1. Introduction

Prospective epidemiological studies follow a cohort of individuals until disease occurrence or death, or until the scheduled end of follow-up. The data for each participant in a cohort study must include three variables: time of entry, time of exit and status at exit. In the R language (R Development Core Team 2010), working with such data is made easy by the Surv class in the survival package (Therneau and Lumley 2010). The survival package also provides modelling functions that use Surv objects as outcome variables, and use the standard S syntax for model formulae. In epidemiological studies with long-term follow-up, there may be more than one time scale of interest. If follow-up time is measured in decades, for example, any analysis of disease risk must take account of the impact of the ageing of the population. In this case, “calendar time” and “age” are both time scales of interest. A time scale may also measure the time elapsed since an important event, such as study entry, first exposure to a risk factor or beginning of treatment.

**1. Introducció**

**Estudis epidemiològics possibles segueixen una cohort d’individus fins a l’aparició de la malaltia o la mort o fins al final previst del seguiment. Les dades de cada participant en un estudi de cohort han d’incloure tres variables: hora d’entrada, hora de sortida i estat de sortida. En el llenguatge R (R Development Core Team 2010), la classe Surv del paquet de supervivència facilita el treball amb aquestes dades (Therneau i Lumley 2010). El paquet de supervivència també proporciona funcions de modelatge que utilitzen objectes Surv com a variables de resultat i que utilitzen la sintaxi S estàndard per a fórmules de model. En estudis epidemiològics amb un seguiment a llarg termini, pot haver-hi més d’una escala d’interès. Si es mesura el temps de seguiment en dècades, per exemple, qualsevol anàlisi del risc de malaltia ha de tenir en compte l'impacte de l'envelliment de la població. En aquest cas, "temps natural" i "edat" són escales d'interès. Una escala de temps també pot mesurar el temps transcorregut des d’un esdeveniment important, com ara l’entrada d’estudi, la primera exposició a un factor de risc o l’inici del tractament.**

The statistical problem of accounting for multiple time scales can be addressed using tools developed in the field of demography in the 19th century. Such tools are also used in descriptive epidemiology to separate time trends in chronic disease incidence rates into age, period and cohort effects. The increasing number of large population-based cohort studies with long-term follow-up has created a demand for these tools in analytical epidemiology. The Epi (Carstensen, Plummer, Laara, and Hills 2010) package contains functions and classes to facilitate the analysis of epidemiological studies in R. Among these, the Lexis class was designed to simplify the analysis of long term follow-up studies by tracking follow-up time on multiple time scales. It also accounts for many possible disease outcomes by having a status variable that is not a simple binary indicator (alive/dead or healthy/diseased) but may take multiple values.

**El problema estadístic de comptabilitzar escales de temps múltiples es pot abordar mitjançant eines desenvolupades en el camp de la demografia al segle XIX. Aquestes eines també s’utilitzen en l’epidemiologia descriptiva per separar les tendències del temps en les taxes d’incidència de malalties cròniques en edats, períodes i efectes de cohort. El nombre creixent d’estudis de cohorts basats en població amb un seguiment a llarg termini ha creat una demanda d’aquestes eines en epidemiologia analítica. El paquet Epi (Carstensen, Plummer, Laara i Hills 2010) conté funcions i classes per facilitar l’anàlisi d’estudis epidemiològics en R. Entre aquests, la classe Lexis va ser dissenyada per simplificar l’anàlisi d’estudis de seguiment a llarg termini mitjançant un seguiment del seguiment. temps màxim a diverses escales de temps. També compta amb molts possibles resultats de malaltia tenint una variable d’estat que no és un simple indicador binari (viu / mort o saludable / malalt), però pot prendre múltiples valors.**

