

# Grau en Estadística

---

**Títol: Anàlisi dels factors associats a les interrupcions del tractament amb metadona de pacients usuaris de drogues**

**Autor: Raquel Pérez Colet**

**Director: Klaus Langohr**

**Departament: Estadística i Investigació Operativa (UPC)**

**Convocatòria: 28/06/2015**



## AGRAIMENTS

Primer de tot vull agrair l'esforç realitzat per la meva família, per a que jo hagi pogut fer la carrera d'estadística ja que sense ells no hagués estat possible. En especial a la meva mare, que sempre que ha pogut m'ha dedicat temps per a què tot quedés millor i fos més fàcil.

A la meva parella, per donar-me forces, per creure en mi sempre d'una forma incondicional, i per tot el recolzament que m'ha donat sempre que ha sigut necessari.

També vull agrair tot l'esforç d'aquells professors que han fet què ens apassionés l'estadística amb classes magistrals.

No hem puc deixar l'Aratnza Sanvisens de la Unitat d'Addiccions de l'Hospital Germans Trias i Pujol. que m'ha facilitat les dades, donant-me l'oportunitat de treballar amb valors reals tot confiant en mi per a treure'n profit d'elles.

I per últim, agrair al meu director del projecte de final de carrera, en Klaus Langohr, l'oportunitat que m'ha donat amb aquest tema de treball i el recolzament que he tingut per part d'ell durant tot el procés.

## **RESUM**

L'essència d'aquest treball és, emprant tècniques de modelització i d'anàlisi de supervivència, descobrir quins factors influeixen en l'adherència al tractament de la metadona i el temps d'adhesió; tot estudiant, observant i analitzant una sèrie de 880 casos amb dependència d'opiacis (heroïna) admesos en la Unitat Hospitalària de Dispensació de Metadona (UHDM), de l'Hospital Universitari de Bellvitge i l'Hospital Germans Trias i Pujol, des de l'inici del programa de 1995 fins a finals de 2013.

## **PARAULES CLAU**

- Opiacis.
- Metadona.
- Via parenteral.
- Regressió logística.
- Anàlisi de supervivència.

## **CLASSIFICACIÓ AMS**

62P10, 62A99 i 62N99.

## ABSTRACT

### ANALYSIS OF FACTORS ASSOCIATED WITH THE INTERRUPTION OF METHADONE TREATMENT OF DRUG USERS.

The essence of this work is the study on the continuity of methadone treatment in patients treated for heroin addiction. Methadone is, usually, administered orally once a day and it is recommended for an indefinite period. Methadone favors the retention of patients in the health system, reduces the transmission of infections such as HIV and parenteral hepatitis C, and improves survival.

In this work, we study the factors associated with the adherence and the continuity in treatment, which are essential to prevent the relapse of heroin addiction. According to estimations, in Spain there are around 80000 people in methadone treatment, 8000 of them in Catalonia.

The aim is to discover what factors are related to treatment adherence and to the time until treatment discontinuation in a study of 880 patients with opioid dependence (heroin) admitted to the hospital unit Methadone Dispensing (UHDM) Hospital Bellvitge University since the program's inception in 1995 until the end of 2013. The first part of this work is about the preprocessing of the data base. Following, the descriptive analysis of the data is presented. Finally, Chapter 5 and 6 are about the logistic regression model used to determine the factors associated with methadone treatment interruptions and the survival analysis methodology used to study time to treatment discontinuation.

## KEYWORDS

- *Opiates.*
- *Methadone.*
- *Parenterally.*
- *Logistic regression.*
- *Survival Analysis*

## Índex

INTRODUCCIÓ.....	10
METODOLOGIA.....	10
COS DEL TREBALL .....	11
1.    Marc teòric .....	11
1.1.    Opiacis .....	11
1.2.    Drogues .....	11
1.3.    Metadona .....	12
1.4.    Tractament amb metadona a Espanya.....	16
2.    Presentació de les dades .....	19
3.    Preprocessament de les dades.....	21
3.1.    Tractament d'observacions inusuals.....	21
3.2.    Recodificació de variables.....	22
3.3.    Noves variables.....	22
4.    Anàlisi descriptiva.....	23
4.1.    Variables categòriques.....	23
4.2.    Variables contínues .....	28
4.3.    Anàlisi bivariant .....	30
4.4.    Anàlisi descriptiva base Dosis.....	32
5.    Modelització.....	34
5.1.    Cerca del model.....	38
5.2.    Model definitiu .....	39
5.3.    Interpretació del model.....	40
5.4.    Diagnosi i validació del model .....	41
6.    Anàlisi de supervivència.....	44
6.1.    Anàlisi no paramètrica.....	45
6.2.    Anàlisi semi-paramètric.....	49
CONCLUSIONS .....	59
LIMITACIONS I FUTURES LINIES D'INVESTIGACIÓ .....	60
BIBLIOGRAFIA .....	61
ANNEX .....	62
i.    Models no presentats al informe .....	62
Models de regressió logística .....	62
Models de Cox.....	70

ii.	Llibreries utilitzades.....	71
iii.	Scripts d'R.....	72
	▪ Preprocessament de les dades .....	72
	▪ Descriptiva.....	73
	▪ Modelització.....	77
	▪ Supervivència .....	80

# Índex de taules

## COS DEL TREBALL

### Marc teòric

TAULA 1. 1 <i>CONVERSIÓ HEROÏNA/METADONA</i> <sup>3</sup> .....	14
---	----

### Presentació de les dades

TAULA 2.1 <i>VARIABLES CATEGÒRIQUES DE LA BASE DE DADES EPISODIS</i> . .....	19
TAULA 2.2 <i>VARIABLES CONTINUES DE LA BASE DE DADES EPISODIS</i> . .....	20
TAULA 2.3 <i>VARIABLES TIPUS DATA DE LA BASE DE DADES EPISODIS</i> . .....	20

### Preprocessament de les dades

#### Anàlisi descriptiva

TAULA 4.1 <i>TAULA DE FREQÜÈNCIES DE VARIABLES CATEGÒRIQUES BÀSIQUES</i> . .....	23
TAULA 4.2 <i>TAULA DE FREQÜÈNCIES DE VARIABLES CATEGÒRIQUES RELACIONADES AMB LES DROGUES</i> . .....	24
TAULA 4.3 <i>TAULA DE FREQÜÈNCIES DE VARIABLES CATEGÒRIQUES RELACIONADES AMB DIFERENTS MALALTIES</i> . .....	25
TAULA 4.4 <i>TAULA RESUM DE LES VARIABLES CONTINUES</i> . .....	28
TAULA 4.5 <i>TAULA FREQÜÈNCIES DE VARIABLES DE L'INICI CONSUM D'HEROÏNA</i> . .....	28
TAULA 4.6 <i>CATEGORIES VARIABLE ÚS DEL PRESERVATIU</i> . .....	31

#### Modelització

TAULA 5.1 <i>MODEL RESULTANT</i> . .....	39
TAULA 5.2 <i>RESULTAT DEL TEST D'OUTLIERS</i> . .....	41
TAULA 5.3 <i>RESULTAT DEL TEST DE MULTICOL·LINEALITAT</i> . .....	41
TAULA 5.4 <i>RESULTAT DEL TEST DE BONDAT DE L'AJUST</i> . .....	41

#### Anàlisi de supervivència.

TAULA 6.1 <i>MEDIANA DE SUPERVIVÈNCIA PER GRUP DE SEXE I VIH</i> . .....	45
TAULA 6.2 <i>ESTIMACIÓ DE LA FUNCIÓ DE SUPERVIVÈNCIA PER GRUP D'ANTECEDENTS PENALS</i> . .....	47
TAULA 6.3 <i>ESTIMACIÓ DE LA FUNCIÓ DE SUPERVIVÈNCIA PER GRUP SEGONS SI HAN TINGUT O NO RELACIONS A CANVI DE DINERS O DROGUES</i> . .....	48
TAULA 6.4 <i>PROVES DE DIFERÈNCIA EN LA SUPERVIVÈNCIA</i> . .....	49
TAULA 6.5. <i>RESULTATS MODEL DE COX</i> . .....	50
TAULA 6.6 <i>TAULA DE HIPOTESIS DE PROPORCIONALITAT DE RISCOS</i> . .....	51
TAULA 6.7 <i>RESULTATS MODEL DEL COX ESTRATIFICAT DEFINITIU</i> . .....	55
TAULA 6.8 <i>TAULA DE HIPOTESI DE PROPORCIONALITAT DE RISCOS</i> . .....	57



# Índex de figures

## COS DEL TREBALL

### Marc teòric

FIGURA 1. 1 CRONOLOGIA DEL TRACTAMENT AMB METADONA.....	18
---	----

### Presentació de les dades

### Preprocessament de les dades

### Anàlisi descriptiva

GRÀFIC 4.1 GRÀFIC DE SECTORS DE LA VARIABLE NÚMERO DE PARELLES EN ELS DARRERS 6 MESOS.....	26
GRÀFIC 4.2 GRÀFIC DE SECTORS DE LA VARIABLE SITUACIÓ LABORAL.....	26
GRÀFIC 4. 3 GRÀFIC DE SECTORS DE LA VARIABLE CONSUM DE COCAÏNA.....	27
GRÀFIC 4.4 HISTOGRAMA DE LA FREQÜÈNCIA D'INDIVIDUS QUE INICIEN EL TRACTAMENT PER ANY.....	29
GRÀFIC 4.5 GRÀFIC DE MOSAIC DE LES VARIABLE I ÚS DEL PRESERVATIU.....	30
GRÀFIC 4.6 GRÀFIC DE MOSAICS DE LES VARIABLE VIH I PARTICIPACIÓ EN EL PROGRAMA DE XERINGUES.....	30
GRÀFIC 4.7 GRÀFIC DE MOSAICS DE LES VARIABLES VIH I ÚS DEL PRESERVATIU.....	31
GRÀFIC 4.8 GRÀFIC DE MOSAICS DE LES VARIABLES SEXE I ÚS DEL PRESERVATIU.....	32
GRÀFIC 4.9 GRÀFIC DE TENDÈNCIA DE LA VARIABLE DOSIS PRESCRITA.....	33

### Modelització

GRÀFIC 5.1 GRÀFIC DE DENSITAT DE LA VARIABLE DOSIS INICIAL.....	35
GRÀFIC 5.2 GRÀFIC DE DENSITAT DE LA VARIABLE EDAT INICI DE TRACTAMENT.....	36
GRÀFIC 5.3 GRÀFIC DE DENSITAT DE LA VARIABLE NOMBRE DE CD4.....	36
GRÀFIC 5.4 GRÀFIC DE RELACIONS ENTRE LES VARIABLES CONTINUES.....	37
GRÀFIC 5.5 BOXPLOT DE LES VARIABLES NOMBRE DE CD4 I PACIENT ACTIU.....	38
GRÀFIC 5.6 BOXPLOT VARIABLE NOMBRE DE CD4 I INFECCIÓ D'VIH.....	39
GRÀFIC 5.7 BONDAT DE L'AJUST.....	42
GRÀFIC 5.8 GRÀFICA DE LA CORBA ROC.....	43

### Anàlisi de supervivència.

GRÀFIC 6.1 FUNCIÓ DE SUPERVIVÈNCIA DEL MODEL NUL.....	45
GRÀFIC 6.2 ESTIMACIÓ DE LA FUNCIÓ DE SUPERVIVÈNCIA PER LES VARIABLES SEXE I VIH.....	46
GRÀFIC 6.3 ESTIMACIÓ DE LA FUNCIÓ DE SUPERVIVÈNCIA PER ELS DOS GRUPS D'ANTECEDENTS PENALS.....	47
GRÀFIC 6. 4 ESTIMACIÓ DE LA FUNCIÓ DE SUPERVIVÈNCIA PER ALS DOS GRUPS DE LA VARIABLE RELACIÓ PER DROGUES O DINERS.....	48
GRÀFIC 6.5 GRÀFIC DE RESIDUS DE SCHOENFELD PEL MODEL DE COX.....	52
GRÀFIC 6.6 GRÀFICA LOG-LOG PER LA VARIABLE SEXE.....	53
GRÀFIC 6.7 GRÀFICA LOG-LOG PER LA VARIABLE CONSUM DE COCAÏNA.....	53
GRÀFIC 6.8 GRÀFICA LOG-LOG PER LA VARIABLE SITACIÓ LABORAL.....	54
GRÀFIC 6.9 GRÀFICA LOG-LOG PER LA VARIABLE INFECCIÓ D'VIH.....	54
GRÀFIC 6.10 GRÀFIC DE RESIDUS DE SCHOENFELD PEL MODEL DE COX ESTRATIFICAT.....	56

## INTRODUCCIÓ

Aquest treball es basa en l'estudi de la continuïtat en els programes de manteniment amb metadona, que és el fàrmac d'elecció per al tractament de l'addicció a l'heroïna. La metadona afavoreix la retenció dels pacients en el sistema de salut millorant la seva qualitat de vida, disminuint la transmissió d'infeccions parenterals com el virus de la immunodeficiència humana (VIH) i la hepatitis C i beneficiant la supervivència.

Al treball estudiarem els factors associats a l'adherència i la continuïtat en el tractament, que són claus per prevenir recaigudes en el consum d'heroïna. S'estima que a Espanya hi ha al voltant de 80.000 persones en tractament de metadona, 8.000 d'elles a Catalunya.

L'objectiu és descobrir quins factors influeixen en l'adherència al tractament i en el temps fins l'adhesió en una sèrie de 880 casos amb dependència d'opiacis (heroïna), admesos en la Unitat Hospitalària de Dispensació de Metadona (UHDM) del Hospital Universitari de Bellvitge, des de l'inici del programa de 1995 fins a finals de 2013; les dades han estat facilitades per l'equip de recerca de la Unitat d'Addiccions de l'Hospital Germans Trias i Pujol. Aquests 880 casos corresponen a 680 individus, ja que hi ha individus que abandonen el tractament i més endavant el tornen a reiniciar, però per no haver de filtrar les dades i perdre informació els tractarem com episodis independents.

Primer realitzarem una part de processament de dades i una anàlisi descriptiva. Després, i una vegada entrant en matèria, farem un estudi de regressió logística per determinar els factors associats a les interrupcions del tractament amb metadona, i una anàlisi de supervivència per explorar el temps fins la interrupció del procés i els components que influeixen en el tall del seguiment. Tots aquests apartats venen precedits d'un marc teòric on s'introdueix el tema de les drogues i els tractaments de manteniment amb metadona.

## METODOLOGIA

Es tracta d'un estudi longitudinal en individus en tractament de metadona, amb dades facilitades en format SPSS i tractades amb R. Es fa una anàlisi inicial de la base de dades que compren la part del *preprocessing* i una primera anàlisi descriptiva, on deixarem la base de dades preparada per entrar en matèria i podrem estudiar les relacions entre variables i les tendències d'aquestes. Després realitzarem un estudi de regressió logística per determinar els factors associats a les interrupcions del tractament amb metadona, i examinarem l'associació de certes variables explicatives i el temps fins que abandonen el tractament. Part de les funcions d'R utilitzades seran exposades en el transcurs del treball.

## COS DEL TREBALL

### 1. Marc teòric

Molta de la informació exposada en aquest apartat ha estat extreta de l'article "Deshabituaçión de opiáceos(I) de Toxicología", de M<sup>a</sup> Carmen Vidal Casero (Vidal, 2002).

#### 1.1. Opiacis

Els opiacis són un conjunt de substàncies alcaloides que es troben a l'opi i que redueixen l'activitat del sistema nerviós central. L'opi s'extreu de l'exsudació lletosa que s'obté en fer un tall a la càpsula d'un tipus de rosella anomenat cascall (*Papaver somniferum* L). També s'utilitza tradicionalment el terme per englobar els derivats de la morfina i, per extensió, els diferents compostos, encara que no provenguin de la morfina ni de l'opi, que actuen de manera semblant a ella, fixant-se en els mateixos receptors.

Alguns opiacis naturalment presents a l'opi són la morfina, la tebaïna, l'hidromorfina (o HM, fàrmac per al dolor crònic intens) l'heroïna (droga il·legal), la metadona (droga legal com a teràpia a la drogodependència) i la codeïna, que s'usa actualment com a analgèsic. En general, s'usen o s'havien usat com a analgèsics o com a drogues d'abús<sup>9</sup>.

#### 1.2. Drogues

El consum d'heroïna, especialment per via parenteral, és la principal causa d'una gran part dels problemes socials i de salut pública Europeus des dels anys 70.

Quan parlem de la subministració de drogues per via parenteral ens referim al consum a través d'una injecció amb tot el que això suposa. D'entre tots els consumidors de drogues els que ho fan per via parenteral són els que tenen major risc de patir malalties o sobredosi. Unes de les malalties que es poden contraure per aquest fet són l'VIH o la hepatitis.

A nivell espanyol el consum de drogues il·legals per grups amplis de població és bastant recent. A més a més, fins que es va començar a consumir l'heroïna per vena, a finals dels anys setanta, les seves repercussions sociosanitàries eren aparentment irrellevants, i tampoc hi havien gairebé referències de problemes importants associats a la utilització d'al·lucinògens o cànnabis durant la transició política; ni de la utilització d'especialitats farmacèutiques amb amfetamina per preparar exàmens, evitar la fatiga o per aprimar<sup>1</sup>.

Al llarg dels anys vuitanta i a començaments dels noranta "l'epidèmia" del consum d'heroïna injectada va ocasionar a Espanya un gran augment de la mortalitat juvenil, la demanda de l'atenció sanitària, i la delinqüència contra la propietat generant intenses alarmes socials, i sent la responsable de que "les drogues" arribessin a percebre's com un

dels tres principals problemes de la població, junt amb l'atur i el terrorisme. A més tots aquests esdeveniments van venir acompanyats d'una ràpida expansió de la infecció del VIH<sup>4</sup>.

