

# DIPLOMATURA D'ESTADÍSTICA

---

**Anàlisi de la progressió a sida o mort en els pacients infectats per l'HIV que inicien tractament antiretroviral d'alta eficàcia de la Cohort PISCIS (1998-2004).**

**Autor: Núria Ortega Serra  
Directora: Anna Esteve Gómez  
Tutor: Klaus Langohr**

UNIVERSITAT POLITÈCNICA DE CATALUNYA  
Biblioteca



1400697714



**Facultat de Matemàtiques  
i Estadística**

UNIVERSITAT POLITÈCNICA DE CATALUNYA




Anàlisi de la progressió a sida o mort en els pacients infectats per l'HIV que inicien tractament antiretroviral d'alta eficàcia de la Cohort PISCIS (1998-2004).


Autor: Núria Ortega Serra  
Directora: Anna Esteve Gómez  
Tutor: Klaus Langohr

Qualificació: 9,5 Bon treball

Llicenciats:

President  
  
R. Homs

Secretari  
A. Juv  
U. Langohr

Vocal  
  
G. J. Sureda

20/07/2007

# 1 Introducció

---

El present document correspon a un Projecte Final de Carrera (PFC) de la Diplomatura d'Estadística. L'estudiant es veu amb la necessitat de realitzar un projecte d'aquesta envergadura per tal d'utilitzar les tècniques adquirides però aplicades encara més a la realitat.

Aquest projecte s'ha realitzat en col·laboració amb un centre extern a la facultat. El Centre d'Estudis Epidemiològics de les ITS i Sida de Catalunya (CEEISCAT) és el centre que em va oferir l'oportunitat de realitzar el projecte. El CEEISCAT és el centre responsable de realitzar la vigilància epidemiològica de la Sida i de les Infeccions de Transmissió Sexual (ITS) a Catalunya, tasca que requereix l'aplicació de diverses tècniques estadístiques per tal de portar a terme el monitoratge de l'epidèmia. La persona que representa al CEEISCAT dins d'aquest projecte és l'Anna Esteve Gómez, desenvolupant el paper de directora.

D'altra banda, com a representant de la UPC està el tutor que és Klaus Langohr. Es va contactar amb ell gràcies a Guadalupe Gómez Melis que va tenir l'agudesa de presentar-nos i així poder formar el bon equip de treball que n'ha sorgit.

La temàtica d'aquest projecte es centra en el camp de la bioestadística. El CEEISCAT coordina, entre d'altres estudis de recerca, una cohort de pacients infectats pel VIH anomenada Projecte per a la Informatització del Seguiment Clínic-epidemiològic de la Infecció per VIH i SIDA (PISCIS). Aquest és el projecte amb el que es realitza l'anàlisi. Per tal d'estudiar les dades de la Cohort es fa ús de les tècniques de supervivència. Es va triar aquesta tècnica per dos motius: el primer, encara que en la diplomatura no s'estudien directament aquests anàlisis per a un bioestadístic és molt necessari tenir coneixement d'aquestes tècniques en cas de treballar com a tal; el segon, per superació personal i ganes d'aprendre aquesta tècnica gràcies a la motivació que em van donar en un seminari els integrants del Grup de Recerca per l'Anàlisi de la Supervivència (GRASS). Les tècniques aplicades d'aquest anàlisi és una descriptiva mitjançant l'estimador de Kaplan-Meier, algunes comparacions entre corbes i una regressió amb el model de Cox per a trobar possibles relacions de la funció de risc amb les variables d'interès i els factors.

El software estadístic utilitzat ha estat el SAS de tal forma s'obté l'anàlisi dins d'un programa tancat. D'una banda, s'ha realitzar tota la preparació de les dades amb l'ús del llenguatge relacionat SQL hostatjat al SAS. D'altre banda, s'ha utilitzat el paquet estadístic per a realitzar l'anàlisi.

## 1.1 Plantejament del problema

Malgrat que España segueix ocupant el segon lloc d'Europa en quant a la incidència de Sida [6], la introducció l'any 1996 dels tractaments antiretrovirals de gran activitat (TARGA) va implicar una disminució de les taxes d'incidència i mortalitat específiques per a aquesta malaltia [13] [14] [16]. Aquesta circumstància ha tingut importants conseqüències clíniques i assistencials, com la millora de qualitat de vida i la supervivència dels afectats.

Actualment, mitjançant cohorts seroprevalents de pacients infectats pel VIH és possible estudiar l'efectivitat del tractament antiretroviral (TARV), donades la fiabilitat i disponibilitat de marcadors biològics de progressió de la infecció [21]. A España, tot i que s'observa una alta prevalença i incidència de la infecció per VIH i del corresponent impacte sanitari, hi ha poques cohort multicèntriques prospectives de pacients infectats pel VIH [9], especialment amb una ampla representació poblacional. Precisament amb el propòsit de contestar a les preguntes d'investigació que ha sorgit després de la introducció del TARV, i de generar informació útil per a l'avaluació de les actuacions sanitàries, l'any 1998 es va decidir implementar un estudi de cohort i prospectiu de pacients adults infectats per VIH reclutats en 10 hospitals de Catalunya i un de les Illes Balears, denominat com a Projecte per a la Informatització del Seguiment Clínic-epidemiològic de la Infecció per VIH i SIDA (PISCIS).

Per altra banda, la gran diversitat i la permanent evolució de les pautes de tractament antiretroviral, així com la creixent prevalença de resistències als antiretrovirals, impliquen la necessitat d'estudis que permetin monitoritzar i avaluar la efectivitat i efectes adversos del TARV, a més de descriure els canvis en la historia natural de la infecció i en les causes de mortalitat d'aquests pacients.

## 1.2 Objectius

El principal objectiu d'aquest projecte és:

Analitzar els factors que influeixen en la progressió a Sida i/o mort en pacients infectats per l'HIV inclosos en la cohort PISCIS que inicien tractament antiretroviral d'alta eficàcia després de l'any 1998 (pacients naïve) i que no presenten antecedents previs de Sida en el moment d'inici del tractament.

Per un altra banda, es podrien considerar com a objectius personals tots aquells que corresponen a l'aprenentatge de totes les tècniques estadístiques i informàtiques que comporta la realització del projecte, que serien l'estudi i aplicació de l'anàlisi de la supervivència, l'ús del llenguatge relacional SQL i l'ús del paquet estadístic SAS.

### 1.3 Hipòtesis del treball

Els investigadors en l'àmbit de la recerca clínic-epidemiològica plantegen les següent hipòtesis:

1. Els principals determinants de la morbiditat i mortalitat associats a la progressió de la infecció són els nivells baixos de limfòcits CD4 (<200cel/ $\mu$ l) i una càrrega viral en sang elevada (>100.000 còpies RNA/ml).
2. Estar coinfectat pel virus de l'hepatitis C influeix en una major morbiditat i mortalitat pel Virus de la Immunodeficiència Humana (VIH).
3. Els pacients que inicien teràpia antiretroviral de gran activitat (TARGA) després de l'any 2000 tenen una menor progressió de la infecció.

### 1.4 Tècniques d'anàlisi de la supervivència

L'anàlisi de la supervivència és una tècnica que treballa en unitats de temps. La seva aplicació més freqüent és en l'àmbit de la medicina i la biologia, però també s'utilitza en altres àrees, com per exemple l'anàlisi de fiabilitat en l'àmbit industrial. L'objectiu a mesurar és el temps que transcorre entre dos events, per exemple el temps que viuen uns pacients envers una malaltia, o bé, el temps que triga una població en risc de contraure la malaltia.

Els objectius de l'anàlisi de supervivència són, d'entre altres, descriure els temps de supervivència, la comparació de les distribucions dels temps de supervivència corresponents a dues o més poblacions i l'estudi entre els temps de vida i les variables explicatives que permeten identificar factors de risc o factors pronòstic. [10]

Els temps de supervivència presenten sovint dues particularitats: per una banda, són els temps censurats (veure apartat 1.4.1), per l'altra banda la variable temps no segueix una distribució normal. Per tal de caracteritzar el model és suficient amb una de les següents funcions: funció de supervivència, funció de distribució, funció de densitat de probabilitat, funció de risc, funció de risc acumulada i la vida residual mitja. La funció de supervivència mesura la probabilitat de que un individu sobrevisqui més de t unitats de temps.

$$S(t) = \text{Pr ob}(T > t) \quad \forall t > 0$$

La funció de risc seria la probabilitat de que a un individu que va sobreviscut fins al moment t li ocorri un esdeveniment en el proper instant. Aquesta funció mostra com canvia el risc amb el temps.

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \text{Pr ob}[t \leq T < t + \Delta t | T \geq t]$$

La funció de risc acumulada és definida com l'integral de la funció de risc:

$$\Lambda(t) = \int_0^t \lambda(u)$$

La vida residual mitja mesura l'esperança de la vida restant.

$$vrm(t) = E(T - t | T > t)$$

### 1.4.1 Tipus de dades

En primer lloc, hi ha dues variables molt importants que són el temps en fer event i la variable censura. El temps en fer event és la variable que mesura per a cada pacient en dies, setmanes, mesos o anys el temps que triga des de la inclusió a l'estudi fins a que es dona la malaltia o bé la mort. En cas de que en algun pacient no es doni aquest esdeveniment es calcula la diferència entre la data de tancament de l'estudi i la data d'inclusió. Aquests últims pacient seran precisament els censurats i per tant cal identificar-los amb una variable binària.

Es disposa de les següents observacions:

- $T_1, T_2, \dots, T_n$  que corresponen als temps observats
- $C_1, C_2, \dots, C_n$  que corresponen a variables aleatòries i idènticament distribuïdes amb una funció de distribució G.
- $Y_i = \min(T, C)$
- $\delta_i = \begin{cases} 1 & \text{si } Y = T \\ 0 & \text{si } Y = C \end{cases}$

La censura pot prendre diverses formes: per la dreta, per l'esquerra, en un interval o la censura doble.

D'una banda, la censura per la dreta pren aquest nom quan per a un pacient no s'ha donat el succés durant el transcurs de l'estudi. Aquesta censura es pot donar degut a l'estudi s'ha finalitzat en un moment determinat i no li ha donat temps al pacient, es perd el seguiment d'alguns pacients, degut a que s'interromp el tractament per a alguns pacients o bé, degut a que el pacient produeix un event per un altre causa.

També n'hi ha diversos tipus: tipus I i tipus II. La censura de tipus I es produeix quan l'event només s'observa si es dona abans que un temps especificat. Mentre que la censura de tipus II es dona quan es decideix finalitzar l'estudi després de que s'hagin donat un nombre determinat d'events.

D'altra banda, la censura per l'esquerra es dona quan s'ha produït l'event en un pacient abans d'entrar en estudi. D'aquests pacients tan sols es sap que han fet event abans d'entrar a l'estudi però no es coneix el moment.

Per últim, la censura es pot donar en un interval que correspon a quan l'event no es pot observar amb exactitud però hi ha prou informació com per ajustar-ho en un interval.

Si a més a més l'origen del temps de supervivència està censurat, com per exemple en cas d'una infecció, es parla de `censura` doble.

A continuació només considerem el cas de censura per la dreta.

### 1.4.2 Estimador Kaplan-Meier

L'estimador de Kaplan-Meier o bé també conegut com l'estimador del límit del producte està basat en la descomposició de la funció de supervivència en un producte de probabilitats condicionades amb intervals aleatoris.

Es donen dues situacions a l'hora de calcular l'estimador: amb empats o sense. Es consideren sense empats quan els temps d'event observats són tots diferents entre ells i per tant, amb empats en el cas contrari.

A continuació sigui  $Y_i$  el temps de supervivència de l'individu  $i$ . Aquest temps pot ser el temps de supervivència exacte o un temps censurat. Per tal de distingir entre temps exacte i censurat, s'ha de definir una variable indicadora de la censura,  $\delta_i$ , associada a  $Y_i$ .  $\delta_i$  és igual a 1 en cas d'una observació exacta i zero en cas contrari.

#### *Estimador de Kaplan-Meier sense empats*

Per tal de calcular aquest estimador es suposa que tots els  $Y_i$  siguin diferents i s'ordenen les observacions ( $Y_{(1)} < \dots < Y_{(n)}$ ). Es defineix intervals aleatoris  $I_i = (\tau_{i-1}, \tau_i]$  on cada interval conté exactament un valor observat. Com s'ha dit anteriorment aquest estimador es calcula a partir de la descomposició en productes.

L'estimador de Kaplan-Meier per a dades sense empats és una funció escalonada on cada salt depen dels temps observats ja siguin d'event o de censura. La seva fórmula és:

$$\hat{S}(t) = \prod_{i: Y_{(i)} \leq t} \hat{p}_i = \prod_{i: Y_{(i)} \leq t} \left(1 - \frac{1}{n_i}\right)^{\delta_{(i)}} = \prod_{i: Y_{(i)} \leq t} \left(\frac{n-1}{n-i+1}\right)^{\delta_{(i)}} ,$$

on  $n$  és el número total d'individus a l'estudi i  $n_j$  és el número d'individus a risc justament abans de  $Y_{(j)}$ .

### *Estimador de Kaplan-Meier amb empats*

Per tal de calcular aquest estimador tan sols cal treballar més la fórmula anterior amb l'afegit que cal suposar que just abans del moment  $\tau_k$  hi ha  $n_k$  pacients vius i en el moment  $\tau_k$  es produeixen  $d_k > 1$  morts. En els casos d'empats entre observacions censurades i observacions no censurades es considerarà sempre que les no censurades han ocorregut abans.

L'estimador pren la següent forma:

$$\hat{S}(t) = \begin{cases} 1 & \text{si } t < Y_{(1)} \\ \prod_{i: Y_{(i)} \leq t} \left(1 - \frac{d_i}{n_i}\right) & \text{si } t \geq Y_{(1)} \end{cases}$$

### *Variància de l'estimador*

La variància d'aquest estimador ve donada per la fórmula de Greenwood:

$$\hat{V}_G(\hat{S}(t))^2 = \sum_{i: Y_{(i)} \leq t} \frac{d_i}{n_i(n_i - d_i)}$$

D'altra banda, però, existeixen altres possibilitats com la proposició que varen fer Aalen i Johansen:

$$\hat{V}_G(\hat{S}(t))^2 = \sum_{i: Y_{(i)} \leq t} \frac{d_i}{n_i^2} \quad (\text{variància de Tsiatis})$$

### *Propietats de l'estimador*

- L'estimador K-M coincideix amb l'estimador que s'obté mitjançant un algorisme de redistribució a la dreta
- Autoconsistent
- Sota certes condicions de regularitat l'estimador K-M és l'estimador de màxima versemblança no paramètric generalitzat
- És un estimador consistent
- Sota certes condicions de regularitat l'estimador convergeix debilment a un procés Gaussià.



### 1.4.3 Comparació entre corbes de supervivència

Com ja s'ha dit anteriorment els dades d'estudi generalment no són normals i per tant, caldrà realitzar proves no paramètriques. En cas de no tenir dades censurades aleshores es podran utilitzar els clàssics mètodes no paramètrics.

En canvi, els següents mètodes es poden aplicar a dades censurades, tan per la dreta com per l'esquerra. Es basen en el nombre d'events en cada temps i el nombre de pacients en risc en cadascun d'aquests temps. Es descriuen a continuació algunes proves per a dues mostres.

