

COLABORACIÓN ESPECIAL**ESTUDIOS LONGITUDINALES: CONCEPTO Y PARTICULARIDADES****Miguel Delgado Rodríguez (1) y Javier Llorca Díaz (2)**

(1) Universidad de Jaén

(2) Universidad de Cantabria

RESUMEN

En la presente revisión se analiza el concepto de estudio longitudinal. Los libros de texto de epidemiología actuales en general no lo definen, mientras que sí lo hacen los tratados de estadística. Se habla más de «datos longitudinales» que de «estudios longitudinales». El estudio longitudinal implica la existencia de medidas repetidas (más de dos) a lo largo de un seguimiento. Sería pues un subtipo de estudio de cohortes que, a diferencia de los de tipo tabla de vida, permite inferencias a nivel individual y analizar cambios en diferentes variables (exposiciones y efectos) y transiciones entre diferentes estados de salud. Las particularidades de este tipo de diseño hace que se tenga que prestar atención especial al control de calidad durante su ejecución, a los abandonos durante el seguimiento, y a los datos perdidos en algunas de las mediciones. El análisis debe tener en cuenta las medidas repetidas y esto es lo que confiere finalmente a un estudio su carácter longitudinal.

Palabras clave: Estudios longitudinales. Estudios de cohortes.

ABSTRACT**Longitudinal Studies: concept and particularities**

In this review the definition of «longitudinal study» is analysed. Most current textbooks on epidemiology do not define a longitudinal study, whereas statistical textbooks do. It is more common to talk about longitudinal data than about longitudinal studies. A longitudinal study implies the existence of repeated measurements (more than two) across follow-up. According to these ideas, a longitudinal study can be considered a subtype of cohort study that, in contrast with life-table cohort studies, allows inference to the subject level, to analyze changes in variables (exposures and outcomes) and transitions among different health states. The characteristics of this design force to paid special attention to quality control during data collection, losses during follow-up, and missing data in some measurements. The statistical analysis should take repeated measures into account, and it is what finally gives the longitudinal character to a study with repeated measurements.

Key words: Longitudinal studies. Cohort studies.

¿QUÉ ES UN ESTUDIO LONGITUDINAL?

La discusión sobre el significado del término longitudinal la resumió Chin¹ en 1989: para los epidemiólogos es sinónimo de estudio de cohortes o seguimiento, mientras que para algunos estadísticos implica medicio-

nes repetidas². Él mismo decide no definir el término longitudinal, por ser difícil encontrar un concepto aceptable para todos, y opta por considerarlo equivalente a «seguimiento», el pensamiento más habitual para los profesionales de la época.

El estudio longitudinal en la epidemiología

En la década de 1980 era muy común utilizar el término longitudinal para separar simplemente la causa del efecto. En oposición al término transversal. Miettinen lo

Correspondencia:
Miguel Delgado Rodríguez
Universidad de Jaén
Edificio B-3
23071-Jaén
Correo electrónico: mdlgado@ujaen.es

define como un estudio cuya base es la experiencia de la población a lo largo del tiempo (al contrario que un corte de la población)³. Consistente con esta idea, Rothman, en su texto de 1986, indica que la palabra longitudinal denota la existencia de un intervalo de tiempo entre la exposición y el comienzo de la enfermedad⁴. Bajo esta acepción, el estudio de casos y controles, que es una estrategia de muestreo para representar esa experiencia de la población a lo largo del tiempo (sobre todo bajo las ideas de Miettinen), también sería un estudio longitudinal. De igual forma es coincidente con esta idea Abramson, que además diferencia los estudios descriptivos longitudinales (estudios de cambio) de los analíticos longitudinales, que incluyen dentro a los estudios de casos y controles⁵. Kleinbaum et al⁶ así mismo definen el término longitudinal en oposición al transversal pero, con un matiz algo diferente, hablan de «experiencia longitudinal» de una población (frente a «experiencia transversal») e implica para ellos la realización de al menos dos series de observaciones a lo largo de un periodo de seguimiento. Excluyen estos últimos autores a los estudios de casos y controles. Kahn y Sempos⁷ tampoco tienen un epígrafe para estos estudios y en el índice por palabras clave, en la entrada «estudio longitudinal» se lee «ver estudio prospectivo».