2. Lexis diagrams and Lexis objects

Figure 1 shows a simple example of a Lexis diagram, named after the demographer Wilhelm Lexis (1837–1914). Each line in a Lexis diagram represents the follow-up of a single individual from entry to exit on two time scales: age and calendar time. Both time scales are measured in the same units (years) so that the follow-up traces a line at 45 degrees. Exit status is denoted by a circle for the 4 subjects who experienced a disease event. The other subjects are disease-free at the end of follow-up. The follow-up line of an individual in a Lexis diagram is defined by his or her entry time on the two time scales of interest (age and calendar time) and the duration of follow up. The Lexis class formalises this representation of follow-up in an R object. Lexis objects are not limited to 2 time scales, but allow follow-up time to be tracked on an arbitrary number of time scales. The only restriction is that time scales must be measured in the same units. To illustrate Lexis objects, we use a cohort of nickel smelting workers in South Wales (Doll, Mathews, and Morgan 1977), which was included as an example by Breslow and Day (1987). The data from this study are contained in the data set nickel in the Epi package .

Follow-up time is defined by entry, a named list of entry times on each time scale, and exit, another named list of exit times. Since duration of follow-up is the same on all time scales, it is only necessary to define the exit time on one of them, in this case age. The three time scales in this Lexis object are:  period, representing calendar time. Date of entry is calculated as date of birth (dob) plus age at entry (agein).  age, representing participant’s age.  tfe, representing time since first employment, which is used as a proxy for first exposure. Entry time on this scale is calculated as the difference between age at entry and age at first employment (age1st). The exit.status argument gives the individual’s status at the end of follow-up. Since this is a study of mortality, the exit status is the cause of death according to the Seventh Revision of the International Classification of Diseases (ICD, World Health Organization 1957). For individuals who were still alive at the end of follow-up, the exit status is 0. The data argument gives a data frame that is the source of all the variables used to define entry time, exit time, status, and so on. The Lexis function transforms this data frame into a Lexis object.

**2. Esquemes de lèxic i objectes de lèxic**

**La figura 1 mostra un exemple senzill d’un diagrama de Lèxic, anomenat així pel demògraf Wilhelm Lexis (1837–1914). Cada línia d’un diagrama de Lèxic representa el seguiment d’un sol individu d’entrada a sortida en dues escales de temps: edat i temps de calendari. Les dues escales de temps es mesuren a les mateixes unitats (anys) de manera que el seguiment traça una línia a 45 graus. L'estat de sortida es marca amb un cercle per als quatre subjectes que van experimentar un esdeveniment de la malaltia. Els altres subjectes no tenen malalties al final del seguiment. La línia de seguiment d’un individu en un diagrama de Lèxic es defineix pel seu temps d’entrada en les dues escales d’interès (edat i temps del calendari) i la durada del seguiment. La classe Lexis formalitza aquesta representació del seguiment en un objecte R. Els objectes de lèxic no es limiten a 2 escales de temps, sinó que permeten fer el seguiment del temps de seguiment en un nombre arbitrari d'escales de temps. L’única restricció és que les escales de temps s’han de mesurar a les mateixes unitats. Per il·lustrar objectes Lexis, es fa servir una cohort de treballadors de fosa de níquel a Gal·les del Sud (Doll, Mathews i Morgan 1977), que va ser inclosa com a exemple per Breslow i Day (1987). Les dades d’aquest estudi es troben al níquel del conjunt de dades del paquet Epi.**

**El temps de seguiment es defineix per l'entrada, una llista de temps d'entrada indicada a cada escala, i sortida, una altra llista de temps de sortida. Com que la durada del seguiment és la mateixa en totes les escales de temps, només cal definir l’hora de sortida en una d’elles, en aquest cas l’edat. Les tres escales de temps d’aquest objecte Lexis són: ˆ període, que representa el temps del calendari. La data d’entrada es calcula com a data de naixement (dob) més l’edat a l’entrada (agein). ˆ edat, representant l’edat del participant. ˆ tfe, que representa el temps des de la primera ocupació, que s’utilitza com a proxy per a la primera exposició. El temps d’entrada en aquesta escala es calcula com la diferència entre l’edat a l’entrada i l’edat al primer treball (primera edat). L’argument exit.status proporciona l’estat de l’individu al final del seguiment. Com que es tracta d’un estudi de la mortalitat, l’estat de sortida és la causa de la mort segons la Setena Revisió de la Classificació Internacional de Malalties (ICD, Organització Mundial de la Salut, 1957). Per a individus que encara eren vius al final del seguiment, l’estat de sortida és 0. L’argument de dades proporciona un marc de dades que és l’origen de totes les variables que s’utilitzen per definir el temps d’entrada, el temps de sortida, l’estat, etc. La funció Lexis transforma aquest fotograma de dades en un objecte Lexis.**