Les hipòtesis nul·la i alternativa són les següents:

$$H_0 : S_1(t) = S_2(t) \quad \forall t$$

$$H_1 : \exists t : S_1(t) \neq S_2(t)$$

Prova del log-rang (Cox i Mantel)

Aquesta prova és òptima per a detectar aquelles alternatives en les que les funcions de risc són proporcionals. O sigui, en aquest cas particular l'hipòtesi es podria plantejar també com:

$$H_0 : \lambda_1(t) = \lambda_2(t) \quad \forall t$$

$$H_1 : \lambda_2(t) = \exp \beta \lambda_1(t)$$

En el cas de comparació de dos tractaments es construiria la següent taula per cada temps  $Y_i$ :

	Tractament 1	Tractament 2	Total
Nombre d'events	$d_{i1}$	$d_{i2}$	$d_i$
Nombre de no events	$R_{i1}-d_{i1}$	$R_{i2}-d_{i2}$	$R_i-d_i$
Nombre de pacients en risc	$R_{i1}$	$R_{i2}$	$R_i$

Aleshores es pot construir l'estadístic considerant que la variable aleatòria  $d_{ij}$  condicionada a  $R_i$ ,  $R_{ij}$  i a  $d_i$  és una hipergeomètrica de paràmetres  $R_i$ ,  $R_{ij}$  i  $d_i$ .

$$LR = \frac{\left( \sum_{i=1}^D \left( d_{i1} - \frac{R_{i1}d_i}{R_i} \right) \right)^2}{\text{Var} \left( \sum_{i=1}^D \left( d_{i1} - \frac{R_{i1}d_i}{R_i} \right) \right)} \quad \text{amb} \quad \text{Var} \left( \sum_{i=1}^D \left( d_{i1} - \frac{R_{i1}d_i}{R_i} \right) \right) = \sum \frac{R_{i1}R_{i2}(R_i - d_i)d_i}{R_i^2(R_i - 1)}$$

La variable aleatòria LR s'aproxima a una  $\chi^2$  quan el nombre total d'individus és suficientment gran.

Prova de Grehan

Aquesta prova considera pesos que depenen fortament del moment en que succeeix un event i de la distribució de la censura i per conseqüent pot conduir a resultats falsos si els patrons de censura de cada població són molt diferents.

Es defineix una funció anomenada *score*:

$$\Phi((Y_{i1}, \delta_{i1}), (Y_{i2}, \delta_{i2})) = \begin{cases} +1 & \text{si } Y_{i1} \leq Y_{i2} \text{ i } \delta_{i1} = 1 \text{ i } \delta_{i2} = 0, \\ +1 & \text{si } Y_{i1} < Y_{i2} \text{ i } \delta_{i1} = 1 \text{ i } \delta_{i2} = 1, \\ -1 & \text{si } Y_{i1} \geq Y_{i2} \text{ i } \delta_{i1} = 0 \text{ i } \delta_{i2} = 1, \\ -1 & \text{si } Y_{i1} > Y_{i2} \text{ i } \delta_{i1} = 1 \text{ i } \delta_{i2} = 1, \\ 0 & \text{en altres casos} \end{cases}$$

L'estadístic de Gehan es defineix com a:

$$U = \sum_{i=1}^n \sum_{l=1}^n \Phi((Y_{i1}, \delta_{i1}), (Y_{l2}, \delta_{l2}))$$

#### 1.4.4 Regressió de Cox

El model de Cox, o bé el model de riscos proporcionals, s'utilitza per analitzar les relacions entre la variable temps i diferents covariants o variables independents. Aquestes es pot distingir entre variables quantitatives i variables categòriques, els factors.

El model bàsic de Cox modelitza la funció de risc en funció de les covariants mitjançant la expressió:

$$\lambda(t | Z) = \exp\{\beta_1 Z_1 + \beta_2 Z_2 + \dots + \beta_p Z_p\} \lambda_0(t),$$

on  $Z = (Z_1, Z_2, \dots, Z_p)$  és el conjunt de totes les covariants, també anomenat perfil,  $\lambda_0(t)$  la funció de risc basal i el factor  $\exp\{\beta_1 Z_1 + \beta_2 Z_2 + \dots + \beta_p Z_p\}$  expressa quantes vegades es major el risc per un event d'un individu amb perfil Z respecte un altre amb perfil  $Z=0$ . Nota que la funció de risc basal és la funció de risc que correspondria a un individu amb valors 0 a totes les covariants.

En el model de Cox el logaritme del quocient de les funcions de risc es relaciona linealment amb els factors pronòstics:

$$\ln(\lambda(t | Z) / \lambda_0(t)) = \beta_1 Z_1 + \beta_2 Z_2 + \dots + \beta_p Z_p$$

A més, el logaritme del quocient de les funcions de risc, així com el quocient de les funcions de risc, es suposa constants independentment de en quin moment es calculi. Per això es coneix el model de Cox com a model de riscos proporcionals.

D'altra banda, la funció de risc basal pot prendre qualsevol forma, per aquest motiu es diu que el model de Cox és semiparamètric.

### *El model de Cox per a mostres censurades per la dreta*

Les variables de censura  $C_i$  poden estar relacionades amb les covariants  $Z_i$  de qualsevol forma. Encara que, cal suposar que les variables aleatòries són condicionalment independents donat  $Z_i$ .

### Incorporació de variables

Com s'ha mencionat abans, la funció de risc pot dependre de variables i factors, entenent com a variables les que prenen valors continus i com a factors com a variables categòriques que prenen un nombre limitat de valors coneguts com els nivells del factor.

En el cas de les variables, s'incorporen en el model de riscos proporcionals assignant a cadascuna d'elles un coeficient  $\beta$ . En canvi, en el cas dels factors s'incorporarà *la quantitat de nivells del factor - 1* coeficients en el model i es suposa que el primer nivell actua com a basal o de referència. En el cas d'introduir més d'un factor al model aleshores també es podria analitzar les possibles interaccions pels efectes corresponents a les diferents combinacions entre ells.

Per a estimar els paràmetres  $\beta = (\beta_1, \beta_2, \dots, \beta_p)$  que quantifiquen el grau d'associació entre la variable temps i  $Z$ , s'utilitza la tècnica de la màxima versemblança.

### *Anàlisi dels residus*

Hi ha diferents definicions per als residus d'un model de Cox: Cox-Snell, martingales, score, deviance i Schoenfeld. Els residus serveixen per a avaluar la bondat d'ajust del model.

### Residus Cox-Snell

Es defineix com:

$$rc_i = \exp\{\hat{\beta}'Z_i\} \hat{\lambda}_0(y_i),$$

on  $\hat{\Lambda}_0(\cdot)$  és la funció de risc acumulada basal estimada segons el model de Cox. Aquests residus no estan distribuïts simètricament al voltant del 0 i no poden ser negatius. Són útils per a la comprovació del ajust global del model.

Si el model és correcte els residus de Cox i Snell formen una mostra censurada per la dreta d'una llei exponencial de mitjana 1. També hi ha un altre versió "millorada" que diferencia entre observacions censurades i no censurades.

Aquest model no dona indicació de la llunyania del model quan el gràfic no és lineal. La distribució exponencial dels residus és vàlida amb els paràmetres poblacionals, però a l'utilitzar les seves estimacions és pot perdre linealitat especialment en mostres petites es produeixen més imprecisions en les cues dretes.

### Residus en martingales

Aquests residus són una versió dels anteriors.

$$r_{M_i} = \begin{cases} 1 - rc_i = 1 - \exp\{\hat{\beta}'Z_i\}\hat{\Lambda}_0(y_i) & \text{si la observació no esta censurada} \\ -rc_i = -\exp\{\hat{\beta}'Z_i\}\hat{\Lambda}_0(y_i) & \text{si la observació esta censurada} \end{cases}$$

Aquests residus són útils per a:

- analitzar els residus com una mesura de l'excés d'events, ja que posaran en evidència aquells individus mal ajustats.
- Per a determinar la millor transformació per a una covariant de forma que expliqui de manera òptima el seu efecte sobre la supervivència.

### Residus score

$$r_{S_{ik}}(t) = \int_0^t \{Z_{ik}(s) - \bar{Z}_k(s)\} d\hat{M}_i(s)$$

$$\bar{Z}_k(t) = \frac{\sum_{i=1}^n J_i(t) Z_{ik} \exp\{\beta'Z_i(t)\}}{\sum_{i=1}^n J_i(t) \exp\{\beta'Z_i(t)\}}$$

$$\hat{M}_i(t) = N_i(t) - \int_0^t J_i(s) \exp\{\beta'Z_i(s)\} d\hat{\Lambda}_0(s)$$

Aquests residus permeten que les covariants depenguin del temps. El conjunt de residus basats en score formen una matriu  $n \times p$ . Cadascun d'aquests residus estableix la diferència entre el valor observat i l'esperat de la covariant  $Z_{ji}$  corresponent a l'individu  $i$ -éssim.

Aquests residus són útils per a comprovar la hipòtesis de riscos proporcionals i per a determinar la influència del individu  $i$ -éssim.

### Residus deviance

Són una transformació dels residus basats en martingales.

$$r_{D_i} = \text{Signo}(r_{M_i}) \left[ -2 \{ r_{M_i} + \delta_i \log(\delta_i - r_{M_i}) \} \right]^{1/2}$$

Aquests residus estan distribuïts al voltant del 0 quan el model és correcte. A més, els hi correspon una distribució aproximadament normal si el model és correcte.

S'utilitzen per a decidir la bondat del model per a cada individu i per a trobar els valors atípics, per avaluar l'efecte de cada individu en el model.

### Residus Schoenfeld

Aquests residus es calculen per al  $k$ -éssim event i el  $i$ -éssim individu quan no hi ha empats:

$$r_{SC_{ik}}(t) = \delta_i J_i(t) \{ Z_{ik} - \bar{Z}_k(T_i) \}$$

Cadascun d'aquests residus estableix la diferència entre el valor observat de la covariant en el  $k$ -éssim temps d'event i el valor esperat de la covariant en aquell moment.

S'utilitzen per a determinar la influència que cada individu té en la estimació dels coeficients i per a una robusta estimació de la variància. A més a més serveixen per comprovar la proporcionalitat dels riscos.

## 1.5 Implicacions ètiques

En el marc clínic de la investigació biomèdica es troben definits els principis bioètics. Encara que en el cas d'una cohort no es correspon tan a investigació sinó que a seguiment epidemiològic, els principis bioètics són igualment aplicables. [3]

Es poden diferenciar entre principis generals i específics. Els principis bioètics generals són aquests quatre:

- Principi de No Maleficència

No fer mai res negatiu a algú encara que així ens ho demani.

- Principi de Justícia

Obliga a tractar a tots els éssers humans amb igual consideració i respecte.

- Principi d'Autonomia

Requereix que els individus, segons la seva capacitat entenedora, tinguin l'oportunitat de decidir si participen o no en un estudi, després de rebre tota la informació necessària.

- Principi de Beneficència

Conté la obligació de maximitzar els beneficis previsibles i de minimitzar els possibles riscos i incomoditats derivats de la investigació clínica.

Els principis bioètics específics fan més referència al servei cap a la societat i són els següents:

- Principi de Transparència Informativa

Valora la veritat i l'objectivitat. Requereix evitar tot allò que afecti a altres i no es faci amb transparència.

- Principi de Respecte a la Vida Privada

Requereix intimitat envers el pacient, especialment en facilitar a ningú l'estat de la seva malaltia.

- Principi de Responsabilitat Informativa

Requereix al metge informar d'una forma responsable, que vol dir pensar a priori les conseqüències que pot tenir el coneixement d'una certa informació sobre els pacients de la societat.

- Principi d'Equitat Informativa

Requereix que tots els pacients en situacions siguin tractats per igual i amb la mateixa informació. Però ningú es mereix una divulgació d'informació errònia ni cap vulneració de confidencialitat.

Per últim, en la bioètica es troben les **normes d'una Bona Pràctica Clínica** o també anomenada **Bona Pràctica d'Investigació Clínica**. Una possible definició de la BPC seria:

*“Es la denominación que ha recibido un conjunto internacional de normas éticas y de calidad científica internacionales que deben regir el diseño, realización, registro de datos y comunicaciones de Ensayos Clínicos. El cumplimiento de dichas normas garantiza que se han protegido los derechos, bienestar y confidencialidad de los sujetos participantes en el ensayo, así como que los datos obtenidos son fiables.”*

En l'aplicació de tots aquests principis en l'estudi amb la Cohort PISCIS es pot observar que hi ha una responsabilitat del seu compliment compartida entre els hospitals reclutats en la cohort i el CEEISCAT que és el centre coordinador.

Encara que hi ha principis que tan sols afecten als hospitals ja que afecten a la relació dels metges de cada hospital amb el seus pacients. Els principis generals que serien aquest cas són el principi de No Maleficència i el principi d'Autonomia. Respecte al principi de No maleficència, com que el CEEISCAT no té contacte directe amb els pacients aleshores no se'l pot afectar negativament ni tampoc ens ho pot demanar. El principi d'Autonomia depèn exclusivament del metge i del pacient, ja que són els únics presents en el moment en que el metge proposa l'entrada a l'estudi.

En canvi, la resta de principis estarien compartits amb el CEEISCAT. El CEEISCAT col·labora en certa manera al igual que els metges a l'hora de maximitzar els beneficis i minimitzar els riscos per als pacients, tenir transparència informativa, respectar la intimitat i la confidencialitat, informar d'una forma responsable i tenir equitat informativa a l'hora de tractar les dades de cada pacient.

Com a actuacions enfront a aquestes implicacions ètiques hi ha, entre d'altres:

- Evitar interpretacions distorsionades de la informació
- Esmentar la procedència de les dades
- Fomentar la publicació dels estudis realitzats
- Garantir la confidencialitat de les dades, especialment si s'han obtingut extraordinàriament.
- Donar una autèntica igualtat d'oportunitats a les persones.
- Garantir informació sobre els problemes de salut més freqüents.

## 2 Descripció de la Cohort PISCIS

---

El PISCIS és un estudi de cohort obert i multicèntric de pacients infectats per VIH. (Jaén A, et al. *Med Clin* 2005). Breument, la cohort PISCIS està constituïda per 10 hospitals de la Comunitat de Catalunya i un de les Illes Balears. Inclou a tots els individus majors de 16 anys, amb serologia confirmada positiva envers el VIH, que són visitats per primera vegada en els hospitals participants en l'estudi a partir del 1 de gener de 1998. Tots els pacients infectats pel VIH que compleixen aquests criteris s'inclouen independentment del estadiatge de la seva infecció. Es realitza el seguiment dels pacients en unitats especialitzades o consultes externes seguint el protocol clínic assistencial habituals per a aquests pacients, normalment amb visites cada 3-4 mesos. El centre d'Estudis Epidemiològics sobre les ITS i Sida de Catalunya (CEEISCAT) actua com a centre coordinador..