Es considera que el major impacte de la mortalitat per sobredosi es va produir als anys 1991-1992 amb més de 1.700 morts l'any, i en més del 90% dels casos estava implicada la injecció d'heroïna. Però hem de tenir en compte que les morts per VIH estan relacionades amb la injecció de drogues, però aquesta malaltia s'hauria agafat de 6 a 11 anys abans. Si ens centrem en dates veiem que la màxima incidència de VIH va ser del 1985 al 1987.

En resum, la dependència de drogues, en concret l'addicció als opiacis té un impacte elevat en la població espanyola i ha sigut reconeguda com un trastorn seriós, inadequat i crònic.

### *Producció, tràfic i confiscació*

Per a fer-nos una idea de la magnitud del problema ressaltem que, al 2009, a Europa hi van haver aproximadament 59.000 confiscacions que es van saldar amb la intercepció de 24 tones d'heroïna, dos terços dels quals van ser notificats per Turquia. Aquest mateix any pel que fa al número de partides de contraban practicades, el Regne Unit i Espanya ocupaven els primers llocs. Hem de tenir en compte que el 2009 va ser un màxim de producció mundial, ja que fins al moment la producció presentava una tendència creixent, i al 2010 va fer una petita davallada, passant de produir 6.900 tones a 3.600 a Afganistan, que segueix sent el principal productor d'opi mundial<sup>6</sup>.

### **1.3. Metadona**

La metadona s'utilitza com a fàrmac substitutiu per aconseguir la reducció/eliminació del síndrome d'abstinència, la desintoxicació reduint/eliminant el desig de consumir i el manteniment bloquejant els efectes euforitzants d'altres opiacis<sup>3</sup>.

S'acostuma a administrar per via oral una vegada al dia i es recomana durant un temps indefinit. La metadona afavoreix la retenció dels pacients en el sistema de salut millorant la seva qualitat de vida, disminuint la transmissió d'infeccions parenterals com el VIH i la hepatitis C i beneficiant la supervivència.

S'entén per manteniment amb metadona l'administració continuada a un individu dependent d'heroïna, d'un fàrmac substitutiu de la mateixa per via oral, en una dosi relativament estable, sota control mèdic, per un període definit superior als 21 dies i sense clar objectiu de desintoxicació.

La metadona arriba al seu màxim d'efectivitat a les 2 hores després del seu subministrament i té una vida mitjana plasmàtica de 25 hores, podent oscil·lar de les 15 a les 47 hores. Quan

es deixa el tractament, si volem fer estudis o un seguiment dels individus, hem de tenir en compte que la metadona roman a la sang durant un temps, podent tenir conclusions errònies per culpa de l'efecte retardat d'aquesta.

A l'hora d'incloure un nou pacient en un tractament amb metadona és molt important que sigui conscient de què és la metadona, quines característiques farmacològiques té i en què consisteix el tractament de manteniment. La informació que se li facilita al futur pacient ha d'incloure les següents indicacions<sup>3</sup>:

- La metadona triga 2 o 3 hores fins que arriba al seu efecte màxim.
- Tot i mantenint una dosi diària fixa, els nivells en sang no s'estabilitzen fins que passen 3 o 5 dies, i per això els increments de la dosi s'han de fer en aquest termini de temps.
- Si es consumeixen drogues (paral·lelament) incrementa el risc d'intoxicació greu.
- Alguns medicaments interfereixen amb la metabolització de la metadona provocant canvis en les concentracions plasmàtiques i per tant facilitant l'aparició d'intoxicació o abstinència.

Els efectes beneficiosos que observem amb el tractament amb metadona són<sup>3</sup>:

- Disminució del consum d'opioides il·legals.
- Disminució de conductes delictives.
- Disminució de les conductes de risc per la infecció per VIH.
- Disminució del risc de mort per sobredosi.
- Millora de la qualitat de vida.
- Retenció en tractament.

Els efectes descrits s'aconsegueixen mentre els pacients romanen al tractament, sent aquest un dels millors paràmetres per valorar l'eficàcia i l'objecte d'estudi d'aquest document.

És important acompanyar el tractament de metadona amb recolzament psicosocial, social i intervencions mèdiques, a més a més de la pròpia prescripció del substitutiu, per augmentar l'efectivitat del tractament.

Quan un pacient es manté estable en el tractament, abstinent d'heroïna de llarga durada i ha canviat de forma clara el seu estil de vida, es pot plantejar la suspensió del tractament<sup>3</sup>, tot i que a vegades el tractament es perllonga de forma indefinida.

### *Formes d'administració de la metadona*

Referent a la dosi es poden utilitzar altes per exercir l'anomenat efecte bloqueig, mentre que en altres casos només s'utilitzen petites dosis per aconseguir una millora simptomàtica durant períodes curts de temps.

No existeix, actualment, la possibilitat de determinar el grau de dependència (tolerància, neuroadaptació, etc.) mitjançant detecció d'opiacis en sang o en orina. Això dificulta, notablement, prescriure una dosi inicial de metadona que compensi adequadament el procés de neurobiologia d'irritabilitat neuromuscular generant el síndrome d'abstinència.

La dosificació òptima hauria d'eliminar el "craving" (ànsia d'opiacis), però segueix qüestionant-se la idoneïtat de la utilització de dosis baixes (50mg o inferiors), o altes (80-120 mg) de metadona en tractaments de manteniment.

Si s'administra la quantitat suficient de metadona el síndrome d'abstinència es veu revertit en un espai curt de temps. En un programa de substitució el que es pretén és administrar una dosi equivalent de l'agonista respecte l'opiaci il·legal, però això no és possible sempre. No tenim una altra manera de saber la quantitat consumida, la periodicitat i altres característiques del consum que preguntant al propi consumidor, ja que no existeix actualment una mesura en sang que quantifiqui aquests detalls. Com ja ens podem imaginar aquest fet porta problemes provocats pel biaix d'informació o falta en la precisió del record i la quantitat de coneixements del consumidor. És per això que l'única manera que tenim, actualment, per saber si s'està subministrant la dosi correcta de metadona als pacients és contrastar el grau de satisfacció i confort dels mateixos en relació a la quantitat de metadona subministrada.

Per fer-nos una idea de l'equivalència entre la dosi d'heroïna i de metadona es presenta la següent taula:

Heroïna	Metadona
¼ g	30-50 mg
½ g	50-70 mg
¾ g	70-90 mg
1 g	90-120 mg

*Taula 1. 1 Conversió heroïna/metadona<sup>3</sup>.*

Existeixen dos programes de manteniment de metadona (PMM):

- PMM de baix llindar d'exigència terapèutica:

L'únic objectiu d'aquest programa és tenir a la població toxicòmana que no vol abandonar el consum d'opiàcis abastida amb metadona, sense exigir l'abandonament de l'heroïna com a condició indispensable, per tal de tenir allunyada la població de la seva anterior vida caòtica, al no tenir la necessitat d'obtenir la dosi diària per pal·liar els símptomes de la deprivació.

Els únics paràmetres a avaluar en aquest programa són la disminució de la mortalitat relacionada amb el consum d'heroïna i el decreixement de la delinqüència.

- PMM d'alt llindar d'exigència terapèutica:

L'objectiu principal és la desintoxicació, però no és vàlid en tots els pacients.

Els paràmetres que són més interessants de cara a l'avaluació del programa són:

- Consum de drogues. L'abstinència de l'heroïna, la sobrietat en l'alcohol, cocaïna i altres psicotròpics segueixen sent condició necessària per a un bon resultat.
- Delinqüència. Referent fonamentalment per la societat general.
- Ús de succedanis (productivitat no remunerada, activitat recreativa, etc.). Identificable amb un bon pronòstic en la rehabilitació social desitjada.

Com més objectius es compleixin, major serà la normalització i la inserció del pacient, tot i estar consumint metadona.

Es recomana el tractament amb metadona en centres hospitalaris per a aquells individus addictes a drogues per via parental, que presentin mal estat físic i/o psicològic i amb un suport familiar i social insuficient per garantir un tractament adequat.

Podríem resumir el tractament amb metadona com substituir un estupefaent opiàci per un altre <<legal>> (desproveït, en principi, dels riscos del mercat negre, de l'adulteració de la substància i de la seva contaminació) prenent control la dependència amb el mínim risc possible. És per això que tot el procediment ha d'estar totalment regulat i controlat.

### *Discussió dosi*

La dosi de metadona, junt amb la duració del tractament, són factors molt importants per assegurar l'efectivitat del programa de manteniment de metadona. Per tant, és molt important la prescripció de dosis òptimes. La dosificació ha de ser individualitzada, ajustant-se en cada cas a l'estat del pacient i als efectes farmacològics desitjats.

L'objectiu general és proporcionar la dosi adequada durant el temps suficient.

Tot i que no existeixen evidències d'un temps òptim de duració del tractament s'aconsella que no sigui inferior a dos anys.

### *Problemes dels tractaments de metadona*

Els tractaments de manteniment amb metadona (TMM) s'han vist afectats pel fet que molts consumidors d'heroïna són també dependents de cocaïna, substància que no es beneficia de l'acció substitutiva de la metadona i per la qual no existeix cap tractament específic efectiu ben comprovat<sup>4</sup>.

Com qualsevol altre tractament, la metadona també presenta efectes secundaris. Els més freqüents, a l'inici del tractament són: sedació, nàusees, vòmits... Però ràpidament es desenvolupa tolerància a aquests efectes. Ocasionalment, es presenten altres efectes secundaris com pruíja (picor constant i desig de rascar-se), hipotensió postural (lleugers marejos), amenorrea (absència de la menstruació) i disminució de la libido.

En subministraments amb dosis elevades, en pacients amb efectes intolerants al tractament o que consumeixen altres substàncies (com fàrmacs diferents o alcohol) també pot provocar depressió respiratòria, que sens dubte seria l'efecte advers més greu.

#### **1.4. Tractament amb metadona a Espanya**

L'objectiu de tota intervenció en desintoxicació ha de ser aconseguir l'eliminació de la simptomatologia que provoca l'abstinència, així com el restabliment de l'estat físic de l'individu per aconseguir una situació lliure de drogues, que posi al mateix en les millors condicions possibles per iniciar el procés de deshabitació, que és el pas següent.

Després de molts estudis s'ha arribat a la conclusió que si no existeix una teràpia coadjuvant d'un altre tipus, a més a més del tractament amb metadona, ja sigui psicoteràpia individual, grupal, familiar o programes de reinserció, els percentatges d'èxit són escassos, al mateix temps que per augmentar el percentatge d'èxits o de retenció en programes s'ha de definir de forma detallada el perfil del pacient que pot beneficiar-se d'aquest tipus de tractament.



Però per arribar a aquestes conclusions s'han anat definint i marcant al transcurs del temps, i cada vegada de manera més exacta, tots els elements a tenir en compte per realitzar degudament un programa de manteniment de metadona.

### *Cronologia*

Fins a 1983 els tractaments amb metadona es feien mitjançant la sol·licitud per part dels pacients d'un carnet d'extradosi que regulaven els col·legis de metges, i que originalment estaven orientats fonamentalment a pacients terminals que rebien opiacis de forma continuada. Paral·lelament, en el sector privat es començaven a realitzar programes de manteniment amb metadona, que consistien en la prescripció de la recepta. Aquest fet va causar estralls en la població, ja que els toxicòmans preferien aquest tractament per sobre dels programes lliures de fàrmacs i drogues, ja que els facilitava l'accés a un opiaci, mal regulat. Com a conseqüència de tot això va començar a circular la metadona per temes extra terapèutics, creant un veritable mercat il·legal.

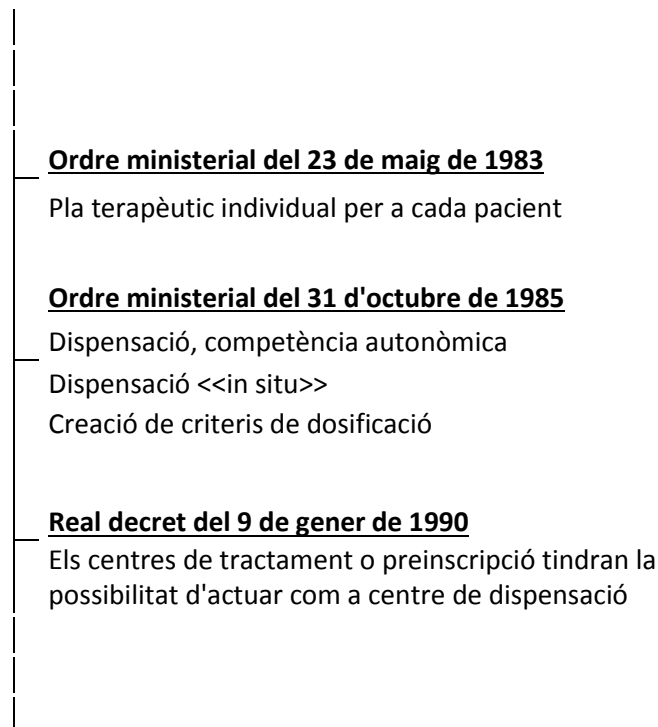
Mitjançant l'Ordre ministerial del 23 de maig de 1983 es va crear una normativa relativa a la prescripció, ús i dispensació de metadona per al tractament de toxicòmans dependents d'opiacis. Aquesta ordre feia referència a la necessitat de realitzar per a cada pacient un pla terapèutic individual, fent responsable de la seva aprovació als serveis corresponents de la Salut Pública. La finalitat última de l'ordre seria determinar la necessitat d'englobar la utilització de la metadona en un pla terapèutic més ampli, en el qual el fàrmac fos considerat com una eina tècnica més que com un mètode terapèutic.

A través de l'Ordre ministerial del 31 d'octubre de 1985 es defineixen les competències en la regulació d'aquests programes. S'estableix que s'haurà de crear en cada comunitat autònoma una comissió parietària, que procurarà el bon compliment dels perceptors contemplats en la normativa. A més a més, l'ordre també determinava que la dispensació de la substància hauria de realitzar-se en centres determinats amb la finalitat d'impedir el desviament de metadona al tràfic il·legal i així assegurar que la dosi prescrita per el terapeuta fos correctament administrada. Per últim, l'Ordre també va incloure uns criteris de dosificació i inclusió i per a poder ser admès en els programes sabien de complir alguns dels següents requisits:

- Majors de 18 anys.
- Absència de patologies psiquiàtriques greus.
- Heroïnòmans amb malalties cròniques greus (infecció per VIH, hepatopaties greus, etc.).
- Heroïnòmans de llarga durada amb diferents fracassos terapèutics de desintoxicació.
- Dones en període de gestació.

Amb el Real Decret del 9 de gener de 1990 es mantenia l'administració en forma de solució extemporània sempre que fos possible, l'exigència de les comissions d'acreditació, avaluació i control de centres o serveis, els criteris per acreditació de centres o serveis de tractaments, l'exigència d'un pla terapèutic individualitzat i en el sistema de notificació del número de pacients en tractament.

Pel que fa a l'admissió en el tractament, s'exigeix prèviament el diagnòstic confirmat de dependència d'opiacis i haver realitzat com a mínim un intent de tractament en una altra modalitat terapèutica, exceptuant aquelles persones dependents d'opiacis que hagin agafat la infecció pel virus de la immunodeficiència humana o es trobin afectats per patologia orgànica severa. Sens dubte, la innovació més important és que els centres de tractament o prescriptors tindran la possibilitat d'actuar com centre de dispensació.



*Figura 1. 1 Cronologia del tractament amb metadona.*

## 2. Presentació de les dades

En aquest treball hem realitzat les tasques corresponents amb dues bases de dades, una que es diu “metadona\_episodis”, que després de les modificacions realitzades en l’apartat següent a aquest s’anomenarà “Episodis” i un altre que es diu “dosis\_prescrites” que l’acabarem anomenant “Dosis”. A continuació es presenta un petit resum del significat de les variables que tenim a la base de dades Episodis:

Aquesta base de dades te 36 variables diferents per a cada un dels 880 individus corresponents, i hi ha tant variables categòriques com contínues com tipus data.

A continuació es presenten les variables categòriques disponibles:

Variables categòriques
Pacient actiu
Sexe
Consum de cocaïna
Antecedents penals
Situació laboral
Nivell estudis adquirits
Ha utilitzat la via intravenosa alguna vegada en la vida?
Ha utilitzat la via intravenosa en els darrers 6 mesos?
En cas d’haver utilitzat la via intravenosa, ha participat en algun programa d’intercanvi de xeringues?
Ha tingut relacions sexuals a canvi de diners o drogues?
Ús del preservatiu en les relacions
Infecció per VIH
Infecció per virus hepatitis B (VHB)
Infecció per virus hepatitis C (VHC)
Prova d’infecció per sífilis ( <i>Veneral Disease Research Laboratory</i> )
prova d’infecció per sífilis ( <i>Treponema Pallidum Haemagglutination</i> )
Antecedent de sífilis
Antecedent de Gonocòccia
Tuberculosis prèvia
Mort (tallant el seguiment a 31/12/2013)
Número de parelles en els darrers 6 mesos

Taula 2.1 Variables categòriques de la base de dades Episodis.