Esto lo refleja el Diccionario de Epidemiología dirigido por Last, que considera el término «estudio longitudinal» como sinónimo de estudio de cohortes o estudio de seguimiento⁸. En el texto clásico de Breslow y Day sobre los estudios de cohortes, el término longitudinal se considera equivalente al de cohorte y se utiliza de manera indistinta⁹. No obstante, Cook y Ware¹⁰ definieron el estudio longitudinal como aquél en el que un mismo individuo es observado en más de una ocasión y lo diferenciaron de los estudios de seguimiento, en los que los individuos son seguidos hasta la ocurrencia de un suceso tal como la muerte o una enfermedad (aunque este suceso es ya la segunda observación).

A partir de 1990, varios textos consideran el término longitudinal equivalente a otros nombres, aunque la mayoría lo omiten. Reflejo de ello es el libro codirigido por Rothman y Greenland¹¹, en el que no hay un apartado específico para los estudios longitudinales dentro de los capítulos dedicados al diseño, y también coincide con esta tendencia la Enciclopedia de Métodos Epidemiológicos¹², que no ofrece una entrada específica para este tipo de estudios. La cuarta edición del Diccionario de Epidemiología de Last reproduce su entrada de ediciones anteriores¹³. Gordis¹⁴ lo considera sinónimo de estudio de cohortes prospectivo concurrente. Aday¹⁵ sigue parcialmente las ideas de Abramson, ya mencionadas, y diferencia los estudios descriptivos (varios estudios transversales secuenciados en el tiempo) de los analíticos, entre los que están los estudios de cohortes prospectivos o longitudinales.

En otros terrenos de la medicina clínica, el sentido de longitudinal se considera opuesto al transversal y se equipara a cohorte, con frecuencia prospectiva. Esto se comprueba, por ejemplo, en las publicaciones centradas en el campo de la menopausia¹⁶.

El estudio longitudinal en la estadística

Aquí las ideas están mucho más claras: un estudio longitudinal es el que implica más de dos mediciones a lo largo de un seguimiento; deben ser más de dos, ya que todo estudio de cohortes tiene este número de mediciones, la del principio y la del final del seguimiento. Éste es el concepto existente en el texto mencionado de Goldstein de 1979. En ese mismo año Rosner era explícito al indicar que los datos longitudinales implican mediciones repetidas en los sujetos a lo largo del tiempo, proponiendo un nuevo procedimiento de análisis para ese tipo de datos¹⁷. Desde ese momento, los artículos en revistas de estadística (por ejemplo¹⁸⁻²²) y textos²³⁻²⁵ son consistentes en el mismo concepto.

Dos obras de referencia en epidemiología, aunque no definen en el apartado correspondiente los estudios longitudinales, coinciden con la noción estadística prevaleciente. En el libro codirigido por Rothman y Greenland, dentro del capítulo Introducción a la modelación de la regresión, el propio Greenland afirma que los datos longitudinales son las mediciones repetidas en los sujetos a lo largo de un periodo de tiempo y que se pueden realizar para exposiciones dependientes del tiempo (por ejemplo, tabaquismo, consumo de alcohol, dieta o presión arterial) o resultados recurrentes (por ejemplo, dolor, alergia, depresión, etc.)²⁶. En la Enciclopedia de Métodos Epidemiológicos, la entrada del «tamaño de muestra» incluye un apartado de «estudios longitudinales» en el que se ofrece la misma información proporcionada por Greenland²⁷.

Conviene matizar que la visión estadística de «estudio longitudinal» parte de un análisis de datos particular (tener en cuenta las medidas repetidas) y que lo mismo sería aplicable a los estudios de intervención, que también tienen seguimiento²⁸.