Per a construir el mínim comú de dades es va dissenyar una base de dades relacional centralitzada mitjançant el Sistema Gestor de Bases de Dades Microsoft Access 2000, estructurada en 8 taules (Altes, malalties, seguiment, tractament, efectes adversos, analítiques, ingressos, baixes) (veure *Annex 1: relació de taules de la Base de Dades PISCIS*) que inclouen totes les variables definides en el protocol. Algunes d'aquestes variables es recullen en el moment de la inclusió en la cohort i altres en cadascuna de les visites subsegüents. Entre elles hi ha variables demogràfiques com la data de naixement, sexe, província de naixement; variables socioeconòmiques com la situació laboral, formació acadèmica, província de residència; variables epidemiològiques com el grup de transmissió, data de primera serologia positiva per al VIH, any probable d'infecció, tractament antiretroviral previ; variables clíniques com les malalties definitòries de Sida i altres relacionades; variables de tractament com la quimioprofilaxis de les infeccions oportunistes, tractaments antiretrovirals, efectes adversos de TARGA, test de resistència antiretrovirals; variables biològiques com el recompte i percentatge de limfòcits CD4 i CD8, càrrega viral del VIH en plasma, serologia pel virus de la hepatitis B (VHB) i virus de la hepatitis C (VHC); variables assistencials com la vacunació per al VHB i pneumococ i ingressos hospitalaris, i variables de pèrdua de seguiment que inclouen la data i el motiu de la baixa, i en el cas de que el motiu sigui la defunció es recull la causa fonamental d'aquest segons el criteri del metge responsable.



## 3 Preparació de les dades per l'anàlisi estadístic

---

### 3.1 Criteris d'inclusió/exclusió per l'estudi

En un estudi com el present és necessari considerar criteris d'inclusió i exclusió dels pacients per tal de poder portar a terme un anàlisi que pugui respondre les hipòtesis inicials de l'estudi. En aquest cas s'ha considerat que complien els criteris d'inclusió aquells pacients que:

1. no havien rebut tractament previ a l'inici del seguiment a la Cohort (naïve).
2. varen iniciar tractament antiretroviral d'alta eficàcia (TARGA).
3. no havien desenvolupat una malaltia definitiva de sida abans de l'inici del tractament.

El primer criteri evita possibles confusions de progressió de la malaltia degudes al tractament previ. El segon criteri, imposa que només se seleccionin els pacients que inicien aquest potent tractament, ja que sinó no es podrien valorar correctament els seus efectes. El tercer criteri és necessari ja que els pacients que varen desenvolupar la sida es consideren en l'última etapa de la malaltia i per tant, corresponen a comportaments extrems i seria convenient estudiar-los separatament.

Per últim, també s'han tingut en compte altres criteris referents a una mínima qualitat en les dades, com són:

4. Mínim seguiment.  
Les dades de les que es disposa són fins a finals del 2004. Per tal d'assegurar un mínim de seguiment (es fa una visita a cada pacient cada 3-4 mesos) per als últims pacients inclosos s'imposa que la data d'inici del tractament sigui anterior al 30 de juny del 2004.
5. Tenir alguna analítica amb valor de CD4.  
Per tal de poder valorar l'estat biològic del pacient s'imposa que tingui un valor d'analítica de CD4 a l'inici del tractament.
6. Treure pacients amb valors erronis.  
En la elaboració del MCD de la base de dades de la cohort es detecten pacients amb valor erronis. Per tal de fer una correcta depuració d'errors no s'inclouen tampoc aquests pacients.

## 3.2 Consultes SQL

### 3.2.1 Criteris inclusió/exclusió

**Criteri 1:** El criteri establert a per a l'estudi és el d'excloure tots aquells pacients que han estat en tractament prèviament (veure consulta SQL *consulta\_criteri\_1* a l'Annex A2.1)

- La taula *mcd\_altes* conté una variable anomenada *Tractament\_previ*. Aquesta variable conté informació sobre si el pacient ha estat en tractament prèviament a l'entrada a la cohort. La codificació d'aquesta variable és:

**Taula 1: Codificació variable *Tractament\_previ***

Codi	Tractament previ
1	Si
2	No
9	Missing

- Per tant, cal incloure tots els pacients que tenen *Tractament\_previ* amb valor 2.

**Criteri 2:** El criteri establert a per a l'estudi és el d'excloure tots aquells pacients que no inicien tractament durant l'estudi prèviament (veure consulta SQL *consulta\_criteri\_2* a l'Annex A2.1)

- La taula *mcd\_altes* conté un parell de variables que contenen informació sobre el tractament que prenen els pacients ja dins en el seguiment en la cohort. Una primera variable és la data d'inici del tractament (*data\_primer\_tractament*). La segona variable conté la pauta d'aquest tractament (*Pauta\_TARGA\_inicial*). S'entén com a pauta la següent codificació:

**Taula 2: Codificació de la variable *Pauta\_TARGA\_inicial***

Codi	Pautes de tractament
0	No tractats
1	Monoteràpia
2	Biteràpia
30	Triteràpia: altres pautes
31	Triteràpia: Pauta 1 o 2 IP + 2 INT
32	Triteràpia: Pauta 2 INT + 1 INNT
33	Triteràpia: Pauta 3 INT

- Els tractaments triteràpia són els que es consideren TARGA. Per tant, tan sols hem d'incloure les pautes 30, 31, 32 i 33; juntament amb tots aquells pacients que tinguin una data per a TARGA, ja que en cas contrari, no inicien TARGA i per tant, estaran exclosos de l'estudi.

**Criteri 3:** El criteri establert a per a l'estudi és el d'excloure tots aquells pacients que varen desenvolupar la sida abans de l'inici del tractament (veure consulta SQL *consulta\_criteri\_3* a l'Annex A2.1)

- En la taula de Seguiments (*mcd\_seguiments*) hi ha la variable *estadiatge* que indica la fase en la que es troba el pacient segons uns estats. La seva codificació és:

**Taula 3: Codificació de la variable *Estadiatge***

Codi	Estadiatge
1	A1
2	A2
3	A3
4	B1
5	B2
6	B3
7	C1
8	C2
9	C3
10	Primo infecció
99	Missing

- L'estadiatge C correspon a la fase on ja es considera que s'ha evolucionat a Sida. Per tant cal excloure tots els pacients que tenen un estadiatge C (7, 8 o 9) en la data d'inici del tractament. No obstant, observant les dades es veu que les persones encarregades d'entrar les dades sovint no omplen aquest camp. Per tant, falta informació sobre l'estadiatge en molts casos. Un altre forma d'aconseguir aquesta informació és a partir de les dades de malalties definitòries. Un pacient es considera que està en estadiatge de Sida quan desenvolupa una malaltia definitòria, aleshores, se li atribueix un estadiatge C. Els estadiatges A i B fan referència a malalties relacionades<sup>1</sup>.

En primer lloc, es consulta els pacients que tenen alguna malaltia definitòria anterior o igual a la data d'inici del tractament (veure consulta SQL *consulta\_criteri\_3.1* a l'Annex A2.1).

<sup>1</sup> Les malalties definitòries i relacionades estan classificades en un llistat tancat amb els seus respectius codis.

En segon lloc, es consulta tots els pacients que tenen un estadiatge C anterior o igual a la data d'inici del tractament (veure consulta SQL *consulta\_criteri\_3.2* a l'Annex A2.1).

**Criteri 4:** El criteri establert en aquesta ocasió és que la data d'inici de TARGA sigui anterior al 30/06/2004, per tal de donar un cert marge als nous pacients. D'altra banda, també es obligat tenir un valor de CD4, ja que sinó no es tindria informació biològica (veure consulta SQL *consulta\_criteri\_4* a l'Annex A2.1)

**Criteri 5:** El criteri establert en aquesta ocasió és excloure tots aquells pacient amb valors erronis i menors de 16 anys d'edat (veure consulta SQL *consulta\_criteri\_5* a l'Annex A2.1)

- La taula *mcd\_altes* conté una variable anomenada *Baixa* que identifica els pacients erronis amb valor 0.

### 3.2.2 Creació de les variables temps de supervivència i censura

#### Variable Sida

**Objectiu 1:** detectar els pacients que desenvolupen la Sida abans del tancament de l'estudi. (veure consulta SQL *consulta\_objectiu\_1* a l'Annex A2.1)

- Aquest punt té cert paral·lelisme amb el punt anterior en el que es buscava els pacients que no tenen Sida al inici, ja que cal aplicar la mateixa metodologia degut a que el problema de l'estadiatge segueix ocorrent.
  - *Primera part: malalties definitòries i estadiatge C durant el seguiment d'estudi.*

En primer lloc, cal detectar tots els pacients que tenen una malaltia definitòria que sigui posterior a la data del primer tractament TARGA (veure consulta SQL *consulta\_objectiu\_1.1\_a* a l'Annex A2.1).

En segon lloc, cal detectar tots els pacients que tenen un estadiatge C també posterior a la data del primer tractament TARGA (veure consulta SQL *consulta\_objectiu\_1.1\_b* a l'Annex A2.1).

- *Segona part: guardar tota la informació per a cada pacient (veure consulta SQL *consulta\_objectiu\_1.2* a l'Annex A2.1)*

En aquesta consulta es guarden les dates de malaltia definitiva i estadiatge C per als pacients que fan Sida durant l'estudi.

### **Variable mort**

**Objectiu 2:** detectar els pacients que moren durant l'estudi. (veure consulta SQL *consulta\_objectiu\_2* a l'Annex A2.1)

- La taula *mcd\_altes* conté una variable anomenada *Baixa*. Aquesta variable conté informació sobre la pèrdua de seguiment del pacient. La codificació d'aquesta variable és:

**Taula 4: Codificació de la variable *Baixa***

<b>Codi</b>	<b>Baixa</b>
0	Pacients amb dades errònies
1	Pacients seguits
2	Pèrdues de seguiment
3	Èxitus (mort)

- Per tant, cal obtenir tots aquells pacients que esdevenen un codi 3 en la variable *Baixa*.

### **Variable esdeveniment (Sida i/o Mort)**

**Objectiu 3:** En aquesta consulta es posa en comú la informació sobre Sida i/o mort a partir de la seva data. Es guarda la variable de la data de l'esdeveniment ja que és el que es necessita per a crear el temps de supervivència posteriorment. (veure consulta SQL *consulta\_objectiu\_3* a l'Annex A2.1).

- També en la consulta es crea automàticament la taula on s'incorporaran les variables de temps de supervivència.

### **Variable data de Sida**

**Objectiu 4:** obtenir la data en que fa Sida un pacient a partir de les dates obtingudes de malaltia definitiva i estadiatge C. A l'hora d'assignar la data de Sida es poden donar dues situacions que s'especifiquen a continuació.

- *Primer pas: la data de Sida és la data de la primera malaltia definitiva en el seguiment d'estudi (veure consulta SQL consulta\_obecjtiu\_4.1 a l'Annex A2.1).*

La primera situació correspon a agafar la data de malaltia definitiva. Això es degut dos possibles motius: la data de malaltia definitiva es la primera de les dues, o no hi ha data d'estadiatge C.

- *Segon pas: la data de Sida és la data del primer estadiatge C en el seguiment d'estudi (veure consulta SQL consulta\_objectiu\_4.2 a l'Annex A2.1).*

La segona situació correspon a agafar la data d'estadiatge C. Això es degut dos possibles motius: la data de d'estadiatge és la primera de les dues, o no hi ha data de malaltia definitiva

### **Variable data de Sida i/o Mort**

**Objectiu 5:** Aquesta data serà la del primer esdeveniment que s'hagi efectuat. Aquest procediment de càlcul és similar al efectuat anteriorment per obtenir la data de Sida.

- *Primer pas: la data de Sida i/o Mort és la data de Sida (veure consulta SQL consulta\_objectiu\_5.1 a l'Annex A2.1).*

La primera situació correspon a agafar la data de Sida. Això es degut dos possibles motius: la data de Sida és la primera de les dues, o no hi ha data de Mort.

- *Segon pas: la data de Sida i/o Mort és la data de mort (veure consulta SQL consulta\_objectiu\_5.2 a l'Annex A2.1).*

La segona situació correspon a agafar la data de Mort. Això es degut dos possibles motius: la data de Mort és la primera de les dues, o no hi ha data de Sida.

### **Variable censura**

**Objectiu 6:** En anàlisi de supervivència cal crear una variable de *censura* que indica si el pacient ha realitzat l'event (sida o mort).

- Primer pas: donar valor a tots aquells pacients censurats (veure consulta SQL *consulta\_objectiu\_6.1* a l'Annex A2.1).
- Segon pas: donar valor a tots aquells pacients no censurats (veure consulta SQL *consulta\_objectiu6.2* a l'Annex A2.1).

### **Variable temps de supervivència**

**Objectiu 7:** Per a l'elaboració del càlcul del temps de supervivència és necessària la data d'event o bé la data d'últim contacte per als pacients censurats, que són els que no realitzen l'event.

- En la taula *mcd\_altes* hi ha una variable anomenada *Últim\_contacte* que conté la última data de la que es té informació sobre el pacient
- Per tant, per calcular el temps de supervivència es realitza la diferència entre la data d'event o la data d'últim contacte amb la data d'inici de tractament que és el primer temps de l'estudi
- *Càlcul temps Sida o mort en dies*

En primer lloc, es calcula el temps de supervivència en dies per a pacients no censurats (veure consulta SQL *consulta\_objectiu\_7.1* a l'Annex A2.1).

En segon lloc, es calcula el temps de supervivència en dies per a pacients censurats (veure consulta SQL *consulta\_objectiu\_7.2* a l'Annex A2.1).

### **3.2.3 Creació d'altres variables**

#### **Trobar l'analítica als 6 mesos de tractament**

Es vol realitzar la comparació dels valors de les analítiques al inici i al cap de 6 mesos de tractament. Els pacients no tenen una analítica exactament als 6 mesos, per tant, cal calcular la més propera a aquest llinard en un rang de 3 mesos (veure consulta SQL *consulta\_criteri4* a l'Annex A2.1).

- Calcular la data als 6 de tractament

En primer lloc, cal calcular la data dels sis mesos (182 dies) a partir de la data d'inici de tractament (veure consulta SQL *consulta\_altres\_1* a l'Annex A2.1).

En la present consulta ja s'estan filtrant les analítiques que estan entre 3 i 9 mesos (91 i 274 dies respectivament).

El resultat de la consulta es guarda en una nova taula anomenada *Anal\_rang* ja que posteriorment es necessitarà donar valors a una nova variable i no seria possible si no es crees la taula.

- Calcular la diferència entre les analítiques dins del rang i la data al cap dels sis mesos de tractament.

S'afegeix la variable diferència (*dif*) sobre la qual es calcularà la diferència de dies entre els 6 mesos i les analítiques filtrades per tal d'obtenir la que té mínima diferència i per tant, la més propera (veure consulta SQL *consulta\_altres\_2.1* i *consulta\_altres\_2.2* a l'Annex A2.1).

Es calcula la diferència en dos passos. El primer pas es calcula la diferència per a analítiques posteriors al llindar dels 6 mesos. El segon pas es calcula la diferència per a analítiques anteriors al llindar dels 6 mesos per tal d'evitar valors negatius en la variable *diferència*.

- Trobar la mínima diferència

Es selecciona l'analítica més propera al llindar dels 6 mesos de l'inici del tractament per a cada pacient (veure consulta SQL *consulta\_altres\_3* a l'Annex A2.1).

Es podria donar que un pacient tingui dues analítiques a la mateixa distancia dels 6 mesos. Per tal d'evitar aquests duplicats d'analítiques s'escull l'analítica la més petita.

