gomez G, Lagakos SW. Estimation of the infection time and latency of AIDS with doubly censored data. *Biometrics* 1994;50:204-12.
 12. Sun J, Liao Q, Pagano M. Regression analysis of doubly censored failure time data with application to censored. *Biometrics* 1999;55:909-14.
 13. Law CG, Brookmeyer R. Effects of mid-point imputation on the analysis of doubly censored data. *Stat Med* 1992;11:1569-78.
 14. Cascade Collaboration. The relationships between the HIV test interval, demographic factors and HIV disease progression. *Epidemiol Infect* 2001;127:91-100.
 15. Gauvreau K, DeGruttola V, Pagano M, Bellocchio R. The effect of covariates on the induction time of aids using improved imputation of exact seroconversion times. *Stat Med* 1994;13:2021-30.
 16. Perez-Hoyos S, Ferreros I, del Amo J, Quintana M, Ruiz I, Cisneros JM et al. Imputación del instante de seroconversión al VIH en cohortes de hemofílicos. *Gac Sanit.* 2003;17:474-82.
 17. 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.
 18. 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.
 19. 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.
 20. Perez-Hoyos S, Aviño MJ, Hernández-Aguado I, González J, Ruiz I. Tiempo libre de SIDA y supervivencia de una cohorte de usuarios de drogas por vía parenteral seroconvertidores frente al VIH. *Gac Sanit.* 1998;13:337-45.
 21. Perez-Hoyos S. Análisis de Supervivencia. Cuadernos de salud pública i administració sanitaria. Valencia: Institut Valencià d'Estudis en Salut Pública; 1997.
 22. Diccicco TJ, Efron B. Bootstrap Confidence Intervals. *Statistical Science* 1996;11:189-228.
 23. Lamarca R, Alonso J, Gomez 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;53:M337-43.
 24. S-Plus 6 for windows. Seattle: Insightful Corporation, 2001.
 25. Stata Corp. Stata Statistical Software: Release 8.0. College Station: Stata Corporation, 2003.
 26. Perez-Hoyos S, Aviño MJ, Hernández-Aguado I, González J. Imputación del momento de la seroconversión para una cohorte de seroprevalentes al VIH. *Revista Española Salud Pública* 1998;72,Suppl:170-1.
 27. Cole SR, Li R, Anastos K, Detels R, Young M, Chmiel JS et al. Accounting for leadtime in cohort studies: evaluating when to initiate HIV therapies. *Stat Med* 2004. (en prensa).