Para finalizar este apartado, en el número monográfico de *Epidemiologic Reviews* dedicado a los estudios de cohortes, Tager, en su artículo centrado en la variable resultado de los estudios de cohortes, clasifica de manera amplia los estudios de cohortes en dos grandes grupos, de «tabla de vida» y «longitudinales»²⁹, aclarando que esta clasificación es algo «artificial». Los primeros son los convencionales, en los que el resultado es una variable discreta, la exposición y la población-tiempo se resumen, se estiman incidencias y la medida principal es el riesgo relativo. Los segundos incorporan un análisis diferente, aprovechándose de las mediciones repetidas en los sujetos a lo largo del tiempo, permitiendo una inferencia, además de poblacional, a nivel individual en los cambios de un proceso a lo largo del tiempo o en las transiciones entre diferentes estados de salud y la enfermedad.

Las ideas anteriores denotan que en epidemiología hay una tendencia a esquivar el concepto de estudio longitudinal. No obstante, resumiendo las ideas comentadas con anterioridad, la noción de estudio longitudinal hace referencia al estudio de cohortes en el que se realizan más de dos mediciones a lo largo del tiempo y en el que se realiza un análisis que tiene en cuenta las diferentes medidas. Los tres elementos claves son: seguimiento, más de dos medidas y un análisis que las tenga en cuenta. Esto puede hacerse de manera prospectiva o retrospectiva, y el estudio puede ser de observación o de intervención.

DIFERENCIAS ENTRE LOS ESTUDIOS LONGITUDINALES Y LOS DE TABLA DE VIDA

En la tabla 1 se resumen las características generales de ambos tipos de diseños. Los estudios de cohortes tipo tabla de vida son los que resumen la exposición y la enfermedad en los grupos que se comparan, por ejemplo, frecuencia del cáncer de pulmón en fumadores y no fumadores. La inferencia proporcionada por estos estudios se refiere a medias poblacionales. Llevan implícita la asunción de que la exposición actúa de manera constante en el tiempo y tiene un efecto también constante por unidad de tiempo a lo largo del seguimiento, y sólo pueden proporcionar una inferencia limitada sobre la dependencia en el tiempo de las asociaciones entre la exposición y el efecto²⁹. Un ejemplo de este tipo de estudio puede ser el *Nurses' Health Study*, con más 120.000 enfermeras en 11 estados de EE.UU., en el que la exposición acumulada a los anticonceptivos orales se valoró como factor de riesgo del cáncer de mama³⁰.

Los estudios longitudinales en cualquier momento se pueden comportar como los estudios tipo tabla de vida. Pueden además realizar inferencias también a nivel individual, valoran el cambio de procesos a lo lar-

Tabla 1

Características de los estudios de cohortes de tabla de vida (TV) y longitudinales (L).

Característica	Comentario
Selección de los sujetos TV: nivel de exposición al inicio de la cohorte L: Además, puede estar no relacionado con el nivel de una exposición única y explícita (estudio de «historia natural»)	Cada grupo de exposición debe ser representativo de la población Lo mismo que los TV Los sujetos deben ser representativos de la población diana. La población diana puede ser sujetos con problemas de salud conocidos
Tiempo TV: lo resumen L: lo resumen y pueden evaluar ciertos efectos explícitamente	Estiman incidencias medias durante el seguimiento, por edades, o épocas Además, pueden separar los efectos de cohorte y edad, modelar las asociaciones relacionadas con el tiempo entre las observaciones a lo largo del seguimiento, y la correlación del efecto sobre varios intervalos de tiempo.
Exposición TV: Basal, resumirla o categorizarla LV: Además, pueden formular explícitamente la dependencia en el tiempo del efecto	Categoriza los sujetos según su exposición inicial o según medidas acumuladas de la misma (medidas durante el seguimiento) y calcular exposiciones promedio durante todo el periodo o fracciones. Además, evalúa los efectos de la nueva exposición durante el seguimiento sobre los acumulados hasta el inicio del mismo. Distingue efectos sobre los individuos de efectos promedio poblacionales. Evalúa la asociación exposición-efecto en cualquier momento de la historia de los sujetos, y la exposición actual sobre perfiles de exposición pasados.
Resultado o efecto TV: Primera ocurrencia del mismo L: Además, tasa de cambio (crecer, declinar) y transiciones entre estados de salud. Correlación del efecto en el tiempo.	Malo para efectos "continuos", que se categorizan perdiéndose información. Las mediciones: tasa, riesgo, y sus cocientes. Además, describe la historia natural de un proceso, de estadios precursores, y elimina la restricción de la primera ocurrencia de un proceso de salud.
Inferencia TV: medias de grupo o poblacionales L: Además, inferencia a nivel individual e historia natural del proceso	Sólo se establecen conclusiones a nivel de grupo, no a nivel individual. Lo ya mencionado.