- Valor de CD4 en la analítica al cap dels 6 mesos

Es separen els valors d'analítiques de CD4 i CV ja que hi ha registres que contenen la informació sobre aquests marcadors per separat. I per tant, aquest efecte genera duplicats. El resultat d'aquesta consulta només dona l'analítica més propera obtinguda anteriorment que tingui un valor per a CD4, ja que si el valor és nul no aporta informació (veure consulta SQL *consulta\_altres\_4* a l'Annex A2.1).



- Valor de CV en la analítica al cap dels 6 mesos

En la següent consulta s'obtenen els valors de CV per a les analítiques obtingudes anteriorment (veure consulta SQL *consulta\_altres\_5* a l'Annex A2.1).

Per tal d'obtenir tota la informació sobre l'analítica als 6 mesos es col·loquen totes les dades en una mateixa taula (veure consulta SQL *consulta\_altres\_6* a l'Annex A2.1).

Finalment, es construeix una taula final que conté tots els valors de les variables d'interès descrites en aquest apartat (veure consulta SQL *consulta\_taula\_final* a l'Annex A2.1).

## 4 Anàlisi estadístic

---

La descripció de les principals variables clínico-epidemiològiques en el moment d'inici del tractament (característiques basals) es va realitzar pel conjunt de pacients inclosos en l'estudi, així com pel subgrup de pacients que varen evolucionar a sida i els que varen morir. Per aquest anàlisi es consideraren les següents variables:

**Taula 5. Variables d'estudi**

<b>Variables socio-demogràfiques:</b>
Edat
Sexe
Grup de transmissió
<b>Variables clínico-biològiques basals:</b>
Càrrega viral (CV)
CD4
Pauta de TARGA inicial
Període d'inici del TARGA
Infecció pel virus de l'hepatitis C (HCV)
Casos incidents de Sida
Morts
<b>Variables clínico-biològiques als 6 mesos:</b>
CD4 als 6 mesos
CV als 6 mesos

Per a l'anàlisi descriptiva de les variables categòriques es varen utilitzar freqüències i percentatges, mentre que per les variables contínues es varen considerar medianes i rangs interquartílics degut a que les seves distribucions no seguien la llei Normal..

La comparació de les freqüències d'algunes variables qualitatives (hepatitis C, nivells de CD4, casos de Sida i Morts) segons els diferents grups de transmissió del virus es va realitzar mitjançant el test de la chi quadrat de Pearson si no existien caselles amb freqüències esperades < 5.

Per a l'estimació de la funció de distribució del temps de supervivència (temps des de l'inici del TARGA fins al desenvolupament de sida i/o mort) es va considerar l'estimador de Kaplan-Meier. D'altra banda, la comparació de les corbes de supervivència segons el sexe, el grup de transmissió i el nivell de limfòcits CD4 es va realitzar mitjançant les proves log-rank, Breslow i Tarone-Ware, per tal de tenir en compte aquells casos en que les corbes de supervivència podrien presentar creuaments, és a dir, que no es complís la hipòtesi de proporcionalitat. Quan per aquelles variables amb més de dues categories les comparacions varen ser

significatives (grups de transmissió i nivells de CD4) es varen realitzar proves de comparacions múltiples (entre els parells de les categories), considerant la correcció de Bonferroni pel nivell de significació.

Finalment, l'estudi dels factors associats a la progressió a sida/mort es va realitzar mitjançant models de Cox semi-paramètrics univariables i multivariables. Totes aquelles variables significatives a un nivell de 0,025 per a l'anàlisi univariable es varen considerar potencials predictors pel model multivariable. La construcció del model multivariable final es va realitzar considerant models aniuats i es va comparar la seva significació mitjançant l'estadístic deviance basat en el logaritme de la versemblança. El diagnòstic de la validesa del model i l'anàlisi dels residus es va realitzar mitjançant els residus Cox-Snell, els residus de Schoenfeld i amb comprovacions gràfiques.

Els paquets estadístics utilitzats per l'anàlisi varen ser el SPSS, el SAS i el programa R pel diagnòstic dels residus dels models de Cox.

## 5 Resultats

---

### 5.1 Anàlisi descriptiu

Del total de 6.927 pacients reclutats a la Cohort PISCIS des de gener de 1998 fins a desembre 2004, 2.068 pacients varen complir els criteris d'inclusió definits per aquest estudi, amb un temps mig de seguiment de 31,5 mesos i rang interquartílic (RIC) de 12,4-53,3 mesos. A la **Taula 6** es descriuen algunes de les característiques clínico-epidemiològiques dels pacients inclosos en l'estudi a l'inici del tractament. Breument, del total 1.561 (75,5%) eren homes, 52,2% eren menors de 35 anys, la via d'infecció més freqüent va ser la sexual (35% i 27,4% a causa de relacions heterosexuals i homosexuals, respectivament) seguida de la d'ús de drogues per via parenteral (32,8%). La mediana del valor de limfòcits CD4 i còpies del virus a l'inici del tractament va ser de 259 cells/ $\mu$ l (RIQ: 139-389) i 70.000 (RIQ 20.000-193.625), respectivament. La prevalença global d'infecció per l'hepatitis C (HVC) va ser del 36,8%.

Durant el seguiment (**Taula 7**) es varen detectar 128 (6,2%) casos incidents de sida i 41 defuncions (2,0%). Dels 128 pacients que varen desenvolupar la sida durant el seguiment, 14 varen morir. Als 6 mesos de l'inici del tractament la mediana dels limfòcits CD4 va augmentar (383 cells/ $\mu$ l; RIQ 249,3-544,5) respecte el moment d'inici del tractament. Addicionalment, la càrrega viral va disminuir (mediana: 122 còpies/ml, RIQ: extrems 54,7-199).

La distribució de la proporció de pacients amb infecció de l'hepatitis C (HVC) i amb CD4 inferiors a 200 cells/ $\mu$ l segons grup de transmissió de l'HIV es mostra a la **Taula 8**. La prova de chi quadrat de Pearson va mostrar que existien diferències significatives en quant a la proporció de pacients amb HVC ( $p < 0,001$ ), sent el grup dels UDVP el que va presentar una proporció més elevada (89%). D'altra banda, les diferències en quant a la proporció de pacients amb nivells de  $CD4 < 200$  va resultar en el llindar de la significació estadística ( $p = 0,054$ ), tot i que el grup dels UDVP també varen presentar la proporció més elevada (40%).

Anàlogament, la distribució dels casos incidents de sida i les defuncions segons grup de transmissió es mostren a la **Taula 9**. La prova de chi quadrat de Pearson no va mostrar diferències significatives en el nombre de casos de sida segons el grup de transmissió ( $p = 0,538$ ). D'altra banda, es varen trobar diferències significatives en quant a la proporció de defuncions ( $p = 0,004$ ) sent el grup de UDVP (3,4%) i el d'altres vies de transmissió (4,9%) els que varen presentar unes proporcions més elevades.

## 5.2 Estimació i comparació de les corbes de supervivència mitjançant l'estimador Kaplan-Meier.

L'estimació de les funcions de supervivència mitjançant l'estimador de Kaplan-Meier segons els diferents nivells dels factors sexe, grup de transmissió i categoria de valors CD4 a l'inici del tractament es mostren a les **Figures 1, 2 i 3**, respectivament. Aquestes mostren que el pitjor pronòstic el presentaven els homes, els pacients infectats per ús de drogues per via parenteral i aquells que varen iniciar tractament amb CD4<200. La proporció del temps lliure de sida/mort en els homes va resultar estadísticament diferent que en les dones (test log-rank,  $p<0,006$ ). Anàlogament les tres proves de comparació de la distribució del temps de supervivència varen mostrar que existien diferències segons el grup de CD4 a l'inici del tractament ( $p<0,001$ ). Aquestes diferències es varen detectar entre els grups <200 i 200-350 ( $p<0,001$ ) i també entre els de <200 i >350 CD4 ( $p<0,001$ ) (**Figura 2**). Per contra, la distribució del temps de supervivència segons grup de transmissió va resultar estadísticament diferent considerant el test del log-rank ( $p<0,034$ ), mentre que no es varen detectar diferències segons les proves de Breslow i Tarone-Ware ( $p>0,15$ ).

## 5.3 Factors associats al risc de progressió a sida o mort. Models de regressió de Cox.

Per determinar els factors associats a un major risc de progressar cap a l'event (sida/mort) es varen realitzar models de Cox bivariats i multivariats (**Taula 10**). Els models univariats varen mostrar que el risc de progressió a sida/mort en pacients entre 35-54 anys va ser de 1,6 (IC 95%: 1,1-2,2) respecte als <35 anys, mentre que en els pacients amb més de 55 anys no es varen trobar diferències estadísticament significatives respecte als més joves. Anàlogament, els homes tenien el doble de risc respecte a les dones (IC 95%: 1,2-2,8). Per la banda de les variables clínic-biològiques varen mostrar més risc els pacients amb nivells de CD4 <200 (IC 95%: 2,7-6,8) i els pacients amb una càrrega viral superior a les 100.000 còpies varen mostrar el doble de risc que la resta (IC 95%: 1,6-3,1). Els pacients infectats pel virus de l'hepatitis C varen mostrar el doble de risc respecte als no infectats (IC 95%: 1,5-2,9). El model multivariats va resultar amb les variables sexe, nivells de CD4, nivells de CV i hepatitis C incloses en el model, i per tant, es van considerar com a variables significatives al temps de supervivència.

## 5.4 Taules i figures

**Taula 6. Característiques clínico-epidemiològiques a l'inici del tractament. Cohort PISCIS (1998-2004).**

	Nombre de pacients (%)
<b>Total de pacients</b>	2.068 (100%)
<b>Variables socio-demogràfiques</b>	
<b>Edat</b>	
< 35 anys	1.079 (52,2)
35-54 anys	869 (42,1)
≥ 55 anys	120 (5,8)
<b>Sexe (home) (n=2067)</b>	1.561 (75,5)
<b>Grup de transmissió (n=2020)</b>	
UDVP	678 (32,8)
Home homosexual/bisexual	566 (27,4)
Heterosexual	723 (35,0)
Altres	41 (2,0)
<b>Variables clínico-biològiques a l'inici del tractament</b>	
<b>Càrrega Viral (CV) copies/ml (n=1968)</b>	70.000 (20.000-193.625,5)*
<b>Càrrega Viral (CV) (n=1968)</b>	
≤ 500 copies/ml	83 (4,2)
> 500 copies/ml	1885 (95,8)
<b>CD4 cells/μl</b>	259 (139-389)*
<b>CD4</b>	
< 200 cells/μl	773 (37,4)
200-350 cells/μl	653 (31,6)
> 350 cells/μl	642 (31,0)
<b>Pauta de TARGA inicial</b>	
1 o 2 IP + 2 INT	840 (40,6)
2 INT + 1 INNT	924 (44,7)
3 INT	168 (8,1)
Altres	136 (6,6)
<b>Període d'inici del TARGA (2001-2004)</b>	1.115 (53,9)
<b>Hepatitis C (HVC) (n=1801)</b>	662 (36,8)

UDVP: Usuaris de drogues per via parenteral

**Taula 7. Característiques clínico-epidemiològiques als 6 mesos de l'inici del tractament i durant el seguiment. Cohort PISCIS (1998-2004).**

	Nombre de pacients (%)
<b>Total de pacients</b>	2.068 (100%)
<b>Variabls clínic-biològiques als 6 mesos:</b>	
<b>CD4</b> (n=1499)	367 (234-530)*
<b>CD4</b> (n=1499)	
< 200 cells/ $\mu$ l	283 (18,9)
200-350 cells/ $\mu$ l	421 (28,1)
> 350 cells/ $\mu$ l	795 (53,1)
<b>CV</b> (n=1499)	166 (53-200)*
<b>Càrrega Viral (CV)</b> (n=1499)	
$\leq$ 500 copies/ml	1249 (83,3)
> 500 copies/ml	250 (16,7)
<b>Variabls de seguiment</b>	
<b>Casos incidents de sida</b>	128 (6,2)
<b>Morts</b>	41 (2,0)
<b>Events (sida o mort)</b>	155 (7,5)
<b>Temps de seguiment</b> (en mesos)	31,5 (12,4-53,3)

\* Mediana (Rang interquartil·lic)

**Taula 8. Distribució d'algunes característiques clínic-biològiques segons grup de transmissió de l'HIV <sup>(1)</sup>.**

	Grup de transmissió					P <sup>(2)</sup>
	UDVP	Homes Homo-/bisexual	Heterosexual	Altres	Total	
	n=678	n=566	n=723	n=41	n=2008	
<b>Hepatitis C</b>	504/568 (89%)	29/513 (5,7%)	112/642 (17,4%)	12/36 (33,4%)	657/1759 (37,5%)	<0,001
<b>CD4&lt;200</b>	268/678 (39,5%)	187/566 (33,7%)	278/723 (38,5%)	19/41 (46,3%)	752/2008 (37,5%)	0,054

**UDVP: Usuaris de drogues per via parenteral**

**(1) Els denominadors indiquen el nombre de casos efectius**

**(2) p-valor del test Chi-quadrat de Pearson**

**Taula 9. Distribució dels casos de sida incidents i les defuncions segons grup de transmissió de l'HIV <sup>(1)</sup>.**

	Grup de transmissió					P <sup>(2)</sup>
	UDVP	Homes Homo-/bisexual	Heterosexual	Altres	Total	
	n=678	n=566	n=723	n=41	n=2008	
<b>Casos de sida</b>	50 (7,4%)	33 (5,8%)	41 (5,7%)	2 (4,9%)	126 (6,3%)	0,538
<b>Morts</b>	23 (3,4%)	5 (0,9)	10 (1,4)	2 (4,9%)	40 (2,0%)	0,004

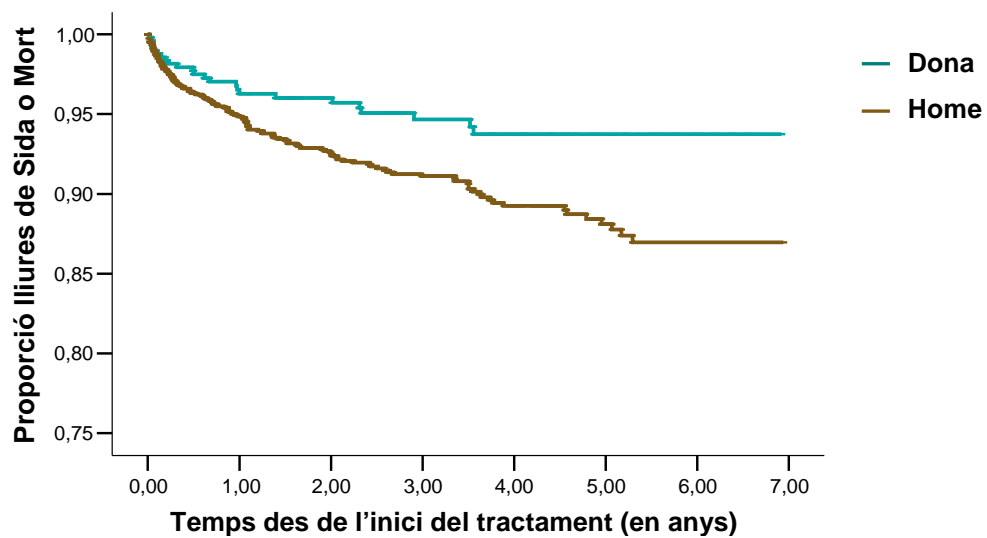
**UDVP: Usuaris de drogues per via parenteral**