Variables continues:

<b>Variables continues</b>
Número d'història clínica del pacient
Identificador de l'episodi
Identificador individu
Dosi inicial de metadona
Edat inici consum heroïna
Edat inici consum heroïna per via parenteral (intravenosa)
Nombre de CD4 (és un marcador biològic indicador de l'estat immunitari i normalment es mesura en els pacients amb VIH)
Edat en el moment de l'inici de tractament

*Taula 2.2 Variables continues de la base de dades Episodis.*

Variables tipus data:

<b>Variables tipus data</b>
Data inici del tractament
Data final del tractament
Data naixement
Data inici tractament (en continu)
Data de naixement (en continu)
Data de la mort (tallant el seguiment a 31/12/2013)*
Darrera vegada vist viu (tallant el seguiment a 31/12/2013)

*Taula 2.3 Variables tipus data de la base de dades Episodis.*

L'altre base de dades de la que disposem es la base Dosis en aquesta només tenim 3 variables diferents, però 22062 observacions. Cada una de les observacions correspon amb les visites dels individus, per cada una d'elles tenim un identificador de l'individu, la data de la visita i la dosi prescrita.

Cal remarcar que els individus de les dues bases de dades son els mateixos i per tant les dues bases de dades es complementen.

### 3. Preprocessament de les dades

En aquest apartat del treball és on realitzem la neteja de les dades, la recodificació d'algunes variables i la creació noves, per poder complir amb els objectius del treball.

Per realitzar aquesta part del treball hem utilitzat entre d'altres la funció “selectByDate” del paquet “openair”<sup>2</sup> d'R, que serveix per agilitzar els treballs realitzats amb variables tipus data.

Quan hem trobat alguna variable inusual o abans de fer qualsevol canvi en les dades hem consultat amb l'equip de recerca que ens ha facilitat les dades.

#### 3.1. Tractament d'observacions inusuals.

En aquest apartat parlem de totes aquelles observacions fora del comú i les decisions que hem pres amb elles. Hem trobat problemes d'aquests en les dues bases de dades disponibles.

Base de dades Dosis:

- Teníem un individu que presentava observacions de l'any 1011 (Ind. 414). Hem corregit l'error posant 2011 en comptes del valor anterior, deduint que era un error de transcripció.
- Observem valor posteriors al moment de tallament de l'estudi, això no ens ha de sorprendre perquè aquesta base de dades se'ns va facilitar tal qual sense tenir en compte fins quin data duraria l'estudi, aquests els vam eliminar.
- Apareixen observacions de l'any 0199. Aquestes les eliminem, perquè ens és impossible deduir en quin any es van fer les revisions i no és de vital importància per l'objectiu que tenim marcat amb aquestes dades.

Base de dades Episodis:

- Ens en adonem de que hi ha individus que tenen una data de final de tractament posterior a la data de mort. Això és degut a que aquests pacients tal dia van deixar d'anar a recollir el tractament, però no se sabia exactament que havia passat amb ells. A falta d'informació s'havia fet una estimació del final de tractament. Posteriorment es van creuar les dades amb el registre oficial de mortalitat i van aparèixer com a morts, però no es va arreglar la variable fi de tractament. És per això que a tots els individus que els hi succeeix aquest fet els assignem com a data de final de tractament la de la mort.

### 3.2. Recodificació de variables.

En aquest apartat hem de comentar que hem truncat la variable edat, ja que tenia molts decimals i pesava més el cost de tractament de la variable que la informació que perdem per truncar-la.

A més a més, per treballar més a gust i que tots els gràfics quedessin bé hem recodificat quatre de les variables categòriques: el consum de cocaïna, la situació laboral, el nivell d'estudis i l'ús de preservatiu en les relacions. La forma com hem recodificat les variables la presentarem junt als gràfics i taules en les que aparegui aquesta variable, per aclarir el seu significat.

### 3.3. Noves variables.

Per a la part de l'estudi de l'anàlisi de supervivència hem hagut de crear dues variables noves, que son "estat", per a saber si un individu esta censurat o no (0 censurat i 1 no censurat), i "sup", que correspon al temps de supervivència fins l'aparició de l'esdeveniment, que en el nostre cas és el moment en el que deixen el tractament. Entendrem com a individu censurat tots aquells individus que moren abans de deixar el tractament o que en la data de final de tractament segueixen actius en aquest, i com a no censurat a tots aquells individus restants que corresponen amb els individus que han deixat el tractament de manteniment amb metadona. La decisió de tractar als individus que tenen com a data final de tractament la data de la mort com a censurats ha estat pròpia, els podíem haver extret de la base de dades o haver tractat com a no censurat considerant que si l'individu no hagués mort, hagués deixat el tractament (per malaltia o qualsevol altre causa).

Després de fer totes les modificacions pertinents creem les noves bases de dades ja anomenades anteriorment, Dosis i Episodis, i les guardem amb format Excel i com a *data frames* en diferents àrees de treball d'R, per treballar a partir d'aquest moment sobre d'aquestes.

## 4. Anàlisi descriptiva

En aquest apartat del treball presentem una anàlisi inicial de les dades, amb taules de freqüències i gràfiques. Cal recordar que l'estudi es realitza sobre una base de dades de 880 individus amb 36 variables de les quals cinc són contínues, set són tipus data i la resta categòriques.

### 4.1. Variables categòriques.

A continuació es mostren les taules de freqüències i de percentatges de les variables categòriques més habituals. Els percentatges de les taules han estat calculats en base al total de respostes, sense tenir en compte totes les "no respostes".

Pacient actiu en l'estudi	n	%	Missings
Sí	113	12,84	0
No	767	87,16	
Sexe	n	%	Missings
Home	735	83,52	0
Dona	145	16,48	
Antecedents presó	n	%	Missings
No	294	38,53	117
Sí	469	61,47	(13,30%)
Situació laboral	n	%	Missings
Sí, fixe	158	20,84	
Sí, ocasionalment	323	42,61	
Mai	39	5,15	122
Atur retribuït	66	8,71	(13,86%)
Atur no retribuït	84	11,08	
Incapacitat	88	11,61	
Nivell d'estudis	n	%	Missings
Analfabet	23	3,03	
Primaris	524	69,04	
FP	180	23,72	121
Batxillerat superior	21	2,77	(13,75%)
Superiors	9	1,19	
Carrera tècnica	2	0,26	
Núm. de parelles ult. 6 mesos	n	%	Missings
0	170	22,91	
1	419	56,47	138
2	80	10,78	(15,68%)
>2	73	9,84	
Mort tallant seguiment al 2013	n	%	Missings
No	631	71,7	0
Sí	249	28,3	

Taula 4.1 Taula de freqüències de variables categòriques bàsiques.

Veiem que deixant de banda les variables observables, tota la resta té una quantitat de missings (dades mancants) molt elevada i això ens pot portar problemes a l'hora de realitzar les anàlisis.

La majoria dels individus de l'estudi (735) són homes i això equival a un 83,52% del total d'observacions. Més de la meitat dels individus (un 61,47%) tenen antecedents penals, tot i que hem de tenir en compte que en aquesta variable tenim 117 dades mancants, un 13,30% del total.

Més de la meitat dels pacients del TMM (tractament de manteniment amb metadona), un 69,04%, tenen estudis primaris i un 23,72% té estudis secundaris. Cal remarcar que un 3,03% dels individus són analfabets, un percentatge més elevat que el d'individus amb batxillerat, que és d'un 2,77%. Només un 1,45% tenen estudis superiors o carreres.

Dels 880 individus que han format part de l'estudi durant els anys analitzats (de 1995 a 2013), en el moment del tallament només hi han 113 individus actius.

Consum cocaïna	N	%	Missings
No	466	62,22	131 (14,89%)
Esporàdicament	191	25,5	
Diàriament	92	12,28	
Via venosa últims 6 mesos	N	%	Missings
Sí	479	62,37	112 (12,73%)
No	289	37,63	
Via venosa	N	%	Missings
Sí	665	86,03	107 (12,16%)
No	108	13,97	
Programari intercanvi	N	%	Missings
Sí	228	30,40	130 (14,77%)
No	522	69,60	
Relacions per diners o drogues	N	%	Missings
Sí	42	5,61	131 (14,89%)
No	707	94,39	
Us preservatiu	N	%	Missings
Sempre	256	34,18	131 (14,89%)
Quasi sempre	34	4,54	
De vegades	57	7,61	
Quasi mai	20	2,67	
Mai	213	28,44	
No té relacions sexuals	169	22,56	

Taula 4.2 Taula de freqüències de variables categòriques relacionades amb les drogues.

Com podem observar totes les variables d'aquesta última taula presenten entre un 12 i un 15% de missings, un nombre bastant elevat.



Dels 880 individus estudiats 665 han consumit alguna vegada heroïna per via intravenosa, dels quals 479 ho han fet en els últims 6 mesos. Només 108 individus no se l'han subministrat per via parenteral i en els últims 6 mesos aquesta xifra a augmentat fins a 289 individus.

Tan sols un 30,4% dels individus han format part de programes d'intercanvis de xeringues, quan la taxa per via parenteral supera el 60% fins i tot en els últims 6 mesos.

En relació amb l'ús del preservatiu veiem que les categories que tenen un major percentatge són "Sempre", "Mai" i "No té relacions sexuals", 34,18%, 28,44% i 22,56% respectivament.

Hepatitis B	N	%	Missings
Sí	425	59,03	160
No	295	40,97	(18,18%)
Hepatitis C	N	%	Missings
Sí	592	80,98	149
No	139	19,02	(16,96%)
Sífilis (VDRL)	N	%	Missings
Positiu	57	10,61	343
Negatiu	480	89,39	(38,98%)
Sífilis (TPHA)	N	%	Missings
Positiu	9	1,67	342
Negatiu	529	98,33	(38,86%)
Antecedents sífilis	N	%	Missings
Sí	26	3,45	127
No	727	96,55	(14,43%)
Antecedents gonocòccia	N	%	Missings
Sí	47	6,36	141
No	692	93,64	(16,02%)
TBC prèvia	N	%	Missings
Sí	93	11,98	104
No	469	60,44	(11,82%)
PPD + PREVI	214	27,58	
VIH	N	%	Missings
Sí	412	52,69	98
No	370	47,31	(11,14%)

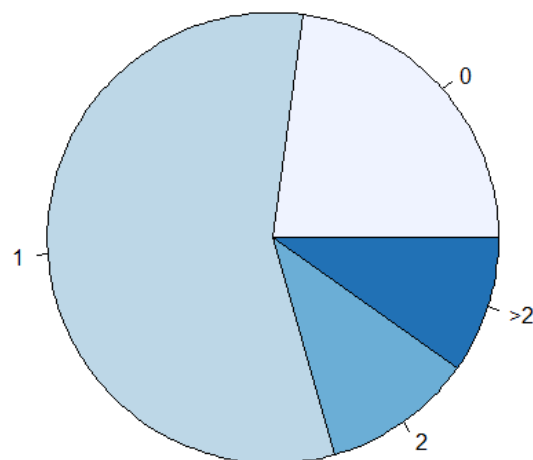
Taula 4.3 Taula de freqüències de variables categòriques relacionades amb diferents malalties.

En aquesta taula hem de ressaltar que en les variables "Sífilis (VDRL)" i "Sífilis (TPHA)" el nostre nivell de no resposta és molt superior que per la resta, però el nivell de respostes positives és del 10,61% i del 1,67% respectivament.

Veiem que una mica més de la meitat dels individus estudiats presenten VIH (412 individus) i/o Hepatitis B (425 individus). No necessàriament els mateixos que pateixen una malaltia han de patir l'altre.

Un 80,98% de les respostes obtingudes dels pacients del tractament de manteniment amb metadona presenten Hepatitis C, deixant només 139 lliures d'aquesta. Un 3,45% mostren antecedents de sífilis, per tant un 96,55% dels individus de l'estudi no indiquen antecedents d'aquesta mena. Només 47 individus tenen antecedent de gonocòccia, per tant 692 pacients no en manifesten.

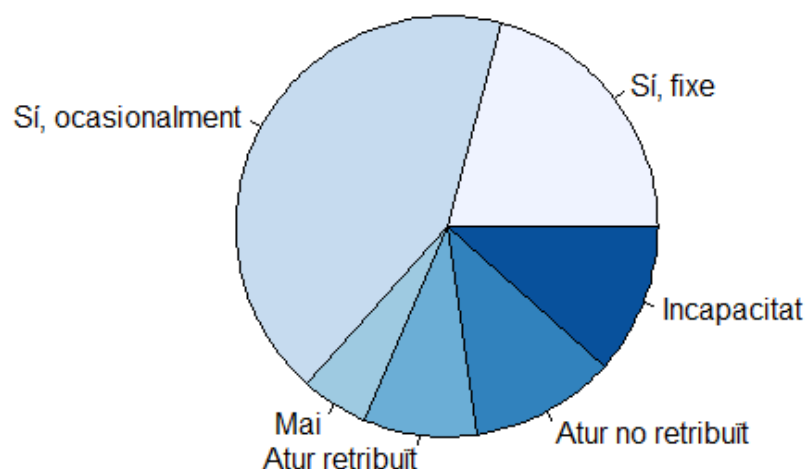
#### Número de parelles en els últims 6 mesos



*Gràfic 4.1 Gràfic de sectors de la variable número de parelles en els darrers 6 mesos.*

Com veiem en el gràfic de sectors la gran majoria d'individus en els darrers 6 mesos han tingut una parella, aquesta majoria equival al 56,47%. Un 22,91% no han tingut cap relació i les altres dues categories es distribueixen d'igual manera en la resta dels individus.

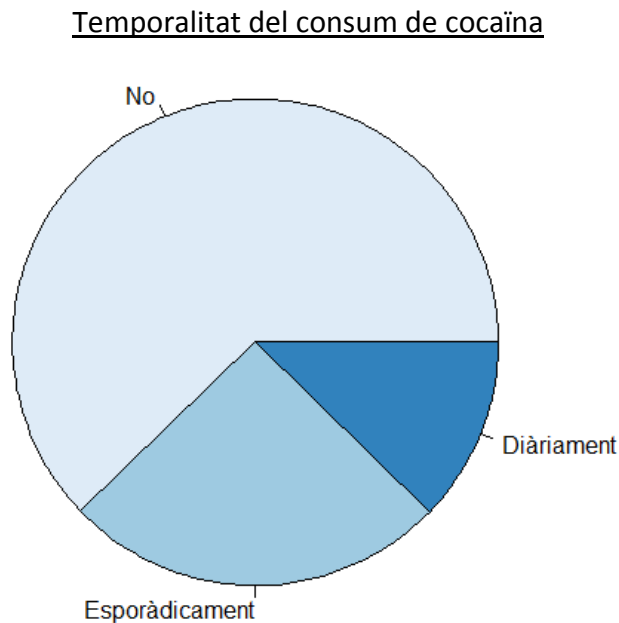
#### Situació laboral



*Gràfic 4.2 Gràfic de sectors de la variable situació laboral.*

Un 42,61% dels individus tenen treball tot i no ser fixe, aquest percentatge equival a 323 pacients de l'estudi. Un 20,84% tenen treball fixe, 158 individus. En l'estudi tenim 150 persones que estan actualment parades, 66 disposen d'algun tipus de compensació econòmica i 84, per tant un 11,08%, no disposen de retribucions.

Dels pacients del estudi 39 no han treballat mai i 88 tenen una situació d'incapacitat, sense dividir entre discapacitats cròniques i no cròniques.



**Gràfic 4. 3 Gràfic de sectors de la variable consum de cocaïna.**

Observem que més de la meitat dels individus, un 62,22%, no consumeix cocaïna, la resta un 25,5% i un 12,28% en consumeixen esporàdicament i diàriament respectivament. Recordem que el consum d'altres substàncies, especialment la cocaïna, dificulta l'èxit del tractament de manteniment amb metadona i en aquest estudi tenim un 37,78% dels individus que amb més o menys regularitat en consumeixen.

Remarcar que aquesta última variable és l'única que tenim que conté relació amb el consum d'altres substàncies, tant il·legals (la cocaïna entre d'altres), com legals (medicaments).

## 4.2. Variables contínues

Tot seguit presentem un resum de les variables contínues:

Variable	Mitjana	Sd	CV(%)	Mínim	Màxim	Mediana	Missings
Dosi inicial de metadona	54,94	38,02	69,19	5	350	50	0
Edat d'inici consum heroïna	19,15	4,83	25,23	9	45	18	8
Edat d'inici per via parenteral	20,24	5,21	25,75	9	40	19	127
Nombre de (CD4)	403,2	344,76	85,49	2	2297	328	517
Edat a l'inici del tractament	33	7,03	21,24	17	60	33	0

Taula 4.4 Taula resum de les variables contínues.

Veiem que les dues variables referents a la situació d'inici del tractament, dosi inicial de metadona i edat al començament del TMM, no tenen *missings*, per contra la variable "nombre de CD4" presenta una quantitat de respostes mancants molt elevades, això succeeix perquè aquesta mesura només es mira en individus amb VIH, a l'hora de crear els models i d'utilitzar-la ho tindrem en compte. Veiem que per la variable edat inici consum d'heroïna els *missings* són insignificants i per tant no suposaran un problema, en canvi per l'edat de començar consumir per via parenteral, la quantitat de *missings* augmenta, però hem de tenir en compte que tenim 108 individus que no han consumit mai per via parenteral hi ha mes a més, per aquesta variable, també teníem 107 *missings*.

Pel que fa al coeficient de variació, la variable que presenta més diversitat, a més a més molt elevada, és la de "nombre de CD4".