2.1. Plotting Lexis objects Not surprisingly, the plot method for Lexis objects plots a Lexis diagram. Figure 2 shows a Lexis diagram for the nickel smelters cohort. The points method is used to annotate the Lexis diagram with the times of all deaths from lung cancer (ICD code 162 or 163):

This example also illustrates the extractor function status, which returns the status at the beginning or (by default) end of each follow-up period. Other extractor functions dur, entry, and exit return respectively the duration of follow-up and the entry and exit times on any given time scale. By default, the plot method chooses the first two time scales of the Lexis object to plot. Other time scales may be chosen using the argument time.scale. A single time scale may be specified:

This produces Figure 3, in which the y-axis is the unique id number and all history lines are horizontal. Such plots may reveal important features of the data. For example, Figure 3 shows that, on the “tfe” time scale, there are many late entries into the study with some participants entering over 20 years after first employment. Due to the method of selection for this cohort, no participant came under observation until 1934, even if they had been working many years in the smelting industry (Breslow and Day 1987).

**2.1. Plotografia d'objectes Lexis No sorprèn que el mètode de traça per a objectes Lexis traça un diagrama de Lexis. La figura 2 mostra un esquema lèxic per a la cohort de les fundidores de níquel. El mètode de punts s’utilitza per anotar el diagrama de Lexis amb els temps de totes les morts per càncer de pulmó (codis ICD 162 o 163):**

**Aquest exemple també il·lustra l'estat de la funció extractora, que retorna l'estat al començament o (per defecte) de cada període de seguiment. Altres funcions d'extractor dur, entrada i sortida de retorn respectivament, la durada del seguiment i els temps d'entrada i sortida en qualsevol escala de temps determinada. Per defecte, el mètode de traça tria les dues primeres escales de temps de l’objecte Lexis a la trama. Es poden triar altres escales de temps amb l'argument time.scale. Es pot especificar una escala de temps única:**

**Es produeix la figura 3, en què l’eix y és el número d’identificació únic i totes les línies d’historial són horitzontals. Aquests complots poden revelar característiques importants de les dades. Per exemple, la figura 3 mostra que, a l’escala de temps “tfe”, hi ha moltes entrades tardanes a l’estudi amb alguns participants que entren més de 20 anys després de la primera ocupació. A causa del mètode de selecció d'aquesta cohort, cap participant va estar sota observació fins al 1934, fins i tot si havien treballat molts anys en la indústria de la fundició (Breslow i el dia 1987).**

2.2. Structure of a Lexis object Lexis objects inherit from the class data.frame. The Lexis object contains all the variables in the source data frame that was given as the data argument to the Lexis function. In addition, a variable is created for each time scale as well as a four variables with reserved names starting with lex. (lex.dur, lex.Cst, lex.Xst, and lex.id).

In this example, the first 3 variables (period, age, and tfe) show the entry times on the 3 time scales. The variable lex.dur shows the duration, lex.Cst and lex.Xst show the current status and exit status respectively, and lex.id shows the unique identifier for each individual.