**(1) Els denominadors indiquen el nombre de casos efectius**

**(2) p-valor del test Chi-quadrat de Pearson**



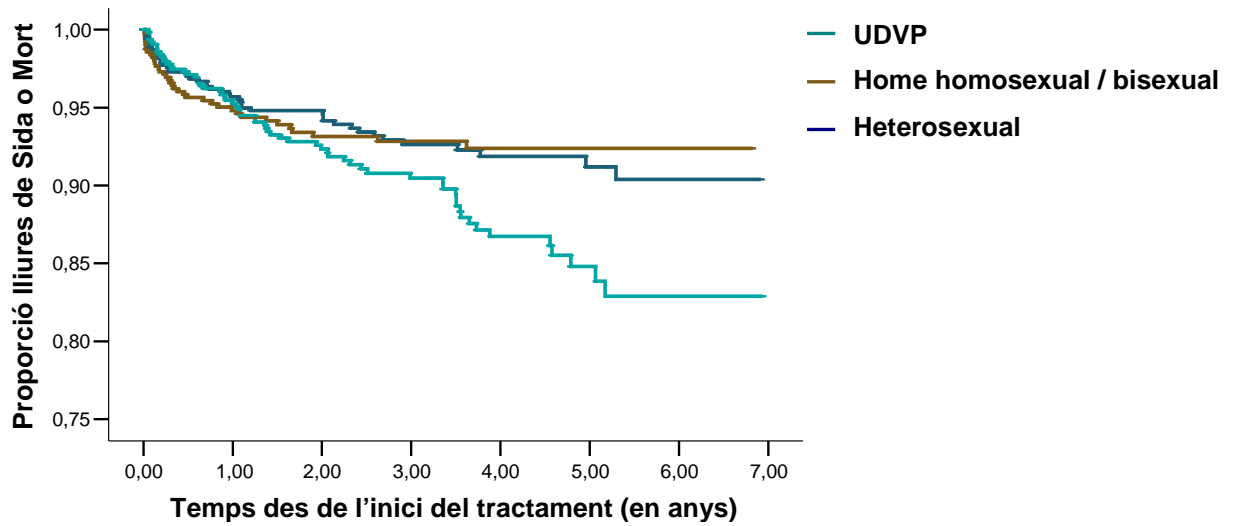
**Figura 1. Comparació de les corbes de supervivència segons sexe.  
Estimador de Kaplan-Meier.**



	Valor	g.ll.*	p
<b>Log Rank</b>	7,520	1	0,006
<b>Breslow</b>	5,751	1	0,016
<b>Tarone-Ware</b>	6,541	1	0,011

\* graus de llibertat de la prova.

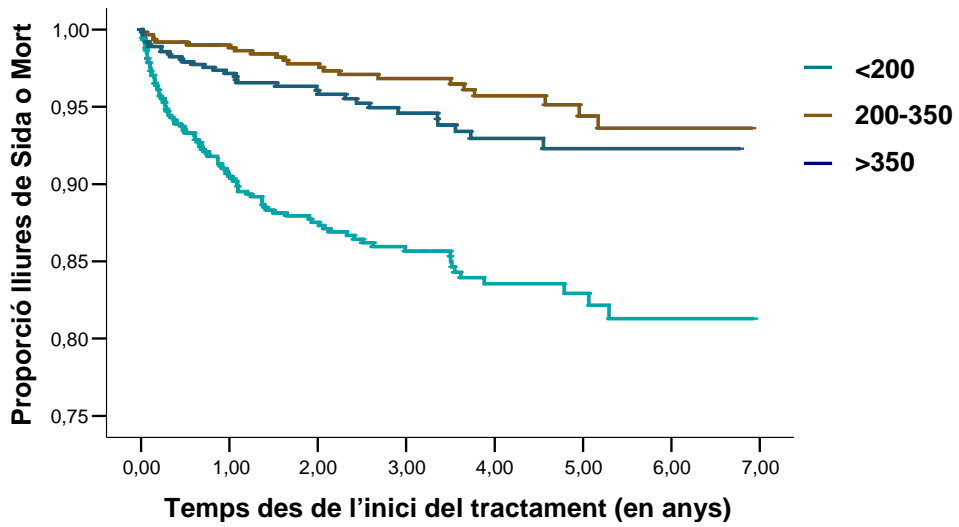
**Figura 2. Comparació de les corbes de supervivència segons grup de transmissió. Estimador de Kaplan-Meier**



	Valor	g.ll.*	p
<b>Log Rank</b>	6,784	2	0,034
<b>Breslow</b>	2,145	2	0,342
<b>Tarone-Ware</b>	3,776	2	0,151

\* graus de llibertat de la prova.

**Figura 3. Comparació de les corbes de supervivència segons grups de valors de linfocits CD4. Estimador de Kaplan-Meier**



	Valor	g.ll.*	p
<b>Log Rank</b>	60,589	2	<0,001
<b>Breslow</b>	65,996	2	<0,001
<b>Tarone-Ware</b>	64,970	2	<0,001

\* grau de llibertat de la prova.

**Taula 10. Factors associats al risc de progressió de l'HIV a Sida o Mort.  
Models de Cox univariables.**

	n	Sida/Mort	Risc	95% IC	p
<b>Grups d'edat</b>					0,019
< 35 anys	1016	59	1	- -	-
35-54 anys	924	83	1,6	(1,1-2,2)	0,008
≥ 55 anys	128	13	1,7	(0,9-3,1)	0,075
<b>Sexe</b>					0,007
Dona	506	24	1	- -	-
Home	1561	131	1,8	(1,2-2,8)	0,007
<b>Grup de transmissió</b>					0,081
Altres	41	4	1	- -	-
UDVP	678	64	1,1	(0,4-3,1)	0,832
Home homo/bisexual	566	37	0,7	(0,3-2,1)	0,537
Heterosexual	723	47	0,7	(0,3-2,1)	0,546
<b>CD4 (cells/μl)</b>					<0,001
>350	642	22	1	- -	-
<200	766	100	4,3	(2,7-6,8)	<0,001
200-350	660	33	1,6	(0,9-2,7)	0,101
<b>CV</b>					<0,001
<100000	1148	58	1	- -	-
≥100000	820	85	2,2	(1,6-3,1)	<0,001
<b>Pauta TARGA</b>					0,145
1 o 2 IP + 2 INT	840	82	1	- -	-
2 INT + 1 INNT	924	54	1,2	(0,6-2,7)	0,609
3 INT	168	12	0,8	(0,4-1,8)	0,617
Altres	136	7	1,2	(0,5-3,0)	0,769
<b>Era TARGA</b>					0,066
Era "early"	953	99	1	- -	-
Era "late"	1115	56	0,7	(0,5-1,1)	0,066
<b>Hepatitis C</b>					<0,001
No	1139	61	1	- -	-
Si	662	69	2,1	(1,5-2,9)	<0,001

**Taula 11. Factors associats al risc de progressió de l'HIV a Sida o Mort.  
Models de Cox multivariable i model final**

	Anàlisi multivariable			Anàlisi multivariable (model final)		
	Risc	95% IC	p	HR*	95% CI**	p
<b>Grups d'edat</b>			0,196			
< 35 anys	1	- -	-	-	- -	-
35-54 anys	1,3	(0,9-1,9)	0,215	-	- -	-
≥ 55 anys	1,7	(0,9-3,5)	0,094	-	- -	-
<b>Sexe</b>			0,138			0,047
Home	1	- -	-	1	- -	-
Dona	0,7	(0,4-1,1)	0,138	1,7	(1,0-2,8)	0,047
<b>Grup de transmissió</b>			0,990			
Altres	1	- -	-	-	- -	-
UDVP	1,1	(0,6-2,1)	0,745	-	- -	-
Home homo/bisexual	1,0	(0,6-1,8)	0,880	-	- -	-
Heterosexual	1,1	(0,3-3,6)	0,903	-	- -	-
<b>CD4 (cells/μl)</b>			<0,001			<0,001
>350	1	- -	-	1	- -	-
<200	0,5	(0,3-0,8)	0,002	4,0	(2,4-7,0)	<0,001
200-350	0,3	(0,2-0,5)	<0,001	1,9	(1,1-3,6)	0,031
<b>CV</b>			0,001			0,002
<100000	1	- -	-	1	- -	-
≥100000	1,9	(1,3-2,8)	0,001	1,8	(1,3-2,6)	0,002
<b>Pauta TARGA</b>			0,223			
1 o 2 IP + 2 INT	1	- -	-	-	- -	-
2 INT + 1 INNT	1,1	(0,4-2,8)	0,835	-	- -	-
3 INT	0,9	(0,3-2,3)	0,795	-	- -	-
Altres	1,8	(0,6-5,3)	0,286	-	- -	-
<b>Era TARGA</b>			0,185			
Era "early"	1	- -	-	-	- -	-
Era "late"	0,8	(0,5-1,1)	0,185	-	- -	-
<b>Hepatitis C</b>			0,005			<0,001
No	1	- -	-	1	- -	-
Si	2,1	(1,2-3,4)	0,005	2,0	(1,4-2,8)	<0,001

**Figura 4: Residus de Cox i Snell**

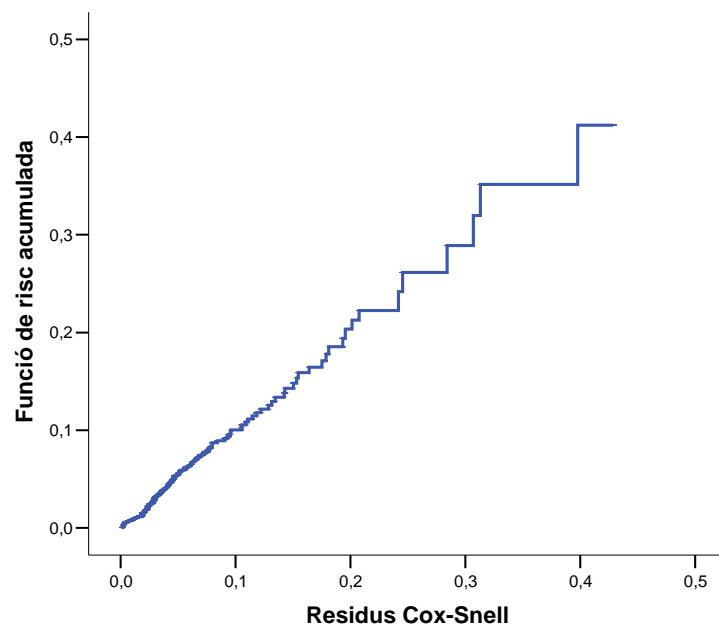
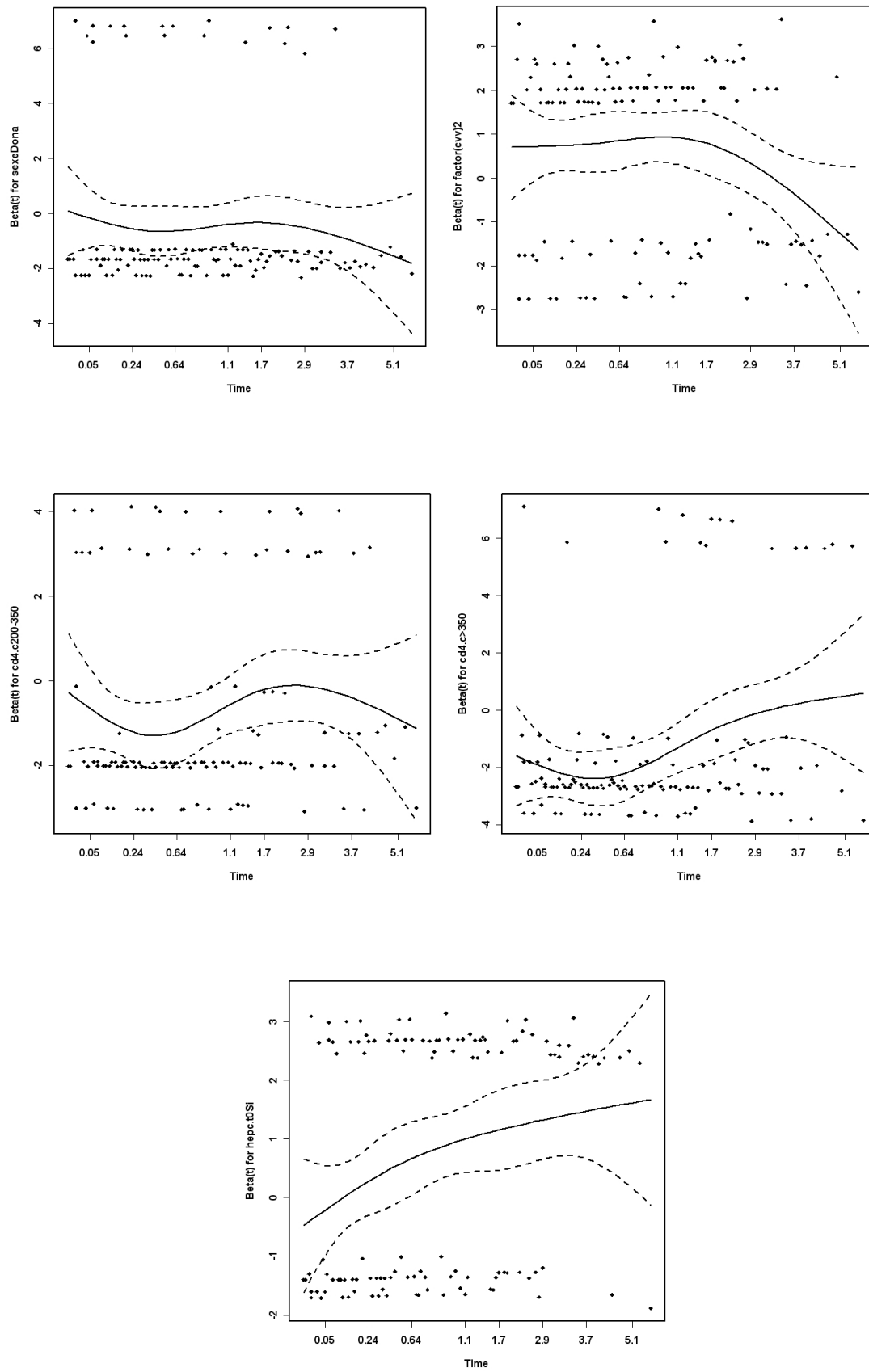


Figura 5: Residus de Schoenfeld



## 6 Discussió

---

### 6.1 Conclusions

L'estudi dels factors de risc associats a la progressió de la malaltia de VIH ens confirma que els pacients amb pitjor pronòstic són aquells que varen iniciar tractament amb nivells de CD4<200 i entre 200 i 350 respecte als que varen iniciar el tractament amb >350 limfòcits. D'altra banda, els pacients que presentaven >100.000 còpies també presentaven pitjor pronòstic. Anàlogament, els pacients que a l'inici del tractament ja estaven infectats per l'hepatitis C tenien més risc de progressar cap a la malaltia o morir. En canvi, el període en el que els pacients iniciaven el tractament no va resultar un factor de risc associat al temps de supervivència.

### 6.2 Limitacions

L'anàlisi efectuat presenta diverses limitacions per a les quals no es poden arribar a conclusions més congruents.