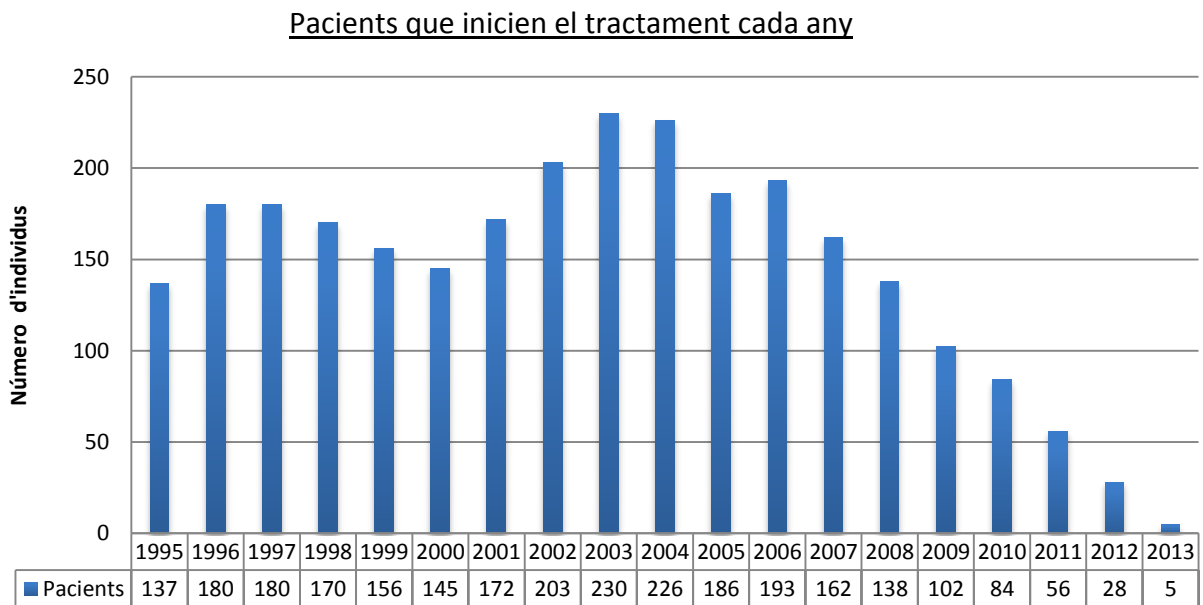
En relació a l'edat d'inici de consum de la droga estudiada, veiem que es situa als 19,15 anys, sent aquesta molt propera a l'edat d'inici del consum per via parenteral, que es situa als 20,24. Com la mediana està a l'esquerra de la mitjana, per a les dues variables, podem deduir que la distribució d'aquestes dues presenta una mica d'asimetria negativa. D'aquestes dues variables cal remarcar que el mínim trobat és de 9 anys, aquesta informació l'hem contrastat amb la responsable de les dades donat que inicialment ens va semblar que podria ser un error, i en adonar-nos que era una dada correcta vam filar més prim en aquests terminis d'edat. A continuació es presenta com es distribueixen les freqüències d'aquestes dues variables dels 9 als 15 anys per a què es pugui observar que no es tracta d'algun cas aïllat.

	Edat inici consum d'heroïna							Edat inici consum per via parenteral						
Edat	9	10	11	12	13	14	15	9	10	12	13	14	15	
Freq.	2	6	1	16	20	58	73	2	2	9	14	34	46	

Taula 4.5 Taula freqüències de variables de l'inici consum d'heroïna.

També veiem que l'edat mitjana d'iniciar el tractament es situa als 33 anys, aproximadament uns 13 anys després de l'edat a la qual comencen a prendre la substància.

En aquest apartat plasmem la quantitat d'individus que inicien el tractament de manteniment amb metadona, cada any, des de 1995 fins al 2013.

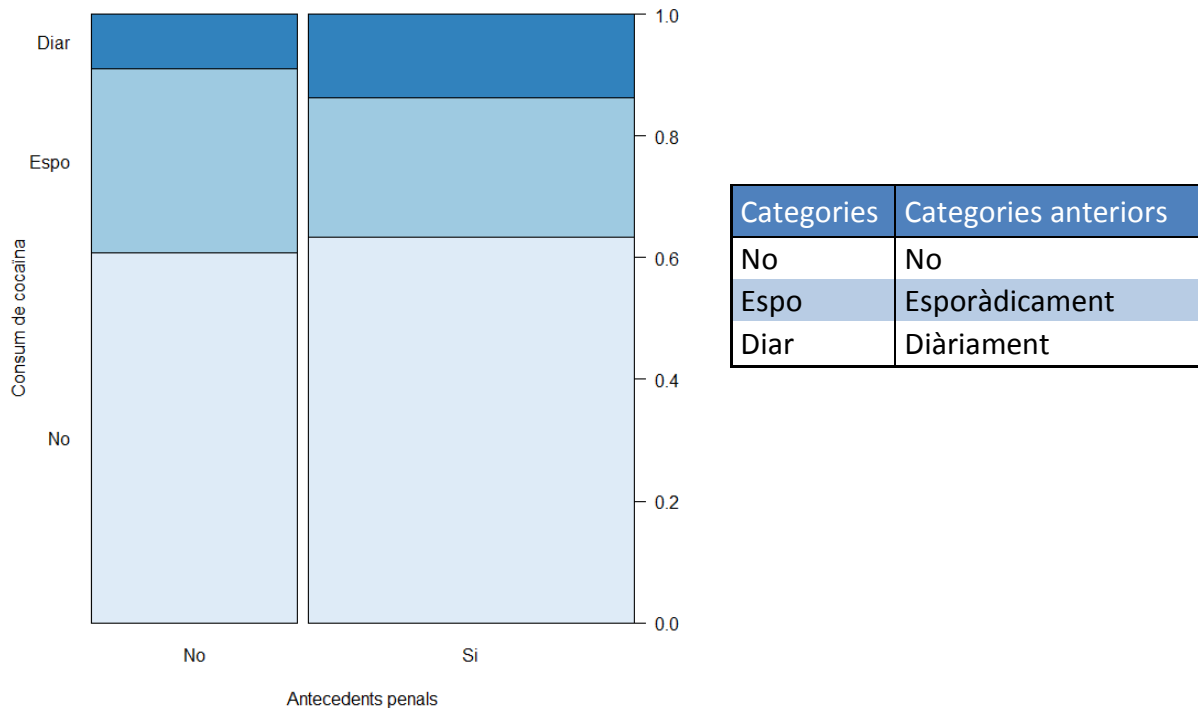


*Gràfic 4.4 Histograma de la freqüència d'individus que inicien el tractament per any.*

En el gràfic anterior es mostra la quantitat d'individus que inicien el tractament cada any, es una variable nova creada per nosaltres a traves de la data d'inici i final de tractament. Veiem que aquesta té dos màxims un al 1996-1997 amb 180 individus/any i el màxim absolut de la gràfica al 2003 amb 230 inscrits al tractament. Cal destacar que a partir de l'últim màxim el de 2006 la tendència de la gràfica és decreixent, fins que al 2013, any de tancament de l'estudi, només inicien el tractament 5 individus.

### 4.3. Anàlisi bivariant

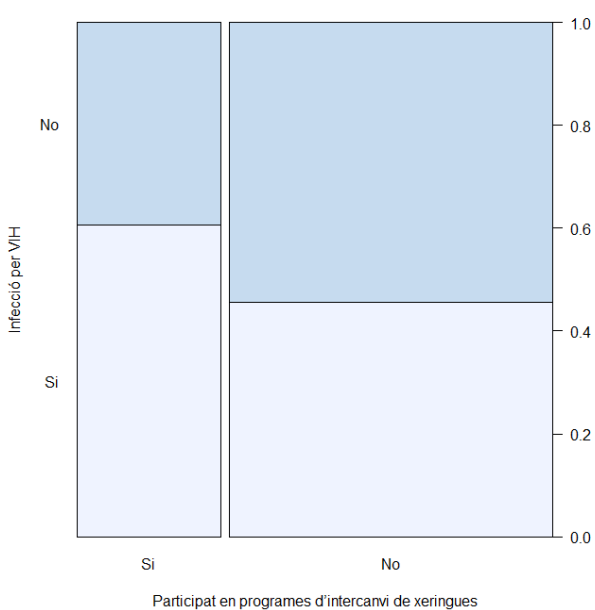
#### Variable Consum de cocaïna creuada amb la variable Antecedents penals



**Gràfic 4.5 Gràfic de mosaic de les variable i ús del preservatiu.**

D'aquest gràfic podem destacar les proporcions de consum de cocaïna les quals són molt semblants tant per els que tenen i per als que no tenen antecedents penals. La proporció de persones que consumeixen diàriament és major per als que tenen els antecedents i la proporció de persones que consumeixen esporàdicament és major per als que no en tenen.

#### Variable VIH creuada amb participació en programes d'intercanvi



Veiem que els individus que participen en programes d'intercanvi tenen una quantitat d'individus amb infecció d'VIH major que els que no participen.

Aquesta observació pot semblar contradictòria, però succeeix perquè les persones amb aquesta malaltia no només la pateixen sinó que en són portadors, i és molt rellevant que coneguin i es conscienciïn de la importància i la gravetat de contagiar-la a algú.

**Gràfic 4.6 Gràfic de mosaics de les variable VIH i participació en el programa de xeringues.**

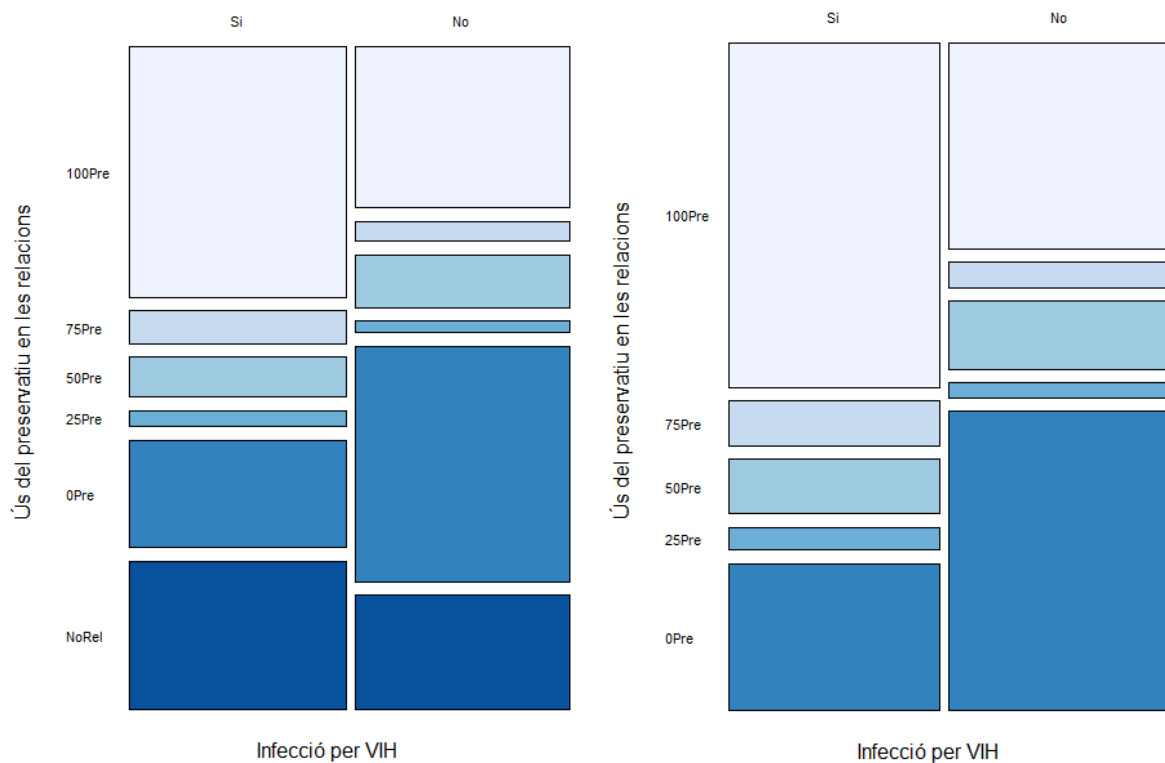
Definició categories variable ús del preservatiu

Categories	Categories anteriors
100Pre	Sempre
75Pre	Quasi sempre
50Pre	De vegades
25Pre	Quasi mai
0Pre	Mai
NoRel	No relacions sexuals

*Taula 4.6 Categories variable ús del preservatiu.*

La taula anterior ens serveix com a glossari per interpretar els dos gràfics següents:

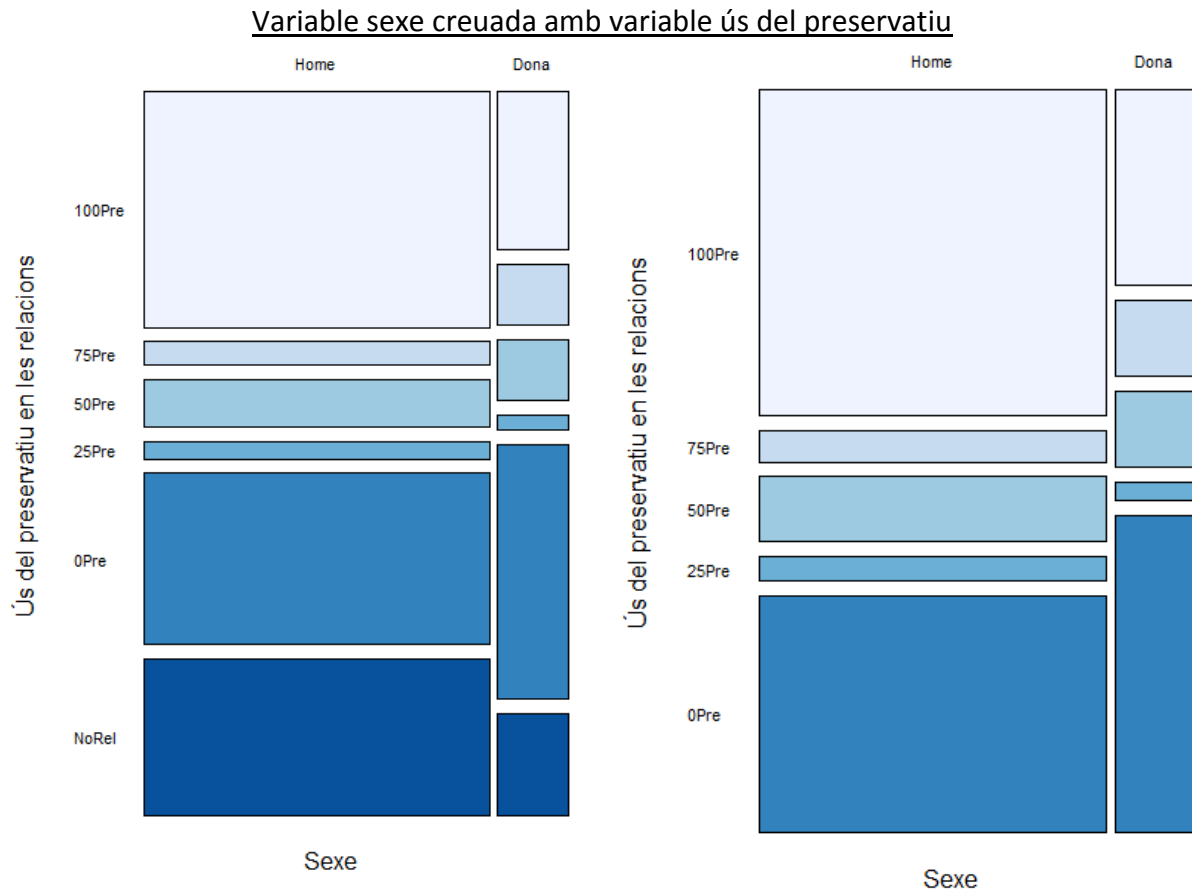
Variable VIH creuada amb variable ús del preservatiu amb i sense els individus que no tenen relacions



*Gràfic 4.7 Gràfic de mosaics de les variables VIH i ús del preservatiu.*

D'aquests dos gràfics se'n desprèn la proporció d'individus que tenen l'VIH segons la freqüència amb la que utilitzen el preservatiu, amb i sense tenir en compte si no tenen relacions sexuals. Veiem que els individus que sí que estan infectats tenen una proporció de "Sempre" més elevada que els que no la pateixen, aquest fet guarda relació amb el comentat anteriorment, no només la sofreixen sinó que en són portadors. La proporció d'individus que no n'utilitzen mai és molt més elevada per als quals no tenen la malaltia, quasi bé doblant la proporció dels que la tenen.

La diferència que hi ha entre les dues gràfiques és que una incorpora als individus que no tenen relacions i l'altre no. El mateix passarà en la gràfica següent:



*Gràfic 4.8 Gràfic de mosaics de les variables sexe i ús del preservatiu.*

Observem, com ja sabíem, que la quantitat de dones es més petita que la d'homes, elles en proporció no utilitzen el preservatiu més que els homes. També volem remarcar que l'índex, en el cas de les dones, que l'utilitzen sempre o quasi sempre, és més petita que la d'homes que l'utilitzen sempre.