**2.2. Estructura d'un objecte Lexis Els objectes Lexis s'hereten de la classe data.frame. L’objecte Lexis conté totes les variables del marc de dades font que es van donar com a argument de dades a la funció Lexis. A més, es crea una variable per a cada escala de temps, així com quatre variables amb noms reservats que comencen amb lex. (lex.dur, lex.Cst, lex.Xst i lex.id).**

**En aquest exemple, les tres primeres variables (període, edat i tfe) mostren els temps d’entrada a les tres escales de temps. La variable lex.dur mostra la durada, lex.Cst i lex.Xst mostren l'estat actual i l'estat de sortida respectivament, i lex.id mostra l'identificador únic per a cada individu**

3. Splitting follow-up time The Cox proportional hazards model, which is the most commonly used model for timeto-event data in epidemiology, does not generalize to more than one time scale. A simpler parametric alternative is to use Poisson regression with a piecewise-constant hazard. In typical applications of Poisson regression the hazard is constant within time bands defined by 5-year periods of age or calendar year. A single individual may pass through several time bands as shown by Figure 4 which shows the follow-up of a single hypothetical individual and reproduces Figure 2.1 of Breslow and Day (1987). The individual represented in this Lexis diagram passes through 5 time bands before the end of follow-up. The total follow-up time is created by a call to the Lexis function:

This creates a simple Lexis object with only one row. The object may be split into separate time bands using the splitLexis function:

Splitting the follow-up time by 5-year calendar periods creates a new Lexis object with 3 rows. The total follow-up time of 11.12 years is divided up into 3 periods of 3.97+5.00+2.15 = 11.12 years. A second call to splitLexis may be used to split the follow-up time along the age axis

The follow-up time is now divided into the 5 separate parts falling in different time bands defined by age and calendar time. Under the Poisson model, these separate follow-up periods make independent contributions to the likelihood and may therefore be treated as if they come from separate individuals (although, if needed, the lex.id variable keeps track of split follow-up times that come from the same individual). This simple example also shows what happens to the entry and exit status when follow-up time is split. It is assumed that an individual keeps their current status (entry status = 0) at each splitting time until the end of follow-up (exit status = 1) in the last interval. A call to the plot method for Lexis objects creates the plot shown in Figure 4.

When a split Lexis object is plotted, the break points are shown as a background grid. The points method annotates the end of each follow-up segment with a circle, showing how the follow-up line is split whenever it crosses either a horizontal or a vertical grid line.

**3. Divisió del temps de seguiment El model de riscos proporcionals de Cox, que és el model més utilitzat per a les dades sobre esdeveniments en episodis en general, no es generalitza a més d'una escala de temps. Una alternativa paramètrica més senzilla és utilitzar la regressió de Poisson amb un perill constant de peça. En aplicacions típiques de regressió de Poisson, el risc és constant dins de franges de temps definides per períodes de cinc anys o any natural. Un sol individu pot passar per diverses bandes de temps tal i com es mostra a la figura 4, que mostra el seguiment d’un sol individu hipotètic i reprodueix la figura 2.1 de Breslow and Day (1987). L'individu representat en aquest diagrama de Lèxic passa per cinc bandes de temps abans de finalitzar el seguiment. El temps de seguiment total es crea mitjançant una trucada a la funció Lexis:**

**Això crea un objecte Lexis senzill amb una sola fila. L'objecte es pot dividir en franges de temps separades mitjançant la funció splitLexis:**

**Dividint el temps de seguiment en períodes de calendari de cinc anys, es crea un objecte Lexis nou amb tres files. El temps total de seguiment d’11,12 anys es divideix en tres períodes de 3,97 + 5,00 + 2,15 = 11,12 anys. Es pot utilitzar una segona trucada a splitLexis per dividir el temps de seguiment al llarg de l’eix d’edat**

**El temps de seguiment ara es divideix en les cinc parts diferents que es divideixen en diferents franges de temps definides per edats i temps de calendari. Sota el model Poisson, aquests períodes de seguiment separats aporten contribucions independents a la probabilitat i per tant es poden tractar com si procedeixen d’individus separats (tot i que, si cal, la variable lex.id fa un seguiment dels temps de seguiment dividits que venen. del mateix individu). Aquest senzill exemple també mostra el que passa a l'estat d'entrada i sortida quan es divideix el temps de seguiment. Se suposa que un individu manté el seu estat actual (estat d'entrada = 0) a cada temps de divisió fins al final del seguiment (estat de sortida = 1) en el darrer interval. Una trucada al mètode de traça per a objectes Lexis crea la trama que es mostra a la figura 4.**

**Quan es dibuixa un objecte Lexis dividit, els punts de ruptura es mostren com a graella de fons. El mètode de punts anota el final de cada segment de seguiment amb un cercle, que mostra com es divideix la línia de seguiment sempre que creua una línia de quadrícula horitzontal o vertical.**

4. Modelling risk on multiple time scales Returning to the cohort of nickel smelters, we now show how time splitting may be combined with Poisson regression.