D'una banda, a partir d'aquests models no es pot inferir en trobar relacions causals entre els factors de risc i el temps de supervivència, ja que es tracta d'una cohort prevalent i no pas d'un estudi aleatoritzat (per exemple, un assaig clínic).

En l'anàlisi tampoc s'han utilitzat variables dependents del temps ja que aleshores l'anàlisi adquireix una complexitat més elevada que no es correspon amb l'objectiu d'aquest projecte. Com a conseqüència d'això, els resultats trobats en relació al període d'inici de TARGA no són concloents i caldria considerar aquesta variable com una variable dependent del temps en el model de Cox.

Donat que la hipòtesis de proporcionalitat dels riscos del model de Cox no es va complir per algunes de les variables predictorres del model multivariable final, caldria valorar la possibilitat de realitzar transformacions d'aquestes variables o considerar models alternatius, com són els models paramètrics per l'estudi del temps de supervivència.



## 7 Sumari de taules i figures

---

<b>Taula 1.</b> Codificació de la variable Tractament Previ	16
<b>Taula 2.</b> Codificació de la variable Pauta Tractament	16
<b>Taula 3.</b> Codificació de la variable Estadiatge	17
<b>Taula 4.</b> Codificació de la variable Baixa	19
<b>Taula 5.</b> Variables d'estudi	24
<b>Taula 6.</b> Característiques clínico-epidemiològiques a l'inici del tractament. Cohort PISCIS (1998-2004).	28
<b>Taula 7.</b> Característiques clínico-epidemiològiques als 6 mesos de l'inici del tractament i durant el seguiment. Cohort PISCIS (1998-2004).	29
<b>Taula 8.</b> Distribució d'algunes característiques clínico-biològiques segons grup de transmissió de l'HIV	30
<b>Taula 9.</b> Distribució dels casos de sida incidents i les defuncions segons grup de transmissió de l'HIV	30
<b>Figura 1.</b> Corbes de supervivència per sexe	31
<b>Figura 2.</b> Corbes de supervivència per grup de transmissió	32
<b>Figura 3.</b> Corbes de supervivència per nivell de CD4	33
<b>Taula 10.</b> Factors associats al risc de progressió de l'HIV a Sida o Mort. Models de Cox univariables.	34
<b>Taula 11.</b> Factors associats al risc de progressió de l'HIV a Sida o Mort. Models de Cox multivariable i model final	35
<b>Figura 4.</b> Residus de Cox i Snell	36
<b>Figura 5.</b> Residus de Schoenfeld	37

## 8 Bibliografia

---



- [1] Benavides, J.; Olaizola, J.M.; Rivero, E. *SQL para usuarios i programadores*. Madrid: Paraninfo, 1992.
- [2] Cobo, E.; Calle, M.L.; Videla, S.; et.al. *Investigación clínica y estadística. Una visión multidisciplinar en estudios de VIH/Sida*. Badalona: Fundació Lluita contra la Sida, 2003.
- [3] Comité de Bioètica de Catalunya. *Consideracions ètiques entorn de la informació sanitària*. Generalitat de Catalunya, departament de Salut. Barcelona, 2004.
- [4] Doménech, J.M. *Fundamentos de diseño y estadística. Descripción de datos de supervivencia: métodos de Kaplan-Meier y Actuarial*. Universitat Autònoma de Barcelona. Barcelona: Signo, 2004.
- [5] Dorca, E.; Sued, O.; Masabeu, A.; Gatell, J.M.; Casabona, J.; Riera, M.; Segura, F.; Altès, J.; Miró, J.M.; Tural, C.; Ferrer, E.; Podzamczer, A.; Jaén, A.; Esteve, A.; García, I.; Villalonga, C.; Vilaró, J.; Navarro, G.; Clotet, B.; Force, L. *Características clínicoepidemiológicas y tendencias en el tratamiento antirretroviral de una cohorte de pacientes con infección por el virus de la inmunodeficiencia humana. Cohorte PISCIS*. Medicina clínica, ISSN 0025-7753, Vol. 124, Nº. 14, 2005, pags. 525-531
- [6] *European Centre for the Epidemiological Monitoring of AIDS. HIV/AIDS surveillance in Europe. End-year report 2002*. Saint-Maurice: Institut de Veille Sanitaire; 2003.
- [7] Fistera. Atención Primaria de Salud en la red. [Consulta: 17 juny 2007]. Disponible a: [http://www.fistera.com/mbe/investiga/supervivencia/analisis\\_supervivencia2.pdf](http://www.fistera.com/mbe/investiga/supervivencia/analisis_supervivencia2.pdf)
- [8] FLS. *Fundació Lluita contra la Sida*. [en línia]. [Consulta: 12 novembre 2006]. Disponible a: <http://www.flsida.org>
- [9] Gaspar, G.; López-Gay, D.; Castilla, V.; Condés, E.; por el Grupo de estudio de la cohorte COMESEM. *Características clínicoepidemiológicas de los pacientes recién diagnosticados de infección por el VIH*. Medicina Clínica. Barcelona, 2002;119:238-9.
- [10] Gómez, G.; Langohr, K. *Anàlisis de supervivència*. Universitat Politècnica de Catalunya. Barcelona, 2004.
- [11] Lafler, K.P. *PROC SQL: beyond the basics usign SAS*. SAS Institute Inc. North Carolina, 2004.





- [12] Langohr, K.; Serrat, C.; Gómez, G. *Tutorial de análisis de supervivencia con R*. Universitat Politècnica de Catalunya. Barcelona, 2007.
- [13] Mocroft, A.; Ledergerber, B.; Katlama, C.; Kirk, O.; Reiss, P.; D'Arminio, A.; et al, for the EuroSIDA study group. *Decline in the AIDS and death rates in the EuroSIDA study: an observational study*. Lancet 2003;362:22-9.
- [14] Moore, R.; Chaisson, E. *Natural history of HIV infection in the era of combination antiretroviral therapy*. AIDS. 1999;13:1933-42.
- [15] *Open Door Clinic*. [en línia]. [Consulta: 12 novembre 2006]. Disponible a: <<http://www.opendoorclinic.org>>
- [16] Palella, F.J.; Delaney, K.M.; Moorman, A.C.; Loveless, M.O.; Fuhrer, J.; Satten, G.A.; et al. *Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection*. N Engl J Med. 1998;338:853-60.
- [17] Pardo, F. *VIH y SIDA*. [en línia]. [Consulta: 12 novembre 2006]. Disponible a: <http://www.ctv.es/USERS/fpardo/home.html>
- [18] Pérez, C. *El sistema estadístico SAS*. Pearson Educación. Madrid, 2001.
- [19] Prairie, K. *The Essential PROC SQL Handbook*. For SAS Users. SAS Institute Inc. SAS Campus Drive, North Carolina, 2005
- [20] Rothman, K. *Epidemiología moderna*. Ediciones Diaz de Santos, S. A. Madrid, 1987.
- [21] Tarwater, P.M.; Mellors, J.; Gore, M.E.; Margolick, J.B.; Phair, J.P.; Detels, R.; et al. *Methods to assess population effectiveness of therapies in human immunodeficiency virus incident and prevalent cohorts*. Am J Epidemiol. 2001;154:675-81.



## ANNEXOS

---

## A 1: RELACIÓ DE TAULES DE LA BASE DE DADES PISCIS

Nom de la taula	Nom de les variables	Tipus de variable
 mcd_altes	Id	Identificador
	Procedencia	Categòrica
	Data_naixement	Data
	Sexe	Categòrica
	Provincia_naixement	Categòrica
	Grup_transmissio	Categòrica
	Formacio	Categòrica
	Laboral	Categòrica
	Data_positiu	Data
	Data_primer_a_visita	Data
	Ultim_contacte	Data
	Dies_seguiment	Numèrica
	Baixa	Categòrica
	Tractament_previ	Categòrica
	Data_primer_tractament	Data
	Pauta_TARGA_inicial	Categòrica
	Data_tract_CD4	Data
	Valor_tract_CD4	Numèrica
	CD4	Numèrica
	Data_tract_CV	Data
	Valor_tract_CV	Numèrica
	CV	Numèrica
	Dies_fins_seguiment	Numèrica
Hepatitis_C	Categòrica	
 mcd_seguiments	Id	Identificador
	Procedencia	Categòrica
	Data_visita	Data
	Pes	Numèrica
	Karnofsky	Categòrica
	Estadiatge	Categòrica
	Ginecologia	Categòrica
	Tipus_visita	Categòrica
	UDVP	Categòrica
	Sida	Categòrica

 mcd_analitiques	Id	Identificador
	Procedencia	Categòrica
	Data_analitica	Data
	CD4	Numèrica
	CD4p	Numèrica
	CD8	Numèrica
	CD8p	Numèrica
	CV	Numèrica
	Cvlog	Numèrica
	Ppd	Categòrica
	Achbs	Categòrica
	Aghbs	Categòrica
	Achbc	Categòrica
	Rna_vhc	Categòrica
	Acvhcr	Categòrica
Acvhce	Categòrica	
 mcd_tractaments	Id	Identificador
	Procedencia	Categòrica
	Data_tractament	Data
	Arv1	Categòrica
	Arv2	Categòrica
	Arv3	Categòrica
	Arv4	Categòrica
	Arv5	Categòrica
	Arv6	Categòrica
	Arv7	Categòrica
	Resistencia	Categòrica
	Total	Categòrica
	Pauta	Categòrica
 Mcd_Malalties_rel	Id	Identificador
	Procedencia	Categòrica
	Data_mal_relacionada	Data
	Malaltia_relacionada	Categòrica
 Mcd_Malalties_def	Id	Identificador
	Procedencia	Categòrica
	Data_mal_definitoria	Data
	Malaltia_definitoria	Categòrica

 mcd_Ingressos	Id	Identificador
	Procedencia	Categòrica
	Data_ingres	Data
	Motiu_ingres	Categòrica
 mcd_baixes	Id	Identificador
	Procedencia	Categòrica
	Data_baixa	Data
	Motiu_baixa	Categòrica
	Motiu_exitus	Categòrica
	Causa_exitus	Categòrica
	Exitus_HIV	Categòrica

## A 2.1: PROGRAMA EN SQL DE PREPARACIÓ DE LES DADES

---

```
/* _____ */
/* _____ */
/*     PROGRAMA PREPARACIÓ DE LES DADES AMB SQL HOSTATJAT AL SAS     */
/* _____ */

/*Es crea la llibreria*/

LIBNAME pfc_as "E:\PFC-PISCIS\Analisi\En SAS\Llibreria2";

/*Inici de la importació de les dades en format ACCESS a SAS*/

proc import table="mcd_altes"
out=pfc_as.mcd_altes
dbms=access;
database="E:\PFC-PISCIS\Analisi\dades_piscis.mdb";
run;

proc import table="mcd_analitiques"
out=pfc_as.mcd_analitiques
dbms=access;
database="E:\PFC-PISCIS\Analisi\dades_piscis.mdb";
run;

proc import table="mcd_baixes"
out=pfc_as.mcd_baixes
dbms=access;
database="E:\PFC-PISCIS\Analisi\dades_piscis.mdb";
run;

proc import table="mcd_ingressos"
out=pfc_as.mcd_ingressos
dbms=access;
database="E:\PFC-PISCIS\Analisi\dades_piscis.mdb";
run;

proc import table="mcd_malalties_def"
out=pfc_as.mcd_malalties_def
dbms=access;
database="E:\PFC-PISCIS\Analisi\dades_piscis.mdb";
run;

proc import table="mcd_malalties_rel"
out=pfc_as.mcd_malalties_rel
dbms=access;
database="E:\PFC-PISCIS\Analisi\dades_piscis.mdb";
run;

proc import table="mcd_tractaments"
out=pfc_as.mcd_tractaments
dbms=access;
database="E:\PFC-PISCIS\Analisi\dades_piscis.mdb";
run;

proc import table="mcd_seguiments"
out=pfc_as.mcd_seguiments
```



```

dbms=access;
database="E:\PFC-PISCIS\Analisi\dades_piscis.mdb";
run;

/*Consulta_criteri_1*/

proc sql;
create table pfc_as.naive as
SELECT mcd_altes.Id, datepart(mcd_altes.Data_naixement)AS
Data_naixement, mcd_altes.Sexe, mcd_altes.Grup_transmissio,
datepart(mcd_altes.Data_primera_visita) AS Data_primera_visita,
datepart(mcd_altes.Data_tract_CV)AS Data_tract_CV, mcd_altes.CV,
datepart(mcd_altes.Data_tract_CD4)AS Data_tract_CD4, mcd_altes.CD4,
datepart(mcd_altes.Data_primer_tractament) AS Data_primer_tractament,
mcd_altes.Pauta_TARGA_inicial, mcd_altes.Hepatitis_C, mcd_altes.Baixa,
datepart(mcd_altes.ultim_contacte)AS ultim_contacte
FROM pfc_as.mcd_altes
WHERE mcd_altes.Tractament_previ=2;
quit;

data pfc_as.naive;
set pfc_as.naive;
format data_naixement date.;
format data_primera_visita date.;
format data_cv_ttr date.;
format data_cd4_ttr date.;
format data_primer_tractament date.;
format ultim_contacte date.;
run;

/*Consulta_criteri_2*/

proc sql;
create view pfc_as.inicien_TARGA as
SELECT *
FROM pfc_as.naive
WHERE (Data_primer_tractament) Is Not Null And (Pauta_TARGA_inicial)
Is Not Null And Pauta_TARGA_inicial>=30;
quit;

/*Consulta_criteri_3*/

/*Consulta_criteri_3.1*/

proc sql;
create table pfc_as.malaltia_def as
SELECT mcd_malalties_def.Id,
Min(datepart(mcd_malalties_def.Data_mal_definitoria))AS Mdef_ant_ttm
FROM pfc_as.mcd_Malalties_def LEFT JOIN pfc_as.mcd_altes ON
mcd_altes.Id=mcd_malalties_def.Id
WHERE Data_mal_definitoria<=Data_primer_tractament
GROUP BY mcd_malalties_def.Id;
quit;

```

```

data pfc_as.malaltia_def;
set pfc_as.malaltia_def;
format Mdef_ant_ttm date.;
run;

/*Consulta_criteri_3.2*/

proc sql;
create table pfc_as.EstadiatgeC as
SELECT mcd_seguiments.Id, Min(datepart(mcd_seguiments.Data_visita)) AS
EstC_ant_ttm
FROM pfc_as.mcd_seguiments LEFT JOIN pfc_as.mcd_altes ON
mcd_seguiments.Id=mcd_altes.Id
WHERE (((mcd_seguiments.estadiatge)=7 Or (mcd_seguiments.estadiatge)=8
Or (mcd_seguiments.estadiatge)=9) AND
((mcd_seguiments.Data_visita)<=Data_primer_tractament))
GROUP BY mcd_seguiments.Id;
quit;

data pfc_as.EstadiatgeC;
set pfc_as.EstadiatgeC;
format EstC_ant_ttm date.;
run;

/*Consulta_criteri_3.3: s'obté la informació total*/

proc sql;
create view pfc_as.Sida_no_inici AS
SELECT mcd_altes.Id, malaltia_def.Mdef_ant_ttm,
estadiatgeC.EstC_ant_ttm
FROM (pfc_as.mcd_altes LEFT JOIN pfc_as.malaltia_def ON
mcd_altes.Id=malaltia_def.Id) LEFT JOIN pfc_as.estadiatgeC ON
mcd_altes.Id=estadiatgeC.Id
WHERE ((Mdef_ant_ttm) Is Null And (EstC_ant_ttm) Is Null);
quit;

/*Consulta_criteri_3.4*/

proc sql;
create view pfc_as.No_sida_inici AS
SELECT Inicien_TARGA.*
FROM pfc_as.Inicien_TARGA INNER JOIN pfc_as.sida_no_inici ON
inicien_TARGA.Id=sida_no_inici.Id;
quit;

/*Consulta_criteri_4*/

proc sql;
create view pfc_as.minim_seg_CD4 AS
SELECT No_sida_inici.*
FROM pfc_as.No_sida_inici
WHERE (Data_primer_tractament) <= '30JUN2004'd And (CD4) Is Not Null;
quit;