#### 4.4. Anàlisi descriptiva base Dosis

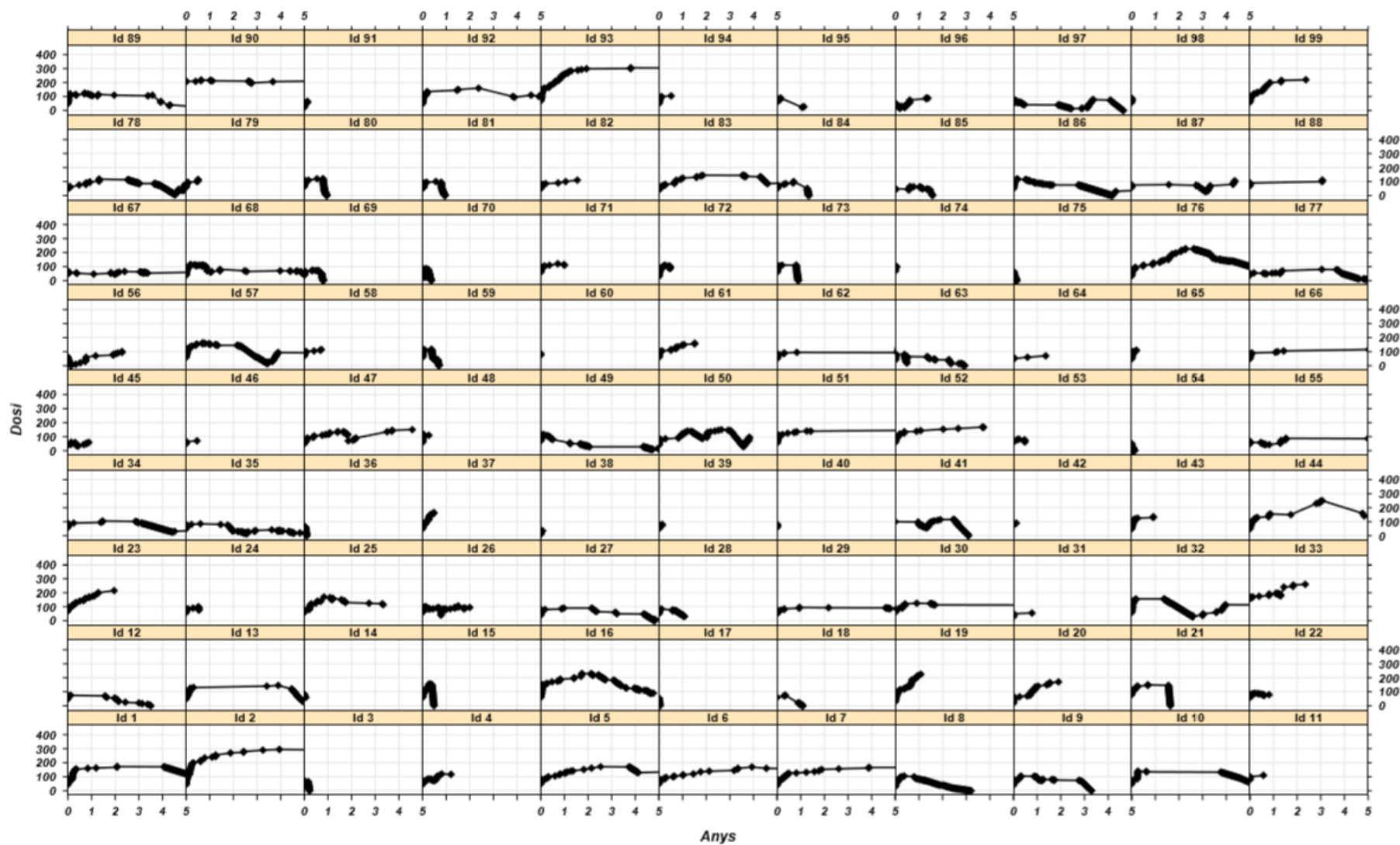
Per finalitzar l'anàlisi descriptiva visualitzem la tendència de les dosis prescrites, per a saber si cal tenir en compte alguna variable nova i per a fer-nos una idea de l'evolució d'aquesta. Aquesta tendència es calcula per a tots els individus, tot i que en l'informe només plasmem una petita mostra. Aquesta variable no la utilitzarem més durant l'estudi.

La gràfica l'hem realitzat amb la base de dades Dosis, que disposa de totes les dosis de metadona prescrites per a tots els pacients cada vegada que s'han realitzat canvis.

Una vegada realitzat tot l'anàlisi descriptiva, iniciem la modelització.



### Gràfica de tendències de la variable dosi prescrita



Gràfic 4.9 Gràfic de tendència de la variable dosis prescrita.

## 5. Modelització

En aquest apartat del treball el que ens interessa és estudiar les variables que influeixen en els pacients per deixar el tractament. El nostre objectiu és descobrir quines són, per que dins del possible, puguem controlar-les i disminuir la quantitat d'individus que el deixen.

Per realitzar les comandes d'aquesta part del treball hem utilitzat el paquet MASS d'R. Aquest ens proporciona un cúmul d'eines per a poder realitzar la modelització desitjada.

Per realitzar aquesta anàlisi hem exclòs els individus difunts, ja que no sabem que passaria amb aquests si seguissin vius. Tampoc tenim informació sobre el motiu de la mort ni si guarda o no relació amb el fet de què els individus abandonin el procés de manteniment amb metadona.

Per tant la nostra variable resposta és la següent:

$Y =$  Pacient actiu (si el individu està actualment actiu en el tractament o no).

Com la nostra variable resposta és binomial, només té dues categories, ajustem un model de regressió logística, que és un tipus de model lineal generalitzat. L'expressió d'aquest model és la següent:

$$Y = \begin{cases} 1 & \text{l'individu deixa el tractament} \\ 0 & \text{l'individu no deixa el tractament} \end{cases}$$

Per tant  $Y \in [0,1]$ , l'expressió del model és la següent:

$$\log \frac{p}{1-p} = \text{logit}(p) \in R = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 \dots \beta_n x_n$$

Om  $p = P(Y=1)$ . L'expressió anterior és equivalent en termes probabilístics a la següent:

$$p = \frac{\exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 \dots \beta_n x_n)}{1 + \exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 \dots \beta_n x_n)}$$

Per tant la regressió logística presentada anteriorment, es la modelització idònia per resultats binomials.

En aquest apartat del treball també utilitzarem l'odds ratio(OR), aquest ens serveix per aproximar el risc. Per calcular-lo dividim l'odds de deixar el tractament sota unes condicions per l'odds de deixar-lo sota les condicions complementàries. .

La manera de calcular-lo es la següent:

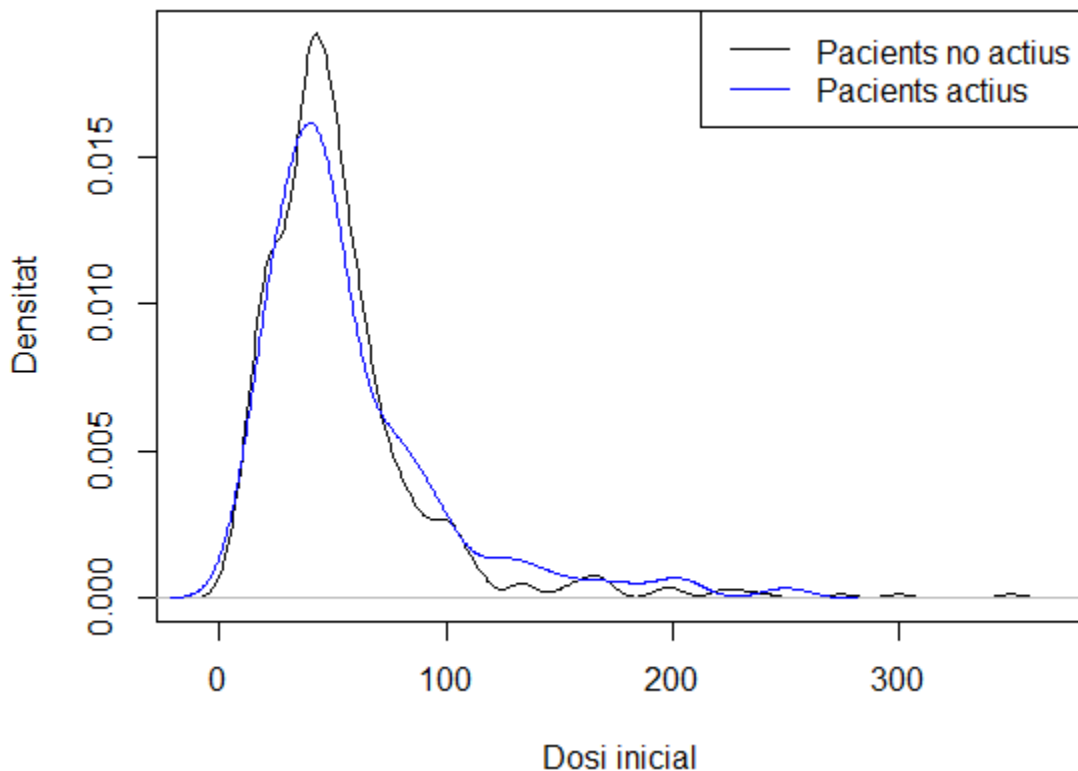
$$\widehat{OR} = \frac{P(Y = 1|X)/(1 - P(Y = 1|X))}{P(Y = 1|\bar{X})/(1 - P(Y = 1|\bar{X}))}$$

On Y és la nostra variable de resposta i X és una de les variables del model.

En definitiva, es pretén estimar la probabilitat que els individus deixin el tractament, segons unes variables explicatives.

Abans d'iniciar la cerca del nostre model, realitzem una primera anàlisi exploratòria de les variables contínues, que creiem que són interessants per incloure en el model:

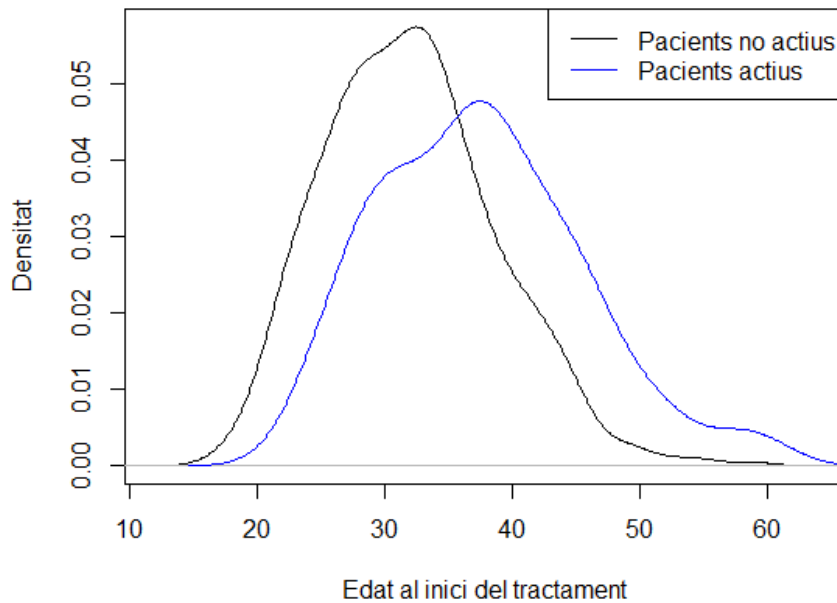
Gràfic de densitat de la variable dosi inicial



*Gràfic 5.1 Gràfic de densitat de la variable Dosis inicial.*

El gràfic anterior mostra la densitat segons la variable Dosi Inicial, diferenciant entre pacients actius i no actius (la línia blava correspon als pacients actius i la negra als pacients no actius). Veiem que tot i no ser exactament iguals les dues línies, tota l'estona es creuen, es per això que sembla que no serà una variable significativa.

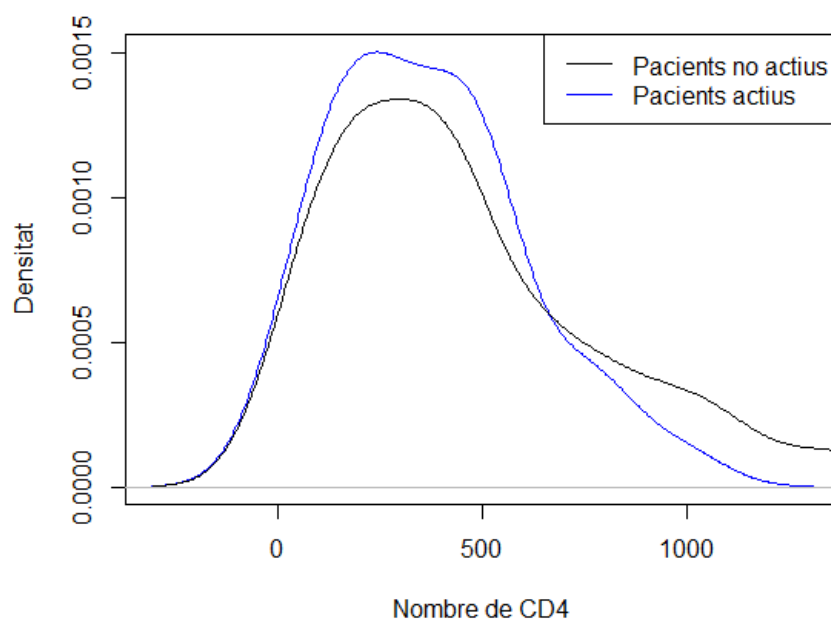
Gràfic de densitat de la variable Edat inici de tractament



*Gràfic 5.2 Gràfic de densitat de la variable Edat inici de tractament.*

En aquesta gràfica veiem la funció de densitat per a la variable Edat al inici del tractament, diferenciant per colors segons si es tracta d'individus actualment actius o no (línia blava = pacients actius, línia negra= pacients no actius). Sembla que hi ha diferències, la tendència de les línies canvia en moments de temps diferents i els centres de la campana i l'alçada no és la mateixa) entre les dues i que per tant pot valdre la pena incloure aquesta variable en el model. En el cas dels pacients no actius la densitat és major quan aquestes inicien el tractament més joves i a mesura que passen els anys la densitat dels actius en el tractament passa per sobre de l'altre.

Gràfic de densitat de la variable Edat inici de tractament

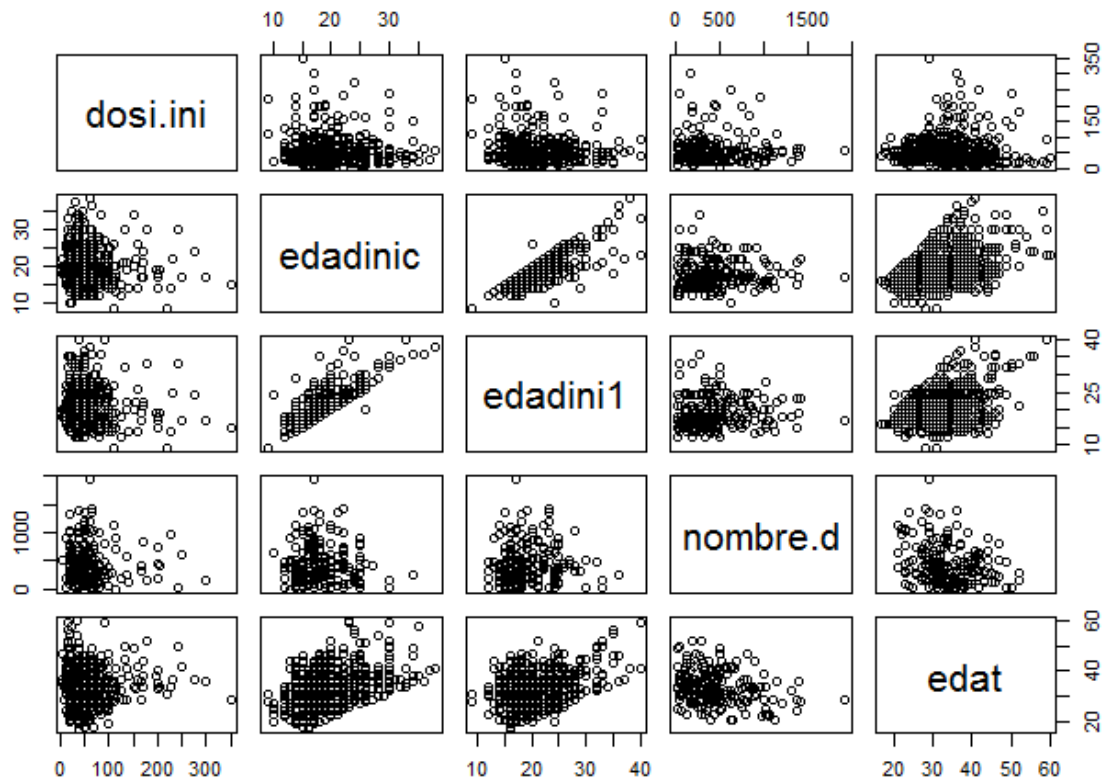


*Gràfic 5.3 Gràfic de densitat de la variable Nombre de CD4.*

En el gràfic anterior veiem com es distribueix la funció de densitat de la variable Nombre de CD4 segons si es tracta d'individus actius en l'estudi (línia blava) o individus no actius en l'estudi (línia negra). Com podem observar, aquestes dues línies tampoc es comporten de la mateixa manera, en el cas dels individus actius la densitat és major en valors més baixos.

També investiguem si tenim problemes de multicol·linealitat entre les nostres variables, i per a fer-ho realitzem el següent gràfic:

Gràfic de relacions entre les possibles variables explicatives del model



*Gràfic 5.4 Gràfic de relacions entre les variables contínues.*

En el gràfic anterior sembla que només hi ha presència de multicol·linealitat entre la variable d'edat d'inici de consum d'heroïna i l'edat d'inici de consum d'heroïna per via parenteral. Aquesta relació entre les dues variables no és d'estranyar, ja que per a que succeeixi la segona ha de succeir la primera. Per evitar problemes de multicol·linealitat el que farem serà, si fos necessari, incloure només una de les dues. En cas de què les dues fossin molt significatives, podríem realitzar canvis en una d'aquestes, com per exemple aplicar logaritmes.