Fuente: ²⁹

go del tiempo y las transiciones entre los distintos estados de salud-enfermedad. Un ejemplo, al igual que el Nurses' Health Study, prolífico en publicaciones, es el MACS (*Multicenter AIDS Cohort Study*), en el que se reclutaron casi 5000 varones en cuatro ciudades estadounidenses³¹. Cuando se miden cambios de una variable en función del tiempo, en el diseño hay que tener en cuenta la duración del seguimiento y el espaciamiento de las mediciones³².

PARTICULARIDADES DE LOS ESTUDIOS LONGITUDINALES

Al realizarse mediciones a lo largo del tiempo, el **control de calidad** juega un papel esencial. Hay que garantizar que todas las mediciones se realicen en el momento oportuno y con técnicas normalizadas. La larga duración de algunos estudios obliga a prestar una atención especial al cambio de personal, al deterioro de los equipos, al cambio de tec-

nologías y a las inconsistencias de las respuestas de los participantes a lo largo del tiempo³³.

Existe una **mayor probabilidad de abandono** durante el seguimiento. Los factores que intervienen en ello son varios³⁴:

- La definición de una población según un criterio inestable. Por ejemplo, vivir en un área geográfica concreta puede motivar que participantes con cambios de domicilio no sean elegibles en fases posteriores.
- Será mayor cuando en los respondedores que no se contactan una vez, no se vuelve a intentar establecer el contacto en fases posteriores del seguimiento.
- El objeto del estudio influye; por ejemplo, en un estudio de ciencia política los no interesados en política abandonarán más.
- La cantidad de atención personal dedicada a los respondedores. Las entrevistas telefónicas y por carta son menos personales que las que se realizan cara a cara, y no se aprovechan para fortalecer los vínculos con el estudio.
- El tiempo invertido por el respondedor en satisfacer la demanda de información de los investigadores. Cuanto mayor sea, mayor será la frecuencia de abandonos.
- La frecuencia del contacto puede también influir, aunque no todos están de acuerdo. Hay estudios que han documentado que un exceso de contactos perjudica el seguimiento, mientras que otros, o no han encontrado relación o ésta es negativa.

Para evitar los abandonos conviene establecer estrategias con el fin de retener y rastrear a los miembros participantes. Debe valorarse al comienzo la voluntad de partici-

pación e informar de lo que se espera de los participantes. Hay que establecer puentes de unión con los participantes mediante el envío de cartas de felicitación, actualizaciones del estudio, etc. La frecuencia de contacto debe ser regular. El personal del estudio debe ser entusiasta, con facilidad de comunicación, que responda rápida y adecuadamente a los problemas de los participantes y adaptable a sus necesidades. No hay que desdeñar dar incentivos que motiven la continuación en el estudio³⁵.