```

```

/*Consulta_criteri_5*/

proc sql;
create view pfc_as.inclosos AS
SELECT minim_seg_CD4.*
FROM pfc_as.minim_seg_CD4
WHERE baixa<>0 AND (data_primera_visita-data_naixement)/365.25>=16 AND
Data_naixement Is Not Null;
quit;

/* CREACIÓ DE LES VARIABLES DE TEMPS DE SUPERVIVÈNCIA I CENSURA */

/*Consulta_objectiu_1*/

/*Consulta_objectiu_1.1_a*/

proc sql;
create table pfc_as.mdef_no_inici AS
SELECT mcd_malalties_def.Id,
Min(datepart(mcd_malalties_def.Data_mal_definitoria)) AS Mdef_pos_ttm
FROM pfc_as.mcd_Malalties_def LEFT JOIN pfc_as.mcd_altes ON
mcd_altes.Id=mcd_malalties_def.Id
WHERE Data_mal_definitoria>Data_primer_tractament AND
Data_primer_tractament Is Not Null
GROUP BY mcd_malalties_def.Id;
quit;

data pfc_as.mdef_no_inici;
set pfc_as.mdef_no_inici;
format Mdef_pos_ttm date.;
run;

/*Consulta_objectiu_1.1_b*/

proc sql;
create table pfc_as.EstC_no_inici AS
SELECT mcd_seguiments.Id, Min(datepart(mcd_seguiments.Data_visita)) AS
EstC_pos_ttm
FROM pfc_as.mcd_seguiments LEFT JOIN pfc_as.mcd_altes ON
mcd_seguiments.Id=mcd_altes.Id
WHERE (((mcd_seguiments.estadiatge)=7 OR (mcd_seguiments.estadiatge)=8
Or (mcd_seguiments.estadiatge)=9) AND
((mcd_seguiments.Data_visita)>Data_primer_tractament)) AND
Data_primer_tractament Is Not Null
GROUP BY mcd_seguiments.Id;
quit;

data pfc_as.EstC_no_inici;
set pfc_as.EstC_no_inici;
format EstC_pos_ttm date.;
run;

```

```

/*Consulta_objectiu_1.2*/

proc sql;
create view pfc_as.amb_sida as
SELECT Inclosos.*, Mdef_pos_ttm, EstC_pos_ttm
FROM (pfc_as.Inclosos LEFT JOIN pfc_as.Mdef_no_inici ON
inclosos.Id=mdef_no_inici.Id) LEFT JOIN pfc_as.EstC_no_inici ON
inclosos.Id=estc_no_inici.Id;
quit;

/*Consulta_objectiu_2*/

proc sql;
create view pfc_as.fan_mort as
SELECT Id, Data_baixa
FROM pfc_as.mcd_baixes
WHERE Motiu_baixa=3;
quit;

/*Consulta_objectiu_3*/

proc sql;
create table pfc_as.taula as
SELECT Amb_Sida.*, Datepart(Data_baixa) AS Data_mort
FROM pfc_as.Amb_Sida LEFT JOIN pfc_as.Fan_mort ON
Amb_Sida.Id=Fan_mort.Id;
quit;
data pfc_as.taula;
set pfc_As.taula;
format data_mort date.;
run;
/* S'afegeixen columnes a la taula creada anteriorment.*/
proc sql;
ALTER TABLE pfc_as.taula
add data_sida date, data_sida_mort date, censura int,
temps_sida_mort_dies int;
quit;

/*Consulta_objectiu_4*/

/*Consulta_objectiu_4.1*/

proc sql;
UPDATE pfc_as.taula SET data_sida = mdef_pos_ttm
WHERE ((mdef_pos_ttm<=EstC_pos_ttm) AND (mdef_pos_ttm) Is Not Null AND
(estc_pos_ttm) Is Not Null) OR ((mdef_pos_ttm) Is Not Null AND
(EstC_pos_ttm) Is Null);
quit;

```

```

/*Consulta_objectiu_4.2*/

proc sql;
UPDATE pfc_as.taula SET data_sida = EstC_pos_ttm
WHERE ((mdef_pos_ttm>=EstC_pos_ttm) And (mdef_pos_ttm) Is Not Null And
(estc_pos_ttm) Is Not Null) Or ((mdef_pos_ttm) Is Null And
(EstC_pos_ttm) Is Not Null);
quit;

/*Consulta_objectiu_5*/

/*Consulta_objectiu_5.1*/

proc sql;
UPDATE pfc_as.taula SET data_sida_mort = data_sida
WHERE ((data_sida<=data_mort) AND (data_sida) Is Not Null AND
(data_mort) Is Not Null) OR ((data_sida) Is Not Null AND (data_mort)
Is Null);
quit;

/*Consulta_objectiu_5.2*/

proc sql;
UPDATE pfc_as.taula SET data_sida_mort = data_mort
WHERE ((data_sida>=data_mort) And (data_sida) Is Not Null And
(data_mort) Is Not Null) Or ((data_sida) Is Null And (data_mort) Is
Not Null);
quit;

/*Consulta_objectiu_6*/

/*Consulta_objectiu_6.1*/
proc sql;
UPDATE pfc_as.taula SET censura = 1
WHERE (data_sida_mort) Is Not Null;
quit;

/*Consulta_objectiu_6.2*/
proc sql;
UPDATE pfc_as.taula SET censura = 0
WHERE (data_sida_mort) Is Null;
quit;

/*Consulta_objectiu_7*/

/*Consulta_objectiu_7.1*/

proc sql;
UPDATE pfc_as.taula SET Temps_sida_mort_dies = (data_sida_mort-
Data_primer_tractament)
WHERE censura=1;
quit;

```

```

/*Consulta_objectiu_7.2*/

proc sql;
UPDATE pfc_as.taula SET Temps_sida_mort_dies = (ultim_contacte-
Data_primer_tractament)
WHERE censura=0;
quit;

/*          CREACIÓ D'ALTRES VARIABLES          */

/*Consulta_altres_1*/

proc sql;
create table pfc_as.anal_rang as
SELECT DISTINCT mcd_tractaments.id, taula.data_primer_tractament,
(datepart(taula.Data_primer_tractament)+182) AS mesos_TARGA,
datepart(mcd_analitiques.Data_analitica) AS Data_analitica,
mcd_analitiques.cd4, mcd_analitiques.cd4p, mcd_analitiques.CV,
mcd_analitiques.cvlog
FROM (pfc_as.mcd_tractaments INNER JOIN pfc_as.taula ON
mcd_tractaments.id=taula.id) INNER JOIN pfc_as.mcd_analitiques ON
taula.id=mcd_analitiques.id
WHERE (datepart(Data_analitica)-data_primer_tractament)>=91 AND
(datepart(Data_analitica)-data_primer_tractament)<=274 AND
Data_analitica Is Not Null AND data_primer_tractament Is Not Null;
quit;

data pfc_as.anal_rang;
set pfc_as.anal_rang;
format mesos_TARGA date.;
format Data_analitica date.;
run;

/*Afegir variable de la diferència*/
proc sql;
ALTER TABLE pfc_as.anal_rang
add dif int;
quit;

/*Consulta_altres_2*/

/*Consulta_altres_2.1*/

proc sql;
UPDATE pfc_as.anal_rang SET dif = (mesos_TARGA-Data_analitica)
WHERE (mesos_TARGA>=Data_analitica);
quit;

```

```

/*Consulta_altres_2.2*/

proc sql;
UPDATE pfc_as.anal_rang SET dif = (Data_analitica-mesos_TARGA)
WHERE (mesos_TARGA<Data_analitica);
quit;

/*Consulta_altres_3*/
proc sql;
create view pfc_as.min_dif as
SELECT Id, Min(dif) AS Mes_propera
FROM pfc_as.anal_rang
GROUP BY id;
quit;

proc sql;
create table pfc_as.data as
SELECT DISTINCT min_dif.Id, anal_rang.data_primer_tractament,
Mes_propera, Min(Data_analitica) AS Analitica_6_mesos
FROM pfc_as.anal_rang RIGHT JOIN pfc_as.min_dif ON
(anal_rang.dif=min_dif.mes_propera) AND (anal_rang.id=min_dif.id)
GROUP BY min_dif.id, mes_propera, anal_rang.Data_primer_tractament;
quit;

data pfc_as.data;
set pfc_as.data;
format Analitica_6_mesos date.;
run;

/*Consulta_altres_4*/

proc sql;
create view pfc_as.anal_6_mesos_cd4 as
SELECT anal_rang.id, data.data_primer_tractament, Data_analitica,
data.analitica_6_mesos, cd4, cd4p, mes_propera
FROM pfc_as.anal_rang RIGHT JOIN pfc_as.data ON (anal_rang.id=data.id
AND Data_analitica=analitica_6_mesos)
WHERE cd4 Is Not Null;
quit;

/*Consulta_altres_5*/

proc sql;
create view pfc_as.anal_6_mesos_cv as
SELECT anal_rang.id, data.data_primer_tractament, Data_analitica, CV,
cvlog, mes_propera
FROM pfc_as.anal_rang RIGHT JOIN pfc_as.data ON
(((anal_rang.id=data.id AND Data_analitica=analitica_6_mesos)))
WHERE (CV) Is Not Null;
quit;

```

```
/*Consulta_altres_6*/
```

```
proc sql;  
create view pfc_as.anal_6_mesos as  
SELECT DISTINCT data.Id, data.data_primer_tractament,  
data.analitica_6_mesos, cd4, cd4p, CV, cvlog, data.mes_propera  
FROM (pfc_as.data INNER JOIN pfc_as.anal_6_mesos_cd4 ON  
data.id=anal_6_mesos_cd4.Id) INNER JOIN pfc_as.anal_6_mesos_cv ON  
data.id=anal_6_mesos_cv.id;  
quit;
```

```
/*          CREACIÓ DE LA TAULA FINAL          */
```

```
/*Consulta_taula_final*/
```

```
proc sql;  
create table pfc_as.taula_final as  
SELECT taula.id, taula.data_naixement, taula.sexe,  
taula.grup_transmissio, taula.data_primera_visita, taula.CV,  
taula.cd4, taula.data_primer_tractament, taula.pauta_TARGA_inicial,  
taula.Hepatitis_C, data_mort, data_sida, data_sida_mort,  
censura, temps_sida_mort_dies, anal_6_mesos.analitica_6_mesos,  
anal_6_mesos.cd4 AS cd4_6mesos, anal_6_mesos.cd4p AS cd4p_6mesos,  
anal_6_mesos.CV AS CV_6mesos, anal_6_mesos.cvlog AS cvlog_6mesos  
FROM pfc_as.taula LEFT JOIN pfc_as.anal_6_mesos ON  
(taula.data_primer_tractament=anal_6_mesos.data_primer_tractament) AND  
(taula.id=anal_6_mesos.id)  
order by taula.id, taula.data_naixement, taula.sexe,  
taula.grup_transmissio, taula.data_primera_visita, taula.CV,  
taula.cd4, taula.data_primer_tractament, taula.pauta_TARGA_inicial,  
taula.Hepatitis_C, data_mort, data_sida, data_sida_mort,  
censura, temps_sida_mort_dies, anal_6_mesos.analitica_6_mesos,  
anal_6_mesos.cd4, anal_6_mesos.cd4p, anal_6_mesos.CV,  
anal_6_mesos.cvlog;  
quit;
```



## A 2.2: PROGRAMA DE L'ANÀLISI ESTADÍSTIC AMB SAS

---

```
/* _____ */
/* _____ */
/*          PROGRAMA D'ANÀLISI DE LES DADES AMB SAS          */
/* _____ */

/* Crear els formats de les variables          */
/* _____ */

proc format;
value sexe
    1="Home"
    2="Dona";
run;

proc format ;
value grups_edat
    1="edat<35"
    2="35-55"
    3="edat>55";
run;

proc format ;
value grups_trans_a
    1="UDVP"
    2="Home homosexual/bisexual"
    3="Heterosexual"
    4="Altres";
run;

proc format ;
value grups_trans_b
    1="UDVP"
    2="Home homosexual/bisexual"
    3="Heterosexual";
run;

proc format ;
value pauta_targa_inicial
    31="1 o 2 IP + 2 INT"
    32="2 INT + 1 INNT"
    33="3 INT"
    30="Altres";
run;

proc format ;
value era_TARGA
    1="Era early"
    2="Era late";
run;

proc format ;
value Hepatitis_C
    0="No"
    1="Si";
run;
```

```

proc format ;
value cd4_inf
    1="cd4<200"
    2="cd4>200";
run;

proc format ;
value cd4_c
    1="cd4<200"
    2="200<cd4<350"
    3="cd4>350";
run;

proc format ;
value cv_c
    1="cv<100000"
    2="cv>100000";
run;

proc format ;
value censura
    0="Censurat"
    1="Fa sida o mort";
run;

proc format ;
value fa_sida
    0="No sida"
    1="Si Sida";
run;

proc format ;
value fa_mort
    0="No mort"
    1="Si mort";
run;

```

```
LIBNAME pfc_as "E:\PFC-PISCIS\Analisi\En SAS\Llibreria";
```

```

/* Creació de variables                                     */
/* _____ */

```

```
/* Sexe*/
```

```

data pfc_as.taula_final;
set pfc_as.taula_final;
format sexe sexe.;
run;

```

```
/* Crear edat d'ingres en la cohort */
```

```

data pfc_as.taula_final;
set pfc_as.taula_final;
edat_ingres_cohort=(datdif(data_naixement, data_primera_visita,
30/360))/365;
run;

```

```

/* Grups d'edat */

data pfc_as.taula_final;
set pfc_as.taula_final;
    if(edat_ingres_cohort<35)      then grups_edat=1;
    if(edat_ingres_cohort>=35 and edat_ingres_cohort<55) then
grups_edat=2;
    if(edat_ingres_cohort>=55)    then grups_edat=3;
    if(edat_ingres_cohort=.) then grups_edat=.;
run;
data pfc_as.taula_final;
set pfc_as.taula_final;
format grups_edat grups_edat.;
run;

/* Grups de transmissió */

data pfc_as.taula_final;
set pfc_as.taula_final;
    if(grup_transmissio=1 OR grup_transmissio=2 OR
grup_transmissio=10 or grup_transmissio=11) then grups_trans_a=1;
    if(grup_transmissio=3 or grup_transmissio=5 or
grup_transmissio=19) then grups_trans_a=2;
    if(grup_transmissio=4) then grups_trans_a=3;
    if(grup_transmissio=6 or grup_transmissio=7 or
grup_transmissio=8 or grup_transmissio=9 or grup_transmissio=12 or
grup_transmissio=14 or grup_transmissio=15) then grups_trans_a=4;
run;

data pfc_as.taula_final;
set pfc_as.taula_final;
format grups_trans_a grups_trans_a.;
run;

data pfc_as.taula_final;
set pfc_as.taula_final;
    if(grups_trans_a=1) then grups_trans_b=1;
    if(grups_trans_a=2) then grups_trans_b=2;
    if(grups_trans_a=3) then grups_trans_b=3;
run;

data pfc_as.taula_final;
set pfc_as.taula_final;
format grup_trans_b grups_trans_b.;
run;

/* Pauta TARGA inicial */

data pfc_as.taula_final;
set pfc_as.taula_final;
format pauta_targa_inicial pauta_targa_inicial.;
run;

/* Crear Eres: extreure l'any de la data d'inici de TARGA */

data pfc_as.taula_final;
set pfc_as.taula_final;
    if (data_inici_TARGA<'01JAN2001'd) then era_TARGA=1;