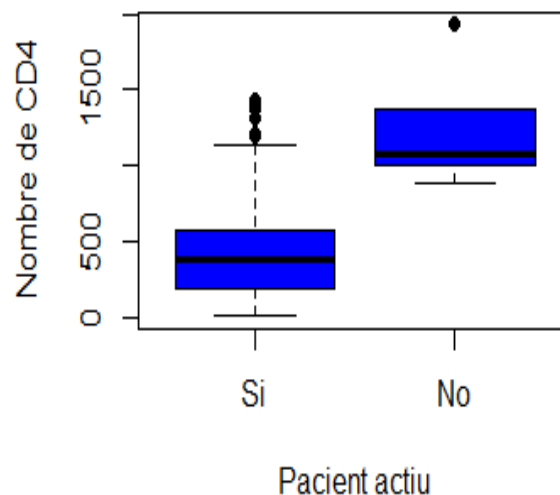
## 5.1. Cerca del model

Per aconseguir un bon model partirem des del model nul i, poc a poc anirem introduint variables de forma additiva, aquest serà el primer pas abans d'establir transformacions, inclusions d'ordre superior o interaccions entre variables si fos necessari. Per decidir si incloem o no una variable en el model ens fixarem en la significació global d'aquest i en la cerca del mínim AIC (criteri d'informació d'Akaike: és una mesura que té en compte la bondat de l'ajust del model i la complexitat d'aquest).

Abans de presentar el model definitiu i analitzar-lo expressem a continuació perquè no incloem la variable Nombre de CD4, ja que per a aquesta tenim en compte més factors que per a la resta.

Primer de tot realitzem un Boxplot de la variable anomenada i la variable resposta:

Boxplot variable Nombre de CD4 i Pacient actiu



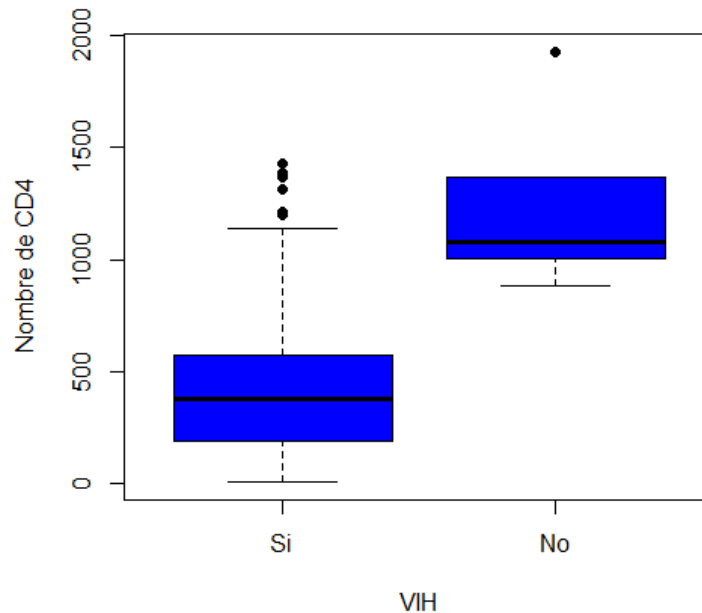
Gràfic 5.5 Boxplot de les variables Nombre de CD4 i Pacient actiu.

Veiem que les mitjanes són diferents segons si estan actualment actius en l'estudi o si no ho estan, també tenim grans diferències en els màxims i en els mínims. És per això que sembla interessant incloure aquesta variable en el nostre model.

Quan fem la prova i la introduïm en el model veiem que la significació d'aquest creix d'una manera notable, i que l'AIC és significativament més petit que l'obtingut fins al moment, però també veiem que perdem moltes dades (aquesta variable presenta 517 *missings*), per tant, deduïm que la millora que tenim del model ve provocada perquè s'ajusta correctament, però per a un nombre d'individus molt inferior. A més a més al incloure-hi la variable veiem que aquesta és no significativa. Per tant deduïm que la variable millora l'ajust però perdem moltes dades i per tant el preu d'incloure-la és massa elevat.

Per a no perdre informació decidim estudiar si aquesta variable guarda relació amb alguna altra, i obtenim el següent:

Boxplot variable Nombre de CD4 i infecció d'VIH



Gràfic 5.6 Boxplot variable Nombre de CD4 i Infecció d'VIH.

Observem que les dues variables plasmades en el gràfic estan relacionades donat que els infectats d'VIH presenten menor nombre de CD4, i els que no ho estan tenen un major nombre de CD4, per això conclouem que en cas de voler incloure una de les dues variables introduïrem VIH.

## 5.2. Model definitiu

Presentem el model definitiu al qual hem arribat seguint els passos comentats anteriorment:

Variable	$\hat{\beta}$	s. e. ( $\hat{\beta}$ )	P(>Z)	$\widehat{OR}$	CI ( $\widehat{OR}$ , 95%)
Intercept	4,63209	0,65681	1.76e-12 ***	-	-
Edat	-0,10513	0,01739	1.51e-09 ***	0,9	( 0,87 ; 0,93 )
Penales:Sí	0,41606	0,23805	0,0805 .	1,21	( 0,95 ; 2,42 )
Viavenos:Sí	0,42858	0,24247	0,0771 .	2,3	( 0,95 ; 2,47 )

(89 observations deleted due to missingness)

AIC: 476.37

Number of Fisher Scoring iterations: 5

Taula 5.1 Model resultant.

$$\text{Lògit } P(Y=1 | X) = 4,63209 - 0,10513 \beta_1 + 0,41606 \beta_2 + 0,42858 \beta_3$$

$$\text{Lògit } P(Y=1 | X) = 4,63209 - 0,10513 \text{ Edat} + 0,41606 \text{ Penales : Sí} + 0,42858 \text{ Viavenos : Sí}$$

Les categories de referència per al nostre model són els individus que no tenen antecedents penals i els que no han consumit per via parenteral.

### 5.3. Interpretació del model

#### *Interpretació en l'escala lineal*

El valor de les interpretacions següents és el valor  $\beta$  obtingut en el model sense sotmetre'l a cap modificació.

- Efecte de la variable Edat a l'inici del tractament: El logit de la probabilitat que els individus deixin el tractament disminueix 0,10513 per cada any més tard que comencin el TMM.
- Efecte de la variable Antecedents penals: El logit de la probabilitat de deixar el tractament augmenta 0,41606 si els individus tenen antecedents penals, respecta la categoria de referència (No antecedents penals) i mantenint constant la resta de variables.
- Efecte de la variable Consum per via venosa: El fet de consumir per a via venosa incrementa el logit de la probabilitat en 0,42858 en escala lineal i mantenint constant la resta de variables.

Per fer les interpretacions dels factors no cal tenir en compte interaccions, ja que el nostre model no en té.

#### *Interpretació de l'OR*

- Efecte de la variable Edat a l'inici del tractament: L'OR estimat per a la variable Edat és de 0,9. Com aquest valor és inferior a 1, tot i que poc, és un factor de protecció i per tant per a cada any més que triguen en iniciar el tractament obtenim millors resultats, sense tenir en compte perquè l'inicien més tard. Sens dubte aquesta variable, en més o menys mesura, està linealment relacionada amb l'edat a la qual comencen a consumir.
- Efecte de la variable Antecedents penals: L'OR de la variable antecedents penals és de 1,21, per tant aquest és un factor de risc. Si els pacients tenen antecedents penals el OR de deixar el tractament augmenta pel factor 1,21 el que equival a un augment d'un 21%.
- Efecte de la variable Consum per via venosa: L'OR de la variable consum per via venosa, és de 2,3, per tant també és un factor de risc i incrementa les probabilitats de deixar el tractament.



Aquestes últimes conclusions poden semblar contradictòries, però hem de tenir en compte que en la nostra variable resposta la categoria de referència són els pacients actualment actius en el tractament i per tant els que no l'han deixat.

#### 5.4. Diagnosi i validació del model

- Estudi d'outliers, que estudia les observacions atípiques respecte l'eix de coordenades:

Resultat test d'Outliers

	R-Student	p-valor
334	-2,381154	0,017259

*Taula 5.2 Resultat del test d'Outliers.*

Veiem que el p-valor és inferior a 0,05 i per tant podem rebutjar la hipòtesi nul·la i deduïm que no tenim outliers significatius, aquesta dada és així perquè el valor dels residus estudentitzats està entre -3 i +3.

-Control multicol·linealitat:

Resultat test de multicol·linealitat

Edat a l'inici del tractament	Antecedents Penals	Consum per via parenteral
1,108312	1,037127	1,071795

*Taula 5.3 Resultat del test de multicol·linealitat.*

Aquesta mesura correspon amb el grau de correlació existent entre les variables explicatives. Recordem que per analitzar els resultats d'aquests test, hem de tenir en compte que sempre hi ha presència de multicol·linealitat, però si els valors del test superessin el 5, tindríem problemes de multicol·linealitat, i en cas de superar el 10 tindríem problemes greus de multicol·linealitat.

Si ens fixem en el model veiem que no és el cas, tenim presència de multicol·linealitat, però no problemes i menys greus.

-Control de bondat de l'ajust:

Homser and Lemeshow Goodness-of-fit

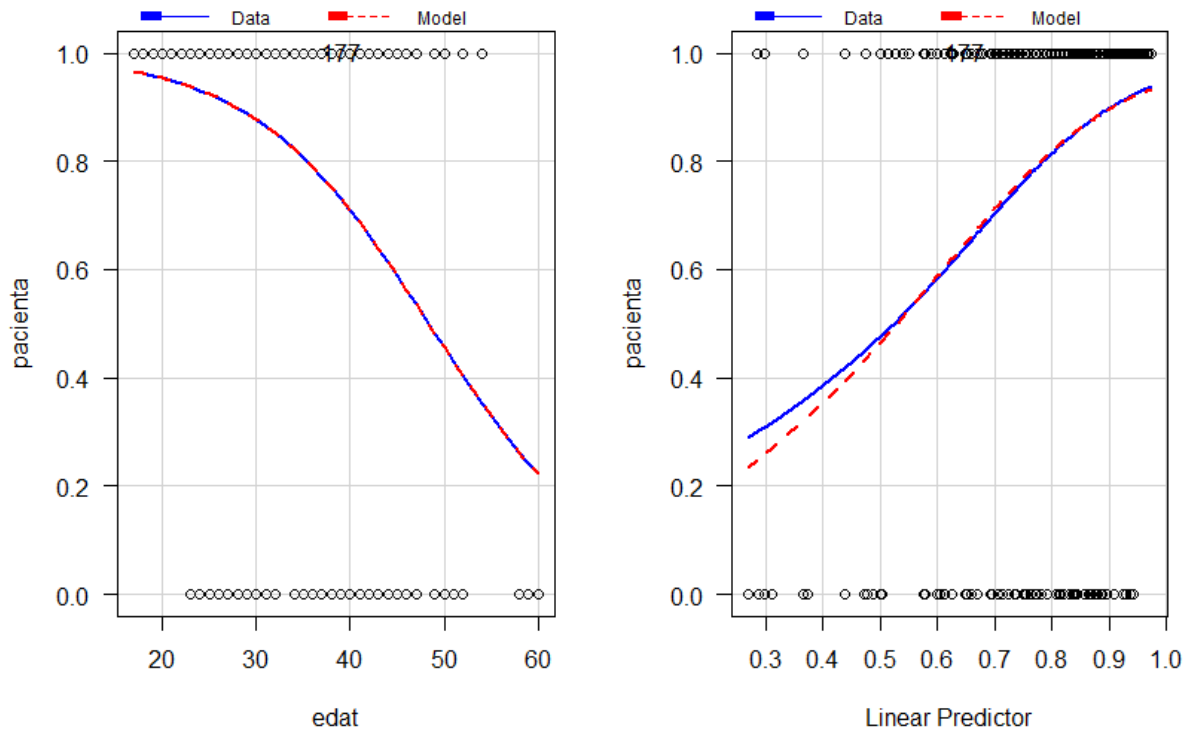
ChiSquare	Df	p-valor
2,894206	8	0,9408201

*Taula 5.4 Resultat del test de bondat de l'ajust.*

El valor obtingut del p-valor del test, és bastant elevat, per això conclourem que hem aconseguit normalitat en la bondat de l'ajust.

A continuació es presenta gràficament la bondat de l'ajust obtingut.

### Gràfic de bondat de l'ajust

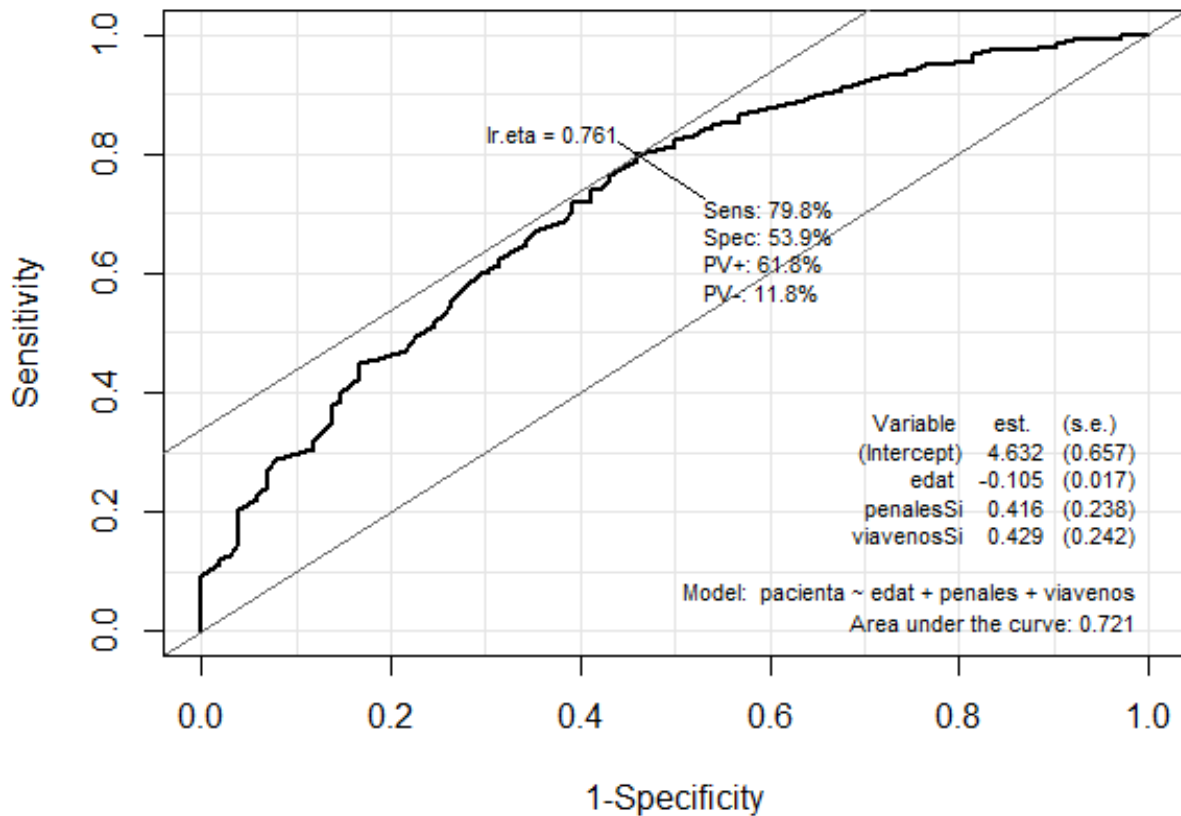


Gràfic 5.7 Bondat de l'ajust.

Al gràfic de l'esquerra està completament superposat per tant l'estimació segons l'evolució de l'edat és bona, i en el gràfic de la dreta que és on es representa el predictor lineal i la nostra estimació veiem que es superposa sobretot a partir del 0,5, però no ens sorprèn perquè abans tenim molt poques dades, per això no és molt preocupant aquest petit desajustament.

Per tal desbrinar la capacitat predictiva del model realitzem la següent corba Roc.

Gràfic de la representació de la corba ROC



Gràfic 5.8 Gràfica de la corba ROC.

Un altre paràmetre per avaluar la bondat de l'ajust és l'àrea sota la Corba ROC, aquesta pot prendre valors entre 1 (que correspondria amb una prova perfecte) i 0,5 ( que correspondria amb una prova amb probabilitats d'encertar igual a les de l'atzar). La nostra corba té una àrea a sota de 0,721. Per tant, podem dir que en un 72,1% dels casos el nostre model prediu correctament els valors de la variable resposta (Pacients actius en el tractament, di utilitzem com a punt de tall una probabilitat estimada de 0,761.

## 6. Anàlisi de supervivència.

Per realitzar les comandes d'aquesta part del treball hem utilitzat el paquet `survival`<sup>7</sup> d'R. Aquest ens proporciona un seguit de funcions molt potents per realitzar anàlisi de supervivència. També necessitem les variables `Estat` (0 censurat, 1 no censurat) i `Sup` (temps fins que abandonen el tractament mesurat en anys), de les quals hem parlat en l'apartat del processament de les dades. Remarquem que la censura que existeix en les nostres dades és per la dreta.

En aquest apartat del treball la nostre variable resposta és el temps de supervivència, aquest correspon amb el temps que estan en tractament.