En tercer lugar, otro problema de mayor calibre frente a otros estudios de cohortes es la **existencia de datos perdidos**. Si se requiere que un participante tenga todas las mediciones hechas, puede producir un problema similar al de los abandonos durante el seguimiento. Para ello se han desarrollado técnicas de imputación de valores perdidos y, aunque se ha sugerido que pueden no ser necesarias si se aplican las ecuaciones de estimación generalizadas (análisis GEE)³⁶, se ha comprobado que otros procedimientos dan mejores resultados, incluso cuando las pérdidas son completamente aleatorias³⁷. Con frecuencia las pérdidas de información son diferenciales y se pierden más mediciones en los pacientes con un peor nivel de salud. Se recomienda en estos casos que la imputación de datos se haga teniendo en cuenta los datos ya existentes del propio individuo al que le faltan³⁸.

Análisis

En el análisis de los estudios longitudinales es posible tratar covariables dependientes del tiempo que pueden a la vez influir sobre la exposición en estudio y ser influidas por ella (variables que se comportan simultáneamente como confundidoras e intermedias entre exposición y efecto). También, de manera similar, permite controlar resultados recurrentes que pueden actuar sobre la exposición y ser ocasionados por ella (se comportan a la vez como confundidores y efectos)²⁶.

El análisis longitudinal se puede utilizar cuando existen mediciones del efecto y/o de la exposición en diferentes momentos del tiempo. Supongamos que la relación entre una variable dependiente Y es función de una variable X que cambia en el tiempo (témpero-dependiente) y otra Z que es estable en el tiempo (témpero-independiente), que se estudian N sujetos en K momentos del tiempo, lo cual se expresa según la ecuación siguiente¹⁷:

$$Y_{it} = \beta x_{it} + z_i \alpha + \varepsilon_{it}$$

donde el subíndice i hace referencia al individuo, el t al momento del tiempo y ε es un término de error (Z no cambia al ser estable y por eso tiene un solo subíndice). La existencia de varias mediciones permite estimar el coeficiente β sin necesidad de conocer el valor de la variable estable, al realizar una regresión de la diferencia en el efecto (Y) sobre la diferencia de valores de las variables independientes:

$$Y_{it} - Y_{i1} = \beta(x_{it} - x_{i1}) + \alpha(z_i - z_i) + \varepsilon_{it} - \varepsilon_{i1} = \beta(x_{it} - x_{i1}) + \varepsilon_{it} - \varepsilon_{i1}$$

Es decir, no es necesario saber el valor de las variables témpero-independientes (o estables) en el tiempo. Esto supone una ventaja sobre otros análisis, en los que hay que conocer dichas variables. El modelo anterior es fácilmente generalizable a un vector multivariable de factores cambiantes en el tiempo.

El análisis longitudinal se realiza dentro del contexto de los modelos lineales generalizados y tiene dos objetivos: adoptar las herramientas convencionales de regresión, en las que se relaciona el efecto con las diferentes exposiciones y tener en cuenta la correlación de las medidas entre sujetos. Este último aspecto es muy importante. Suponga que analiza el efecto del crecimiento sobre la presión arterial; los valores de presión arterial de un sujeto en los distintos exámenes realizados dependen del valor inicial o basal y por ello hay que tenerlo en cuenta.

Por ejemplo, el análisis longitudinal se podría realizar en una cohorte infantil en la que se valora como exposición principal la deficiencia de vitamina A (que puede cambiar en el tiempo) sobre el riesgo de infección (que puede ser múltiple a lo largo del tiempo), controlando la influencia de la edad, el peso y la talla (variables témpero-dependientes). El análisis longitudinal se puede clasificar en tres grandes grupos³⁹:

- a) Modelos marginales: combinan las diferentes mediciones (que son cortes en el tiempo) de la prevalencia de la exposición para obtener una prevalencia media u otra medida resumen de la exposición a lo largo del tiempo, y la relaciona con la frecuencia de la enfermedad. El elemento longitudinal es la edad o la duración del seguimiento en el análisis de regresión. Los coeficientes de este tipo de modelos se transforman en una razón de prevalencias poblacionales; en el ejemplo de la vitamina A y la infección sería la prevalencia de infección en niños con deficiencia de vitamina A dividida por la prevalencia de infección en niños sin déficit de vitamina A.
- b) Los modelos de transición realizan una regresión del resultado presente sobre valores pasados y sobre las exposiciones pasadas y presentes. Un ejemplo de ellos son los modelos de Markov. Los coeficientes del modelo se transforman directamente en un cociente de incidencias, esto es, en RRs; en el ejemplo sería el RR del déficit de vitamina A sobre la infección.
- c) Los modelos de efectos aleatorios permiten que cada individuo tenga parámetros de regresión únicos, y existen procedimientos para resultados normalizados, binarios y datos de persona-tiempo. Los coeficientes del modelo se transforman en una odds ratio

referida al individuo, que se asume constante en toda la población; en el ejemplo sería la odds de infección en un niño con déficit de vitamina A frente a la odds de infección en el mismo niño sin deficiencia de vitamina A.

Los modelos lineales, logístico, de Poisson y muchos análisis de supervivencia se pueden considerar casos particulares de modelos lineales generalizados. Hay procedimientos que permiten las entradas tardías o en momentos diferentes y de manera desigual en la observación de una cohorte.

Además de los modelos paramétricos indicados en el párrafo anterior, es posible el análisis mediante métodos no paramétricos; por ejemplo, el uso de análisis funcional con *splines* ha sido revisado recientemente^{40,41}.

Se han mencionado varios textos específicos sobre análisis de datos longitudinales. Uno de ellos ofrece incluso ejemplos con las rutinas a escribir para realizar correctamente el análisis usando diferentes paquetes estadísticos convencionales (STATA, SAS, SPSS)²⁵.

BIBLIOGRAFÍA

- Chinn S. Longitudinal studies: objectives and ethical considerations. *Rev Epidém Santé Publ* 1989; 37: 417-29.
- Goldstein H. The design and analysis of longitudinal studies. Londres: Academic Press, 1979.
- Miettinen OS. Theoretical epidemiology: principles of occurrence research in medicine. Nueva York: Wiley; 1985.
- Rothman KJ. Modern epidemiology. Boston: Little-Brown; 1986.
- Abramson JH. Classification of epidemiologic research. *J Clin Epidemiol* 1989; 42: 819-20.
- Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic research. Belmont: Lifetime Learning Publications; 1982.
- Kahn HA, Sempos CT. Statistical methods in epidemiology. Nueva York: Oxford University Press; 1989.
- Last JM. A dictionary of epidemiology. 2ª ed. Nueva York: Oxford Univ. Press; 1988.
- Beslow NE, Day NE. Statistical methods for cancer research. Volume II- The design and analysis of cohort studies. Lión: IARC Scientific Publications; 1987.
- Cook NR, Ware JH. Design and analysis methods for longitudinal research. *Annu Rev Public Health* 1983;4:1-24.
- Rothman KJ, Greenland S, editores. Modern Epidemiology. 2ª ed. Filadelfia: Lippincott-Raven; 1998.
- Gail MH, Benichou J, editores. Encyclopedia of epidemiologic methods. Chichester: Wiley; 2000.
- Last JM. A dictionary of epidemiology. 4ª ed. Nueva York: Oxford Univ. Press; 2000.
- Gordis L. Epidemiology. Filadelfia: Saunders; 1996. p. 119.
- Aday LA. Designing and conducting health surveys. 2ª ed. San Francisco: Jossey-Bass Publishers; 1996. p. 29-30.
- Collins A, Landgren B-M. Longitudinal research on the menopause: methodological challenges. *Acta Obstet Gynecol Scand* 2002; 81: 579-80.
- Rosner B. The análisis of longitudinal data in epidemiologic studies. *J Chron Dis* 1979; 32: 163-73.
- Louis TA. General methods for analysing repeated measures. *Stat Med* 1988; 7: 29-45.
- Ware JH, Lipsitz S. Issues in the analysis of repeated categorical outcomes. *Stat Med* 1988; 7: 95-107.
- Landis JR, Miller ME. Some general methods for the analysis of categorical data in longitudinal studies. *Stat Med* 1988; 7: 29-45.
- Zeger SL, Liang KY. An overview of methods for the analysis of longitudinal data. *Stat Med* 1992; 11: 1825-39.
- Carlin JB, Wolfe R, Coffey C, Patton GC. Analysis of binary outcomes in longitudinal studies using weighted estimating equations and discrete-time survival methods: prevalence and incidence of smoking in an adolescent cohort. *Stat Med* 1999; 18: 2655-79.