The timeBand function returns information about the time band on a given time scale. It can label the time bands in many different ways, according to the type argument. For Poisson regression, it is easiest to return a factor.

Factors are labelled in the same way as for the cut function, as can be seen from the selected output for subject 8. These factors may then be used as predictor variables in a Poisson regression model that separates the effects of age from time since first employment:

Since no deaths occur from lung cancer before age 40 in this cohort, we have removed the lowest level of the age factor from the model using the subset argument to the glm function.

4. Model de risc a diverses escales de temps Tornant a la cohort de fundidores de níquel, ara mostrem com es pot combinar la divisió del temps amb la regressió de Poisson.

La funció TimeBand retorna informació sobre la banda de temps en una escala de temps determinada. Pot etiquetar les franges horàries de moltes maneres diferents segons l’argument tipus. Per a la regressió de Poisson, és més fàcil tornar un factor.

Els factors estan etiquetats de la mateixa manera que per a la funció de tall, com es pot veure a la sortida seleccionada per al tema 8. Aquests factors poden ser utilitzats com a variables predictores en un model de regressió de Poisson que separa els efectes de l'edat del temps des del primer treball. :

Com que no es produeixen morts per càncer de pulmó abans dels 40 anys en aquesta cohort, hem eliminat el model més baix del factor edat del model mitjançant l’argument del subconjunt a la funció glm.

5. Time splitting on an event Lexis objects also allow follow-up time to be split on an event. We illustrate this using data from a cohort of patients who were exposed to Thorotrast (Andersson, Vyberg, Visfeldt, Carstensen, and Storm 1994; Andersson, Carstensen, and Storm 1995), a contrast medium used for cerebral angiography in the 1930s and 1940s that was later found to cause liver cancer and leukaemia (IARC 2001). Data on the cohort are contained in the data set thoro in the Epi package. We convert the thoro data frame into a Lexis object using the data of injection of Thorotrast (injecdat) as the data of entry, and using time scales of calendar time (“cal”) and age (“age”). The cal.yr function from the Epi package is used to convert the Date variables to numeric calendar years.

For these data, the exit status may take three values (1 = dead, 2 = alive, 3 = lost to followup). We give all subjects an entry status of 2 (alive) in the call to the Lexis function. The summary method for Lexis objects prints a table showing the transitions between entry and exit states and a second table showing the transition rates.

In this study, 1966 out of 2470 participants (80%) died before the end of follow up in 1992, and the overall mortality rate was 4% per year. For participants in the cohort who developed liver cancer during follow-up, the variable liverdat contains the date of diagnosis of liver cancer. For other participants, this date is missing. We can use the cutLexis function to split follow-up time for liver cancer cases into pre-diagnosis and post-diagnosis:

Unlike the splitLexis function, the cutLexis function allows the status variable to be modified when follow-up time is split. The new.state = 4 argument means that the date we are splitting on is the date of transition to state 4 (liver cancer case). This is reflected in the updated transition table printed by the summary method, which shows 35 incident liver cancer cases (transitions 2 → 4), all of whom died during follow-up (transitions 4 → 1). The summary output also shows that the Lexis object has data on 2468 persons instead of the 2470 in the original object. In fact the 2 individuals who were dropped have no follow-up time in the study: their exit date is the same as the entry date. The cutLexis function automatically drops follow-up intervals with zero length. Survival from liver cancer can be analyzed by selecting only the rows of the Lexis object that represent follow-up after diagnosis, when the participant is in state 4.