El model principal d'aquest apartat és el model de Cox, aquest consta d'una part paramètrica i una no paramètrica. L'expressió d'aquesta és la següent:

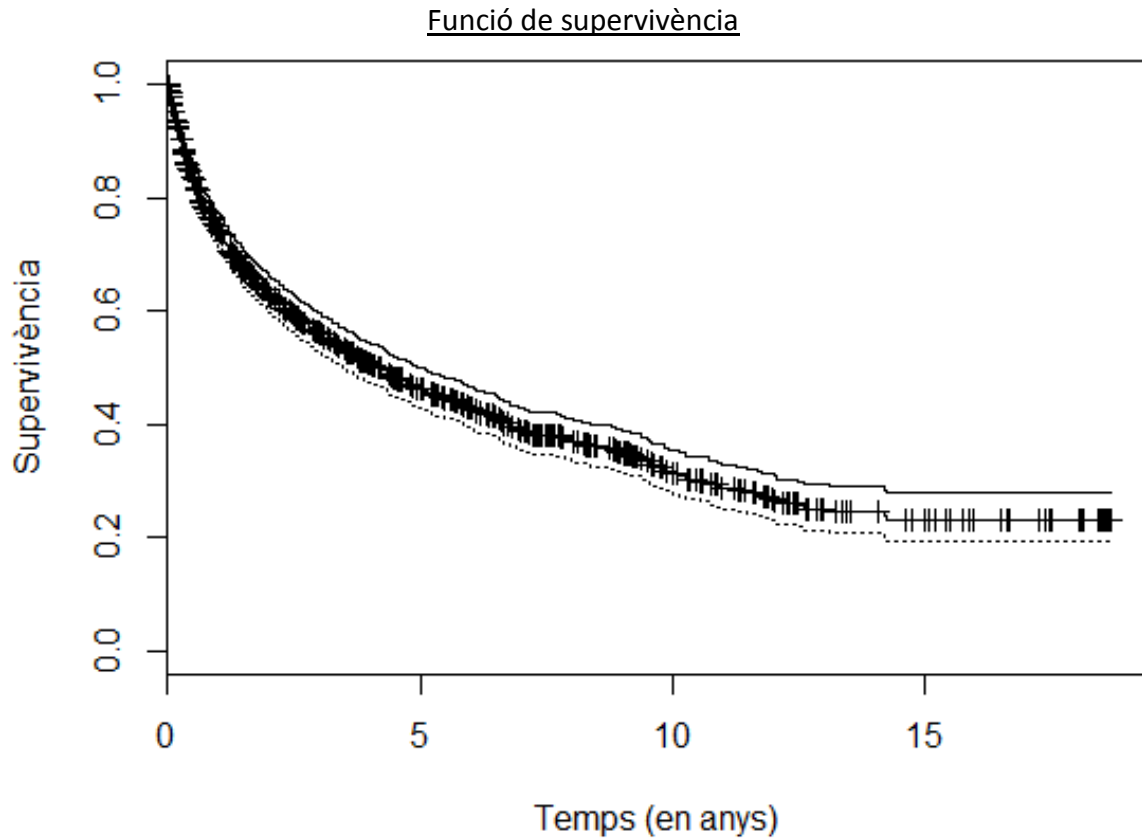
$$h(t, \mathbf{X}) = h_0(t_i) \cdot \exp\left(\sum_{j=1}^p \beta_j x_j\right)$$

On  $h_0(t)$  és la funció de risc, que dóna el risc instantani de patir l'esdeveniment d'interès en un moment  $t$  donat que no ha ocorregut abans,  $t$  és el temps fins l'aparició de l'esdeveniment i la resta de variables és igual als altres models.

Aquesta expressió es divideix en dues parts, la funció de risc basal  $h_0(t_i)$ , que és l'única part del model que depèn del temps per tant la no paramètrica, i  $\exp\left(\sum_{j=1}^p \beta_j x_j\right)$  que depèn del vector de covariables i no del temps, per tant la part paramètrica.

## 6.1. Anàlisi no paramètrica

Primer de tot estimem les funcions de supervivència a través de l'estimador de Kaplan-Meier, comencem presentant la supervivència de tota la mostra:



Gràfic 6.1 *Funció de supervivència del model nul.*

Podem observar que aproximadament el 50% dels individus estan en el tractament uns 4 anys, també que només un 30% (aproximadament) estan 12 o més anys en tractament.

Seguim amb l'informe presentant la funció de supervivència per diferents grups segons les categories de les variables:

Funció de supervivència per grups segons les variables Sexe i VIH que hem estimat:

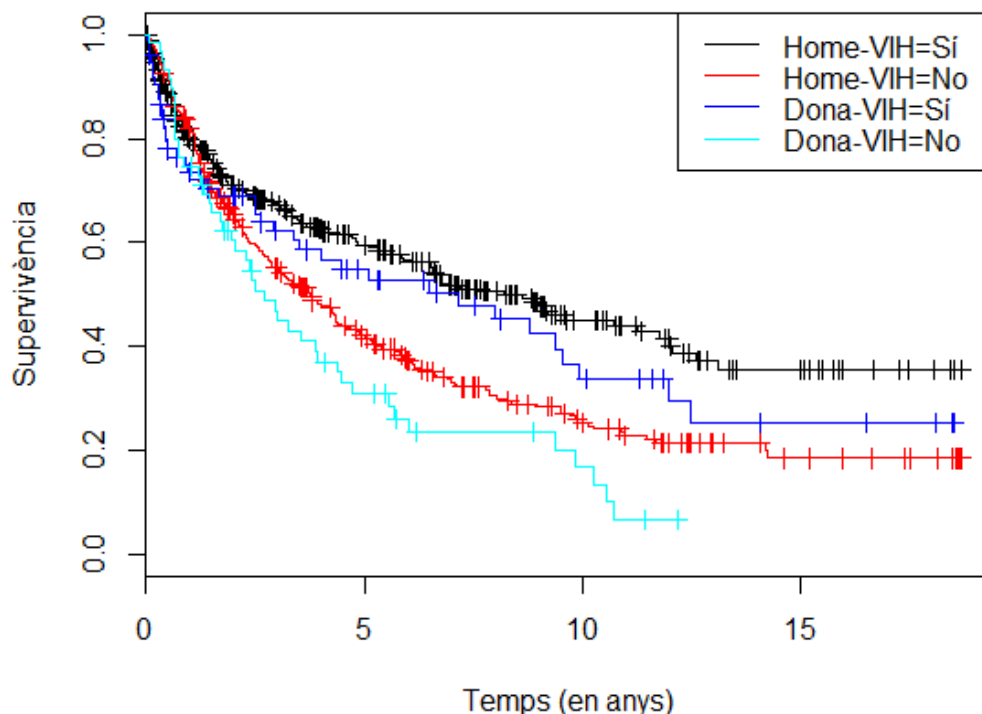
	Individus	Esdeveniments	Mediana	0,95LCL	0,9UCL
Home amb VIH	338	153	8,22	6,52	11,78
Home sense VIH	311	211	3,76	2,89	4,58
Dona amb VIH	74	40	7,17	3,37	11,96
Dones sense VIH	59	46	2,73	1,97	4,47

Taula 6.1 *Mediana de supervivència per grup de Sexe i VIH.*

Veiem que tot hi haver-hi més homes que no estan infectats per VIH (311), respecte la quantitat dels que sí que ho estan (338), més individus del primer tipus han deixat el tractament (211 front 153). Gràcies a les medianes d'aquests dos tipus d'individus sabem que la probabilitat estimada de que deixin el tractament abans del 3,76 i els 8,22 anys és d'un 50%, respectivament. En el cas de les dones passa el mateix, la quantitat que no tenen la malaltia (59) es inferior al que sí la tenen(74), però l'esdeveniment succeeix més vegades per les que no la tenen (40 front 46). Les medianes en aquest cas ens indiquen que tenim una probabilitat estimada del 50% de que les dones amb VIH deixin el tractament abans dels 7,17 anys, i abans del 2,73 anys per les dones sense la malaltia.

A continuació representem les 4 funcions de supervivència, estimades:

Gràfica de la funció de supervivència per Sexe i VIH



*Gràfic 6.2 Estimació de la funció de supervivència per les variables Sexe i VIH*

Veiem que a l'inici de la gràfica les dones amb VIH tenen una funció de supervivència per sota de les altres tres. A mesura que avancen els anys cada cop les funcions es separen més, al final de la gràfica veiem que els individus que presenten una major supervivència són els homes que sí que tenen la malaltia. Recordem que en aquesta anàlisi, supervivència vol dir mantenir-se en el TMM. Sembla que podem descartar la igualtat entre funcions.

També hem fet una estimació de la funció de supervivència segons si els individus tenen o no antecedents penals:

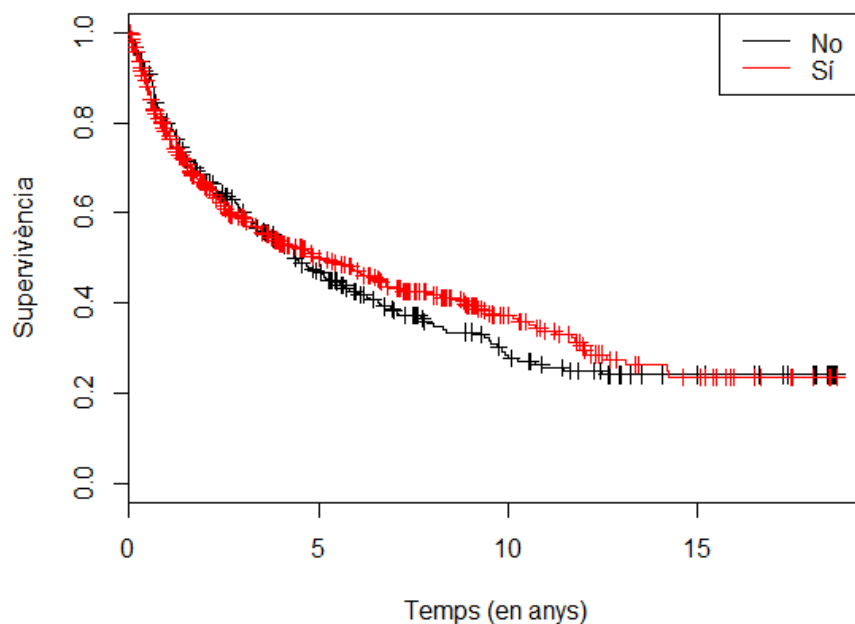
	Individus	Esdeveniments	Mediana	0,95LCL	0,9UCL
Sí	294	187	4,39	3,57	5,73
No	469	257	5,12	3,74	6,78

*Taula 6.2 Estimació de la funció de supervivència per grup d'antecedents penals.*

Veiem que hi ha més individus que no tenen antecedents penals(469), dels quals 257 han presentat l'esdeveniment. Per contra tenim 294 individus amb antecedents penals dels quals la majoria també han presentat l'esdeveniment. Però tot i haver-hi diferències en les proporcionalitats, el temps de supervivència per als dos és molt semblant.

Per acabar de decidir si deduïm o no diferències representem les dues funcions de supervivència estimades seguidament:

Gràfica de la funció de supervivència segons Antecedents Penals



*Gràfic 6.3 Estimació de la funció de supervivència per els dos grups d'antecedents penals.*

Les dues corbes són bastant paral·leles, comportant-se de la mateixa forma, per tant, no inclourem aquesta variable en el model de Cox que calcularem més endavant, i deduïm que no hi ha diferències entre les dues corbes de supervivència.

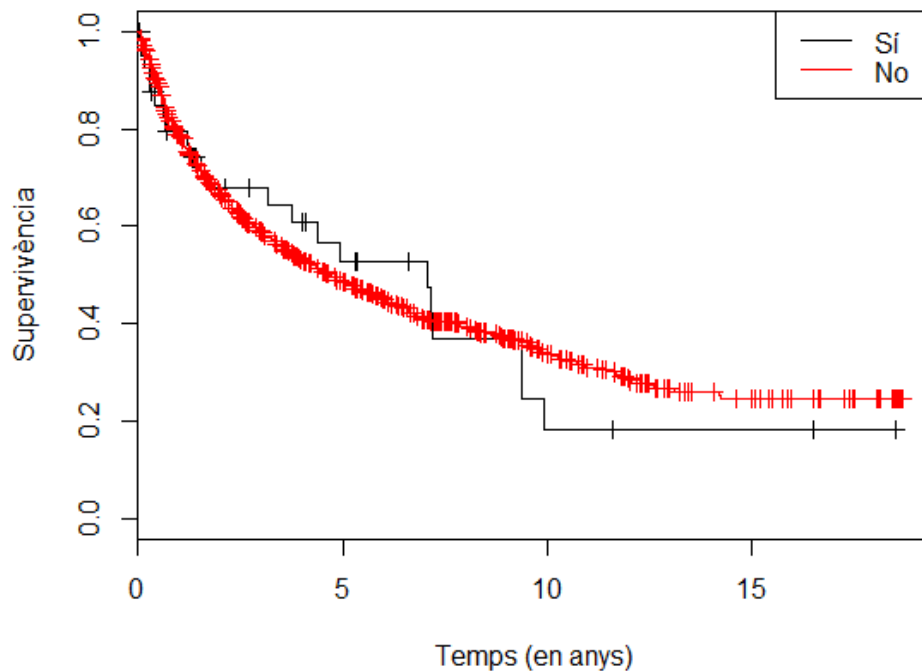
Per últim presentem unes estimacions de la funció de supervivència segons si els pacients del tractament han tingut o no relacions a canvi de diners o drogues:

	Individus	Esdeveniments	Mediana	0,95LCL	0,9UCL
Sí	42	22	7,08	3,76	9,93
No	707	415	4,70	3,93	5,83

*Taula 6.3 Estimació de la funció de supervivència per grup segons si han tingut o no relacions a canvi de diners o drogues.*

Veiem que tenim molts menys individus que sí han tingut relacions a canvi de diners o drogues (42), que els que no ho han fet (707), aquest fet ens podria donar problemes si incloguéssim aquesta variable en algun model de Cox. Les proporcionalitats dels individus que presenten l'esdeveniment és molt semblant per als dos grups.

Gràfica de la funció de supervivència segons si han tingut o no relacions a canvi de diners o drogues



*Gràfic 6. 4 Estimació de la funció de supervivència per als dos grups de la variable Relació per Drogues o Diners.*

Observem que les corbes constantment es creuen i finalment decidim que no tenim respostes suficients per al grup dels que presenten l'esdeveniment com per a tenir en compte la variable.

Seguidament procedim a fer la comparació de dos o més funcions de supervivència, per fer-ho realitzarem la prova del Logrank-test i el test de Peto.



Els resultats es presenten en forma de taula i després es comenten:

Variable	p-valor	
	Logrank	cPeto
Sexe	0,0831	0,147
Consum cocaïna	0,00157	0,001
Situació laboral	0,00148	0,00633
VIH	4,86e-06	0,000482
Antecedents penals	0,539	0,872

*Taula 6.4 Proves de diferència en la supervivència.*

Per un nivell de significació del 0,05 obtenim que les variables Consum de Cocaïna, Situació Laboral i infecció d'VIH són significatives, tant per el test de Peto com per el test del Logrank, per tant amb aquest nivell de significació podem decidir que hi ha diferències en les funcions de supervivència segons els grups de cada variable.

La variable Antecedents penals dóna altament no significativa i per tant, com ja havíem comentat anteriorment, no podem afirmar que hi ha diferències entre les funcions de supervivència entre els individus que sí que tenen antecedents penals i els que no en tenen.

La variable Sexe no és significativa amb un nivell del 0,05, per cap dels dos test, però amb un nivell de significació de 0,1 sí que ho és en el test del Logrank, és per això que si incloem o no aquesta variable en el model ho decidirem més endavant.

A continuació es representen les funcions de supervivència per als dos casos, dins de la significació, les més extremes:

## 6.2. Anàlisi semi-paramètric

Per realitzar l'anàlisi semi-paramètrica ho farem a través d'un Model de Cox. El model de Cox pressuposa proporcionalitat de riscos. És realment important comprovar que es dona aquesta proporcionalitat, perquè si no el model no seria vàlid.

El model de Cox és ideal per a quan no sabem quina distribució segueix el temps de supervivència, ja que els resultats ajustats tendiran a aproximar-se als del model paramètric correcte. Aquest model ens permet realitzar una anàlisi sense haver d'establir uns paràmetres que es podria donar el cas que fossin incorrectes, per tant si no estem segurs de quina distribució segueixen les nostres dades el millor seria realitzar un model de Cox. Això sí, si coneixem el model al que s'ajusten les dades el millor seria realitzar un anàlisi paramètrica.

A diferència de l'apartat de regressió logística aquí no hem anat afegint variables de forma additiva, aquí hem posat directament les variables que, amb tot el realitzat fins al moment,

ens semblaven que serien significatives i interessants. Totes les variables que hem afegit al model han sigut fixes (que no poden variar), per això hem decidit descartar la variable dosi inicial.

A continuació presentem el primer model que ha semblat interessant:

Variable	$\hat{\beta}$	$e^{\hat{\beta}}$	Risc	s.e. ( $\hat{\beta}$ )	Z	p-valor	CI (95%)
Consum cocaïna: Esporàdicament	0,103	1,1089	↑	0,1145	0,903	0,366413	0,8860 1,3880
Consum cocaïna: Diàriament	0,517	1,6771	↑	0,1542	3,352	0,000801	1,2396 2,2690
Situació laboral: Sí, ocasionalment	-0,058	0,9436	↓	0,1257	-0,462	0,643805	0,7376 1,2071
Situació laboral: Mai	-0,323	0,7241	↓	0,2698	-1,196	0,231541	0,4267 1,2288
Situació laboral: Atur retribuït	-0,064	0,9382	↓	0,2125	-0,300	0,764088	0,6186 1,4230
Situació laboral: Atur no retribuït	-0,291	0,7475	↓	0,1933	-1,506	0,132114	0,5118 1,0917
Situació laboral: Incapacitat	-0,638	0,5284	↓	0,2208	-2,889	0,003864	0,3428 0,8145
VIH: No	0,439	1,5506	↑	0,1078	4,069	4,72e-05	1,2553 1,9155
Sexe: Dona	0,272	1,3130	↑	0,1262	2,157	0,030980	1,0252 1,6816

Codi de significació: 0 0.001 0.01 0.05 0.1 1

Likelihood ratio test	52,63	on	9	df,	p=3,436e-08
Wald test	50,47	on	9	df,	p=8,786e-08
Score (logrank) test	51,71	on	9	df,	p=5,131e-08

Taula 6.5. Resultats model de Cox.