23. Dwyer JH, Feinleib M, Lippert P, Hoffmeister H. Statistical models for longitudinal studies of health. Nueva York: Oxford University Press; 1992.
24. Diggle PJ, Heagerty P, Liang KY, Zeger SL. Analysis of longitudinal data. 2ª ed. Oxford: Oxford University Press; 2002.
25. Twisk JWR. Applied longitudinal data analysis for epidemiology. A practical guide. Cambridge: Cambridge University Press; 2003.
26. Greenland S. Introduction to regression modeling. En: Rothman KJ, Greenland S, editores. Modern Epidemiology. 2ª ed. Filadelfia: Lippincott-Raven; 1998. p. 359-432.
27. Liu G. Sample size for epidemiologic studies. En: Gail MH, Benichou J, editores. Encyclopedia of epidemiologic methods. Chichester: Wiley; 2000. p. 787-788.
28. Galbraith S, Marschner IC. Guidelines for the design of clinical trials with longitudinal outcomes. *Controlled Clin Trials* 2002; 23: 257-73.
29. Tager IB. Outcomes of cohort studies. *Epidemiol Rev* 1998; 20: 15-28.
30. Lipnick RJ, Buring JE, Hennekens CH, Rosner B, Willett W, Bain C, et al. Oral contraceptives and breast cancer. A prospective cohort study. *JAMA* 1986; 255: 58-61.
31. Muñoz A, Kirby AJ, He YD. Long-term survivors with HIV-infection: incubation period and longitudinal patterns of CD4+ lymphocytes. *J Acquir Immune Defic Syndr Hum Retrovirol* 1995; 8: 496-505.
32. Schlesselman KK. Planning a longitudinal study. II. Frequency of measurement and study duration. *J Chron Dis* 1973; 26: 561-70.
33. Whitney CW, Lind BK, Wahl PW. Quality assurance and quality control in longitudinal studies. *Epidemiol Rev* 1998; 20: 71-80.
34. Deeg DJH, van Tilburg T, Smit JH, de Leeuw ED. Attrition in the Longitudinal Aging Study Amsterdam: The effect of differential inclusion in side studies. *J Clin Epidemiol* 2002; 55: 319-28.
35. Hunt JR, White E. Retaining and tracking cohort study members. *Epidemiol Rev* 1998; 20: 57-70.
36. Twisk J, de Vente W. Attrition in longitudinal studies: how to deal with missing data. *J Clin Epidemiol* 2002; 55: 329-37.
37. Touloumi G, Babiker AG, Pocock SJ, Darbyshire JH. Impact of missing data due to drop-outs on estimators for rates of change in longitudinal studies: a simulation study. *Stat Med* 2001; 20: 3715-28.
38. Engels JM, Diehr P. Imputation of missing longitudinal data: a comparison of methods. *J Clin Epidemiol* 2003; 56: 968-76.
39. Samet JM, Muñoz A. Evolution of the cohort studies. *Epidemiol Rev* 1998; 20: 1-14.
40. Guo W. Functional data analysis in longitudinal settings using smoothing splines. *Stat Meth Med Res* 2004; 13: 49-62.
41. Zhang H. Mixed effects multivariate adaptive splines model for the analysis of longitudinal and growth curve data. *Stat Meth Med Res* 2004; 13: 63-82.