**5. Dividir el temps en un esdeveniment Els objectes Lexis també permeten dividir el temps de seguiment d'un esdeveniment. Ho il·lustrem utilitzant dades d’una cohort de pacients exposats a Thorotrast (Andersson, Vyberg, Visfeldt, Carstensen i Storm 1994; Andersson, Carstensen i Storm 1995), un medi de contrast utilitzat per a l’angiografia cerebral a la dècada de 1930 i 1940 que més tard es va trobar que causava càncer de fetge i leucèmia (IARC 2001). Les dades de la cohort es troben al conjunt de dades del paquet Epi. Convertim el fotograma de dades thoro en un objecte Lexis utilitzant les dades d'injecció de Thorotrast (injecdat) com a dades d'entrada, i utilitzant escales de temps de calendari ("cal") i edat ("age"). La funció cal.yr del paquet Epi s'utilitza per convertir les variables de data en anys de calendari numèrics.**

**Per a aquestes dades, l'estat de sortida pot prendre tres valors (1 = mort, 2 = viu, 3 = perdut per al seguiment). Posem a tots els subjectes un estat d’entrada de 2 (viu) a la crida a la funció Lexis. El mètode de resum dels objectes Lexis imprimeix una taula que mostra les transicions entre els estats d'entrada i sortida i una segona taula que mostra les taxes de transició.**

**En aquest estudi, el 1966 sobre 2470 participants (el 80%) van morir abans de finalitzar el seguiment el 1992, i la taxa de mortalitat global va ser del 4% anual. Per als participants de la cohort que va desenvolupar càncer de fetge durant el seguiment, la variable liverdat conté la data de diagnòstic del càncer de fetge. Per a altres participants, aquesta data no troba. Podem utilitzar la funció cutLexis per dividir el temps de seguiment dels casos de càncer de fetge en pre-diagnòstic i post-diagnòstic:**

**A diferència de la funció splitLexis, la funció cutLexis permet modificar la variable d'estat quan es divideix el temps de seguiment. El nou argument.state = 4 significa que la data en què dividim és la data de transició a l'estat 4 (cas de càncer de fetge). Això es reflecteix a la taula de transició actualitzada impresa pel mètode resum, que mostra 35 casos de càncer de fetge incident (transicions 2 → 4), tots els quals van morir durant el seguiment (transicions 4 → 1). El resultat resum també mostra que l'objecte Lexis té dades sobre 2468 persones en lloc del 2470 de l'objecte original. De fet, les dues persones caigudes no tenen cap moment de seguiment a l'estudi: la data de sortida és la mateixa que la d'entrada. La funció cutLexis baixa automàticament els intervals de seguiment amb longitud zero. La supervivència del càncer de fetge es pot analitzar només seleccionant les files de l’objecte Lexis que representen el seguiment després del diagnòstic, quan el participant es troba en l’estat 4.**

The results are shown in Figure 5. Survival from liver cancer in this cohort is very short, except for one case who survives more than 10 years after diagnosis. Such an anomalous result may prompt further checking of the data to ensure that the date of diagnosis or date of death had not been mis-coded. The same technique of splitting follow-up by event times can also be applied to multi-state disease models, in which arbitrarily complex transitions between disease states are possible. The additional machinery in the Epi package to handle this more complex situation is the subject of a companion paper (Carstensen and Plummer 2011).

**A la Figura 5. Els resultats es mostren a la Figura 5. La supervivència del càncer de fetge en aquesta cohort és molt curta, tret d’uns casos que sobrevisquin més de 10 anys després del diagnòstic. Aquest resultat anòmal pot demanar una revisió posterior de les dades per assegurar-se que la data del diagnòstic o la data de la defunció no s’havien codificat erròniament. La mateixa tècnica de dividir el seguiment en temps d'esdeveniment també es pot aplicar a models de malalties multiestatals, en què és possible transicions arbitràriament complexes entre estats de malaltia. La maquinària addicional del paquet Epi per gestionar aquesta situació més complexa és objecte d’un document d’acompanyament (Carstensen i Plummer 2011).**