```

```

        if (data_inici_TARGA>='01JAN2001'd) then era_TARGA=2;
        if (data_inici_TARGA=.) then era_TARGA=.;
run;

data pfc_as.taula_final;
set pfc_as.taula_final;
format era_targa era_targa.;
run;

/* Hepatitis C */

data pfc_as.taula_final;
set pfc_as.taula_final;
format Hepatitis_C hepC.;
run;

/* Crear CD4<200 */

data pfc_as.taula_final;
set pfc_as.taula_final;
    if(cd4<200) then cd4_inf=1;
    if(cd4>=200) then cd4_inf=2;
    if (cd4=.) then cd4_inf=.;
run;

data pfc_as.taula_final;
set pfc_as.taula_final;
format cd4_inf cd4_inf.;
run;

/* CD4 categoritzada*/

data pfc_as.taula_final;
set pfc_as.taula_final;
    if(cd4<200) then cd4_c=1;
    if(cd4>=200 AND cd4<=350) then cd4_c=2;
    if(cd4>350) then cd4_c=3;
    if (cd4=.) then cd4_c=.;
run;

data pfc_as.taula_final;
set pfc_as.taula_final;
format cd4_c cd4_c.;
run;

/* CV */

data pfc_as.taula_final;
set pfc_as.taula_final;
    if(CV<100000) then cv_c=1;
    if(CV>=100000) then cv_c=2;
    if (CV=.) then cv_c=.;
run;

data pfc_as.taula_final;
set pfc_as.taula_final;
format cv_c cv_c.;

```

```

run;

/* Censura */

data pfc_as.taula_final;
set pfc_as.taula_final;
format censura censura.;
run;

/* Crear fa sida */

data pfc_as.taula_final;
set pfc_as.taula_final;
    if(data_sida=.) then fa_sida=0;
    else fa_sida=1;
run;

data pfc_as.taula_final;
set pfc_as.taula_final;
format fa_sida fa_sida.;
run;

/* Crear fa mort */

data pfc_as.taula_final;
set pfc_as.taula_final;
    if(data_mort=.) then fa_mort=0;
    else fa_mort=1;
run;

data pfc_as.taula_final;
set pfc_as.taula_final;
format fa_mort fa_mort.;
run;

data pfc_as.taula_final;
set pfc_as.taula_final;
temps_sida_mort_anys=temps_sida_mort_dies/365;
run;

data pfc_as.taula_final;
set pfc_as.taula_final;
temps_sida_mort_mesos=temps_sida_mort_dies/30;
run;

```

```

/* Descriptiva univariant                                     */
/* _____ */

ODS listing close;
ods pdf file='E:\Sortida SAS\Descriptiva Univariant.pdf';

proc freq data=pfc_as.taula_final;
tables grups_edat sexe grups_trans_a;
title 'DESCRIPTIVA UNIVARIANT: variables socio-demogràfiques';
run;

proc univariate;
var CV cd4;
TITLE 'DESCRIPTIVA UNIVARIANT: variables clínico-epidemiològiques a
l'inici del tractament';
run;

proc freq data=pfc_as.taula_final;
tables cv_c cd4_c pauta_targa_inicial era_targa Hepatitis_C;
run;

proc univariate;
var CV_6mesos cd4_6mesos;
TITLE 'DESCRIPTIVA UNIVARIANT: variables clínico-epidemiològiques als
6 mesos';
run;

proc freq data=pfc_as.taula_final;
tables fa_sida fa_mort censura;
title 'DESCRIPTIVA UNIVARIANT: variables de seguiment';
run;

proc means ql median q3 ;
var temps_sida_mort_mesos;
run;

ods pdf close;
ods listing;

/* Descriptiva bivariant                                     */
/* _____ */

ODS listing close;
ods pdf file='E:\Sortida SAS\Descriptiva bivariant.pdf';

proc freq data=pfc_as.taula_final;
tables Hepatitis_C*grups_trans_a fa_sida*grups_trans_a
fa_mort*grups_trans_a censura*grups_trans_a cd4_inf*grups_trans_a;
run;

ods pdf close;
ods listing;

```

```

/* Anàlisi de supervivència */
/* _____ */

ODS listing close;
ods pdf file='E:\Sortida SAS\Kaplan-Meier.pdf';

proc lifetest data=pfc_as.taula_final plots=(S);
time temps_sida_mort_anys*censura(0);
strata sexe;
Title 'Estimació funció de supervivència (Kaplan-Meier): SEXE';
run;

proc lifetest data=pfc_as.taula_final plots=(S);
time temps_sida_mort_anys*censura(0);
strata grups_trans_b;
Title 'Estimació funció de supervivència (Kaplan-Meier): GRUP DE
TRANSMISSIÓ';
run;

proc lifetest data=pfc_as.taula_final plots=(S);
time temps_sida_mort_anys*censura(0);
strata cd4_c;
Title 'Estimació funció de supervivència (Kaplan-Meier): NIVELL DE
LINFÒCITS CD4';
run;
ods pdf close;
ods listing;
ODS listing close;
ods pdf file='E:\Sortida SAS\Cox univariant.pdf';

proc phreg data=pfc_as.taula_final;
model temps_sida_mort_anys*censura(0)=sexe;
Title 'Model de Cox univariat: SEXE';
run;

proc phreg data=pfc_as.taula_final;
model temps_sida_mort_anys*censura(0)=grups_edat;
Title 'Model de Cox univariat: GRUPS SEGONS EDAT';
run;

proc phreg data=pfc_as.taula_final;
model temps_sida_mort_anys*censura(0)=grups_trans_a;
Title 'Model de Cox univariat: GRUP DE TRANSMISSIÓ';
run;

proc phreg data=pfc_as.taula_final;
model temps_sida_mort_anys*censura(0)=cv_c;
Title 'Model de Cox univariat: CARREGA VIRAL CATEGORITZADA';
run;

proc phreg data=pfc_as.taula_final;
model temps_sida_mort_anys*censura(0)=CD4_C;
Title 'Model de Cox univariat: CD4 CATEGORITZAT';
run;

proc phreg data=pfc_as.taula_final;
model temps_sida_mort_anys*censura(0)=pauta_targa_inicial;
Title 'Model de Cox univariat: PAUTA TARGA INICIAL';
run;

proc phreg data=pfc_as.taula_final;

```

```
model temps_sida_mort_anys*censura(0)=Era_TARGA;  
Title 'Model de Cox univariat: ERA TARGA';  
run;  
  
proc phreg data=pfc_as.taula_final;  
model temps_sida_mort_anys*censura(0)=Hepatitis_C;  
Title 'Model de Cox univariat: HEPATITIS C';  
run;  
  
ods pdf close;  
ods listing;
```



### A 3: PLANIFICACIÓ I COSTOS DEL PROJECTE

A continuació es detalla la distribució de les hores dedicades en aquest projecte juntament amb els seus costos associats.

Distribució de les hores totals de dedicació i la seva distribució en el temps de treball:

Fases		Hores
INICIACIÓ	Cerca informació malaltia	3
	Cerca informació AS	3
	Títol	1
	Objectius	1,5
	Justificació	1
	Descripció de l'estudi	2
	Estudi AS	25
	Planificació	3
	Implicacions ètiques	2
EXECUCIÓ	Anàlisi exploratoria general	1
	Preparació dades: SQL	30
	Traduir SQL amb SAS	15
	AS Descriptiva	8
	AS Estimació KM	6
	AS Model de Cox	10
	Traduir AS amb SAS	5
	Anàlisi dels resultats	4
	Possibles interpretacions	1
SEGUIMENT	Redactar teoria AS	6
	Redactar descripció de l'estudi i cohort	1,5
	Redactar preparació dades	8
	Redactar anàlisi dades	6
	Redactar resultats	15
	Redactar conclusions	1
	Bibliografia	1
	Repàs, retocs i impressió	10
	Preparació Power Point	30
<b>Total</b>		<b>200</b>

Fases		MARÇ																																
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31		
INICIACIÓ	Cerca informació malaltia																																	
	Cerca informació AS																																	
	Títol																																	
	Objectius																																	
	Justificació																																	
	Descripció de l'estudi																																	
	Estudi AS																																	
	Planificació																																	
	Implicacions ètiques																																	
	EXECUCIÓ	Anàlisi exploratoria general																																
Preparació dades: SQL																																		
Traduir SQL amb SAS																																		
AS Descriptiva																																		
AS Estimació KM																																		
AS Model de Cox																																		
Traduir AS amb SAS																																		
Anàlisi dels resultats																																		
Possibles interpretacions																																		
SEGUIMENT	Redactar teoria AS																																	
	Redactar descripció de l'estudi i cohort																																	
	Redactar preparació dades																																	
	Redactar anàlisi dades																																	
	Redactar resultats																																	
	Redactar conclusions																																	
	Bibliografia																																	
	Repàs, retocs i impressió																																	
Preparació Power Point																																		
<b>Lliurament</b>																																		
<b>Defensa</b>																																		

Fases		ABRIL																														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
INICIACIÓ	Cerca informació malaltia																															
	Cerca informació AS																															
	Títol																															
	Objectius																															
	Justificació																															
	Descripció de l'estudi																															
	Estudi AS																															
	Planificació																															
	Implicacions ètiques																															
EXECUCIÓ	Anàlisi exploratoria general																															
	Preparació dades: SQL																															
	Anàlisi exploratoria: AS																															
	Aplicació AS																															
	Anàlisi dels resultats																															
	Possibles interpretacions																															
SEGUIMENT	Redactar teoria AS																															
	Redactar descripció de l'estudi i cohort																															
	Redactar preparació dades																															
	Redactar anàlisi dades																															
	Redactar resultats																															
	Redactar conclusions																															
	Bibliografia																															
	Repàs, retocs i impressió																															
	Preparació Power Point																															
<b>Lliurament</b>																																
<b>Defensa</b>																																

Fases		MAIG																															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	
INICIACIÓ	Cerca informació malaltia																																
	Cerca informació AS																																
	Títol																																
	Objectius																																
	Justificació																																
	Descripció de l'estudi																																
	Estudi AS																																
	Planificació																																
	Implicacions ètiques																																
EXECUCIÓ	Anàlisi exploratoria general																																
	Preparació dades: SQL																																
	Anàlisi exploratoria: AS																																
	Aplicació AS																																
	Anàlisi dels resultats																																
	Possibles interpretacions																																
SEGUIMENT	Redactar teoria AS																																
	Redactar descripció de l'estudi i cohort																																
	Redactar preparació dades																																
	Redactar anàlisi dades																																
	Redactar resultats																																
	Redactar conclusions																																
	Bibliografia																																
	Repàs, retocs i impressió																																
	Preparació Power Point																																
<b>Lliurament</b>																																	
<b>Defensa</b>																																	

Fases		JUNY																														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
INICIACIÓ	Cerca informació malaltia																															
	Cerca informació AS																															
	Títol																															
	Objectius																															
	Justificació																															
	Descripció de l'estudi																															
	Estudi AS																															
	Planificació																															
	Implicacions ètiques																															
EXECUCIÓ	Anàlisi exploratoria general																															
	Preparació dades: SQL																															
	Anàlisi exploratoria: AS																															
	Aplicació AS																															
	Anàlisi dels resultats																															
	Possibles interpretacions																															
SEGUIMENT	Redactar teoria AS																															
	Redactar descripció de l'estudi i cohort																															
	Redactar preparació dades																															
	Redactar anàlisi dades																															
	Redactar resultats																															
	Redactar conclusions																															
	Bibliografia																															
	Repàs, retocs i impressió																															
	Preparació Power Point																															
<b>Lliurament</b>																																
<b>Defensa</b>																																

Fases		JULIOL																														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
INICIACIÓ	Cerca informació malaltia																															
	Cerca informació AS																															
	Títol																															
	Objectius																															
	Justificació																															
	Descripció de l'estudi																															
	Estudi AS																															
	Planificació																															
Implicacions ètiques																																
EXECUCIÓ	Anàlisi exploratoria general																															
	Preparació dades: SQL																															
	Anàlisi exploratoria: AS																															
	Aplicació AS																															
	Anàlisi dels resultats																															
	Possibles interpretacions																															
SEGUIMENT	Redactar teoria AS																															
	Redactar descripció de l'estudi i cohort																															
	Redactar preparació dades																															
	Redactar anàlisi dades																															
	Redactar resultats																															
	Redactar conclusions																															
	Bibliografia																															
	Repàs, retocs i impressió																															
	Preparació Power Point																															
Lliurament																																
Defensa																																

Els costos atribuïbles a aquest projecte han estat:

<b>Descripció</b>	<b>Cost</b>
Hores dedicades (30 € per hora)	6.000,00 €
Ordinador (PC)	800,00 €
Versió de Windows	400,00 €
SAS Académico	6.300,00 €
SPSS for Windows	750,00 €
Microsoft Office	200,00 €
Impressions varies	7,00 €
Impressions copies projecte final	200,00 €
Enquadernació PFC	100,00 €
<b>Total</b>	<b>14.757 €</b>

## AGRAÏMENTS

---

En primer lloc, moltes gràcies a l'esforç dels meus tutors, Anna i Klaus, per anar a un ritme de treball molt productiu, per tots els bons consells que m'heu donat i tot el que m'heu ensenyat.

En segon lloc agrair a Jordi Casabona el seu consentiment d'ús de les dades del Centre d'Estudis Epidemiològics de les ITS i Sida de Catalunya (CEEISCAT) i el seu suport.

També agrair a l'equip del CEEISCAT pel seu suport i en especial a Berta Ortiga pel seu suport a nivell mèdic, a Eva Puchol i Eva Loureiro pel seu suport a nivell informàtic, a Alexandra Montoliu pel seu suport a nivell estadístic; a Anabel Romero per haver resolt dubtes que han anat sorgint i a Jéssica Sánchez pel seu suport com a documentalista.

No oblidar-me d'agrair la seva col·laboració a Mercè Codina i a Cisco Soler de la biblioteca de la FME pel seu suport bibliotecari i d'elaboració de documents. També agrair a tot l'equip de la biblioteca el seu interès i suport en aquest projecte.

D'altra banda, agrair molt especialment tot el suport moral que m'ha donat la meva família, la meva mare, el meu germà i l'Antonio; la meva parella, Miguel Àngel Solé; i les meves companyes de facultat, Elisabet Peralta i Ana Gómez.

Per últim, dir que personalment dedico aquest projecte i tot l'esforç amb el que l'he tirat endavant als meus pares i al meu germà.

Sincerament, gràcies a tots.