Categories de referència:

- Consum de cocaïna: No.
- Situació laboral: Sí, fixe.
- VIH: Sí.
- Sexe: Home.

L'expressió de l'estimació del nostre model és la següent:

$$h(t, \mathbf{X}) = h_0(t) \cdot \exp(\beta_1 \text{Consum cocaïna esporàdicament} + \beta_2 \text{Consum cocaïna diàriament} + \beta_3 \text{Situació Laboral: Sí Ocasionalment} + \beta_4 \text{Situació Laboral: Mai} + \beta_5 \text{Situació Laboral: Atur retribuït} + \beta_6 \text{Situació Laboral: Atur no retribuït} + \beta_7 \text{Situació Laboral: Incapacitat} + \beta_8 \text{VIH: No} + \beta_9 \text{Sexe: Dona})$$

#### Validació del model

Abans de fer la interpretació de les variables anem a comprovar si es compleix la hipòtesi de proporcionalitat de riscos. Amb una funció que prova aquesta proporcionalitat mitjançant la creació d'interaccions amb el temps mitjà en anys.

Variable	Rho	Chisq	P
Consum cocaïna: Esporàdicament	-0,0103	0,0445	0,8329
Consum cocaïna: Diàriament	0,0114	0,0536	0,8169
Situació laboral: Sí, ocasionalment	-0,1195	6,0900	0,0136
Situació laboral: Mai	-0,0893	3,3763	0,0661
Situació laboral: Atur retribuït	-0,0959	3,9050	0,0481
Situació laboral: Atur no retribuït	-0,0655	1,7969	0,1801
Situació laboral: Incapacitat	-0,0299	0,3878	0,5335
VIH: No	0,0787	2,7083	0,0998
Sexe: Dona	0,0587	1,4274	0,2322

Taula 6.6 Taula de Hipotesis de proporcionalitat de riscos.

Veiem que amb una significació de 0,05 dues de les categories de situació laboral no compleixen la hipòtesi de proporcionalitat de riscos. A més a més amb una significació de 0,1 la variable VIH tampoc la compleix. Anem a fer la representació gràfica dels residus per donar suport a la idea argumentada.

En la gràfica 6.8 efectivament veiem que la categoria "Sí, Ocasionalment" de la variable Situació Laboral no compleix la hipòtesi de proporcionalitat de riscos, ja que els residus presenten una tendència negativa, per tant depèn del temps.

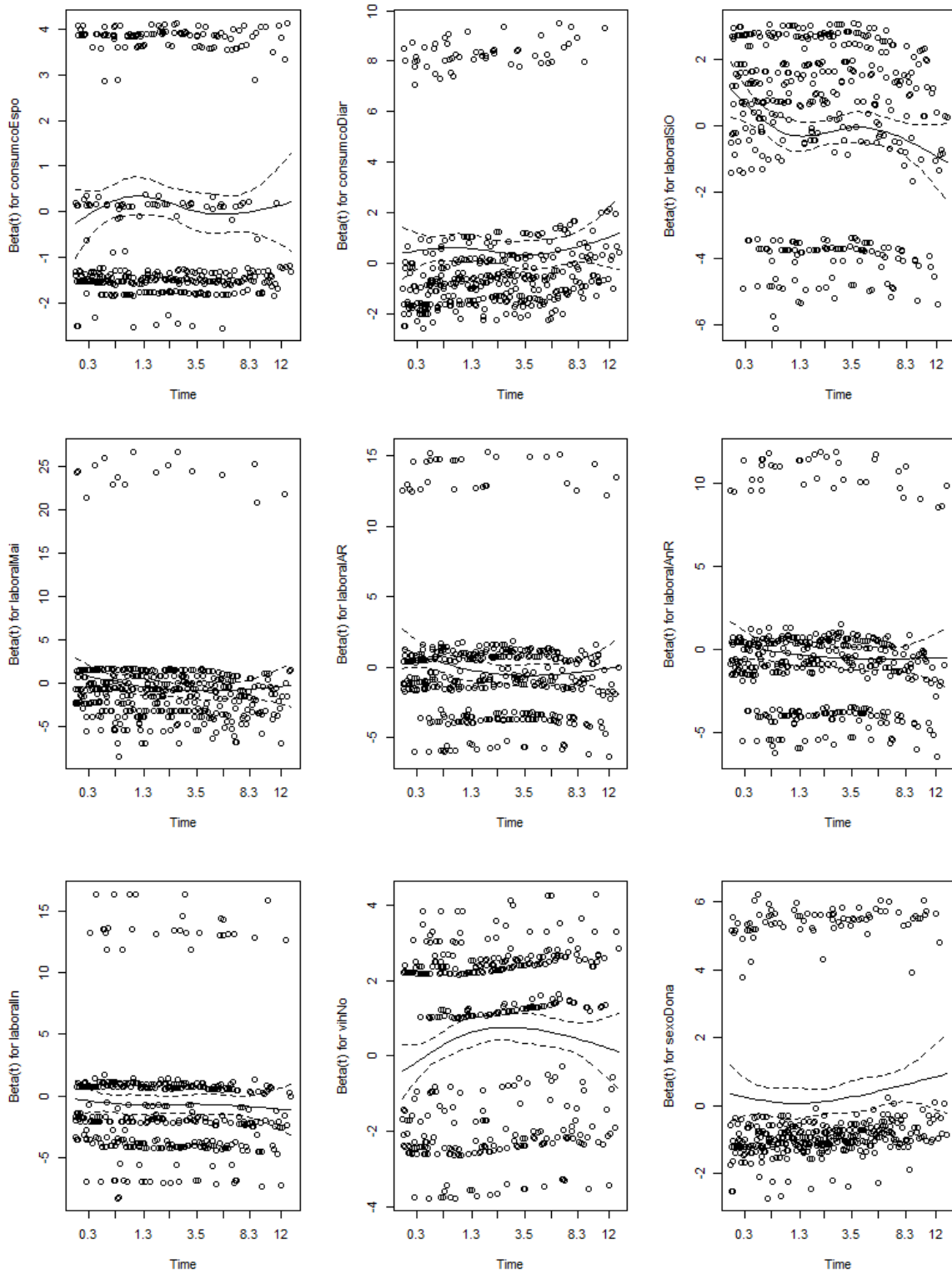
Els residus de la variable VIH, com ja esperàvem, presenten també una forma inusual però menys marcada que la variable anomenada anteriorment.

També comentem que els residus de la resta de variables és comporten de manera horitzontal com era d'esperar.

Per tant a continuació se'ns presenta un problema, què fer amb la variable Situació laboral.

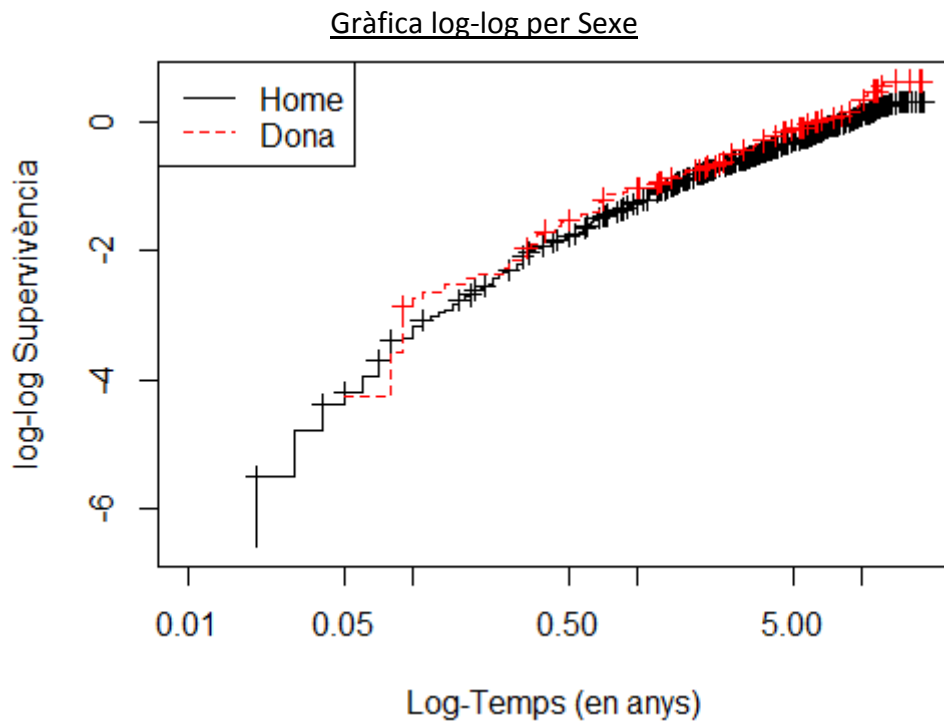
Abans de prendre qualsevol decisió sobre les variables del model, representarem les seves gràfiques log-log per a veure si es comporten paral·lelament o és creuen.

## Gràfic de residus de Schoenfeld



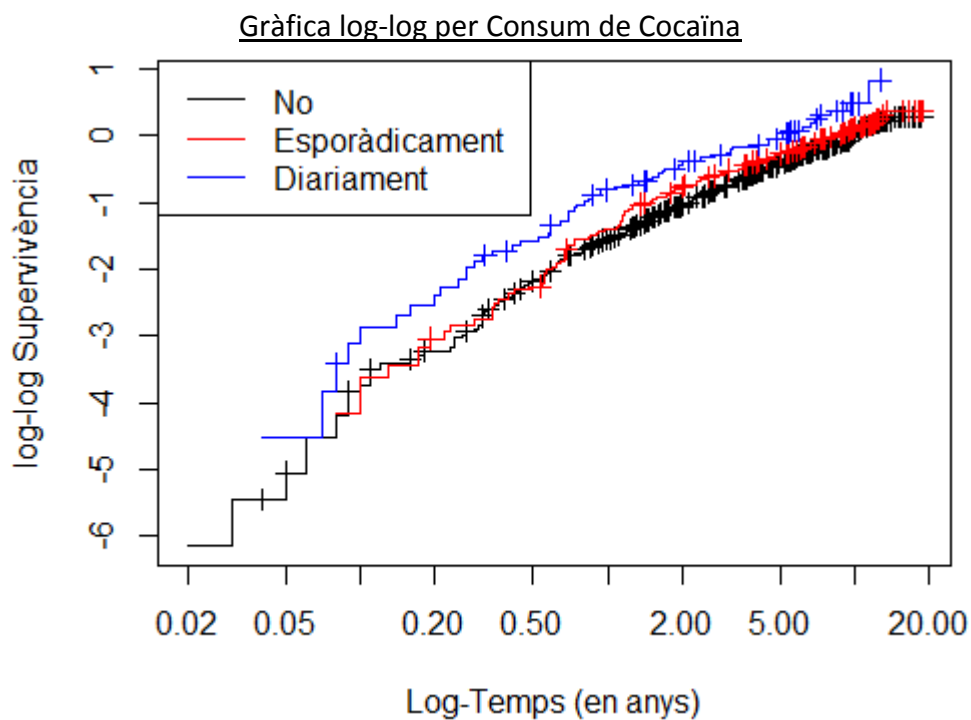
Gràfic 6.5 Gràfic de residus de Schoenfeld pel model de Cox.

En les gràfiques de log-log busquem que les corbes siguin mes o menys paral·leles:



Gràfic 6.6 Gràfica log-log per la variable Sexe.

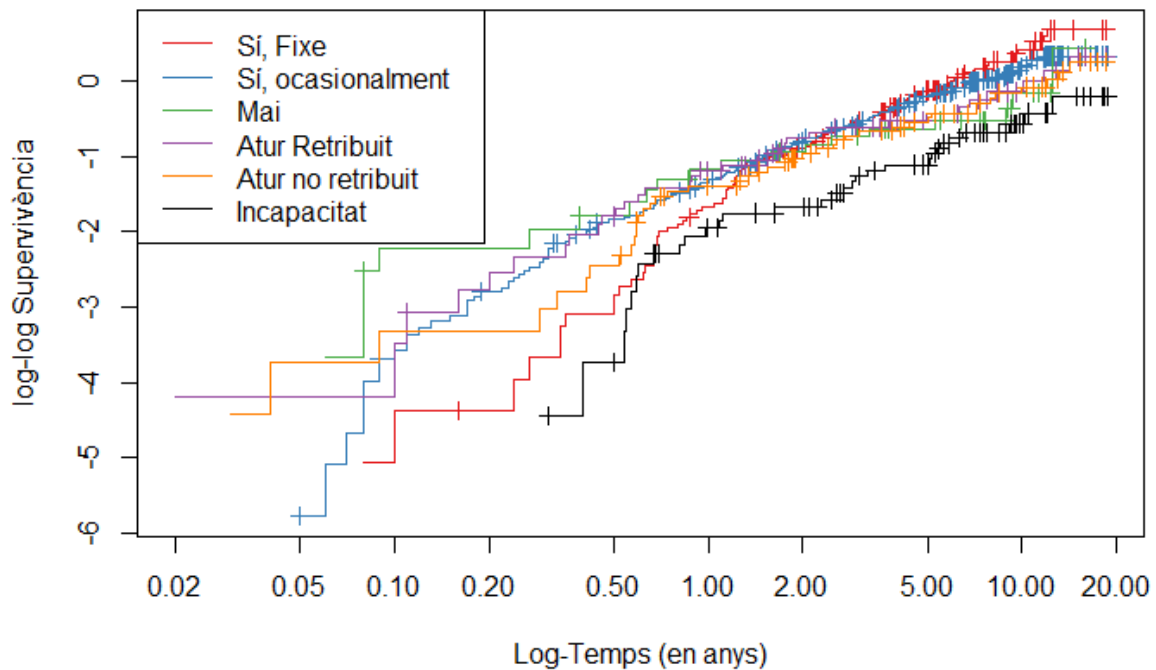
Efectivament són bastant paral·leles i per tant segueixen la hipòtesi de proporcionalitat de riscos.



Gràfic 6.7 Gràfica log-log per la variable Consum de cocaïna.

Realment són paral·leles i no es creuen tal i com ens esperàvem.

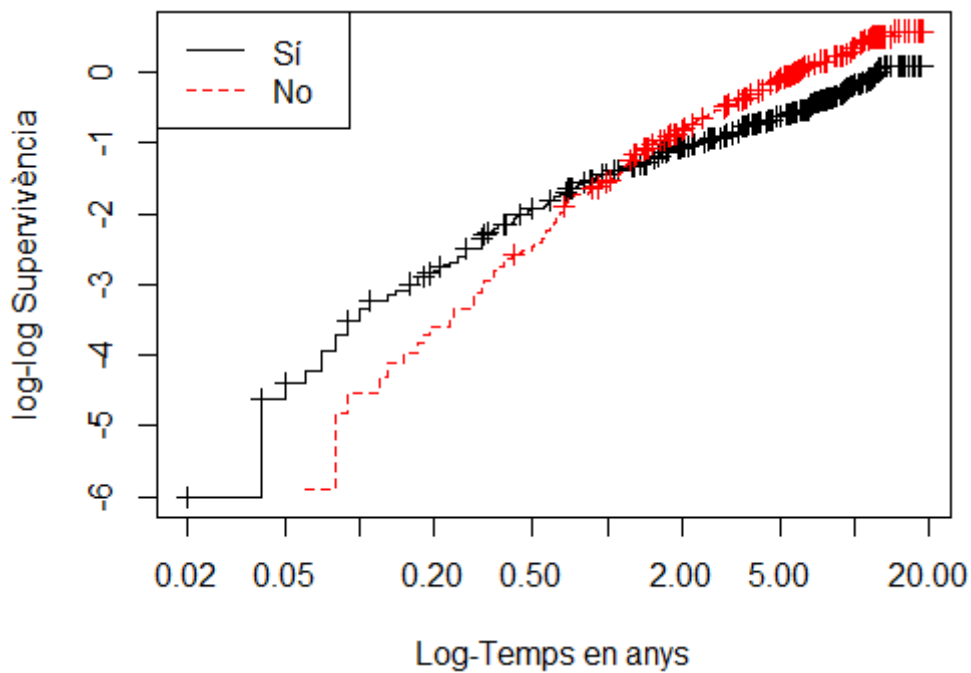
Gràfica log-log per Situació laboral



*Gràfic 6.8 Gràfica log-log per la variable Situació Laboral.*

Com veiem no és comporten de manera homogènia, i la corba corresponent a la categoria “Sí, ocasionalment” és creua amb quasi tota la resta de corbes. Com ja sabíem aquesta variable no compleix la hipòtesi de proporcionalitat de riscos.

Gràfica log-log per infecció d'VIH



*Gràfic 6.9 Gràfica log-log per la variable infecció d'VIH*

Tot i ser bastant paral·leles les corbes arriba un punt en el que es creuen, ho haurem de tenir en compte.

Com ja hem comentat anteriorment desitgem treure la variable Situació Laboral del model, però quan ho fem ens trobem que la significació del model baixa molt i la variable VIH deixa de complir la hipòtesi de proporcionalitat de riscos, amb una significació de 0,05, és per això que introduïrem la variable situació Laboral en forma d'estrat, i aquest serà el nostre model definitiu.

### Model definitiu

Finalment tenim un Model de Cox estratificat, i el presentem a continuació:

Variable	$\hat{\beta}$	$e^{\hat{\beta}}$	Risc	s.e. ( $\hat{\beta}$ )	z	p-valor	CI (95%)
Consum cocaïna: Esporàdicament	0,1084	1,1145	↑	0,1152	0,941	0,346497	0,8893 1,397
Consum cocaïna: Diàriament	0,5239	1,6885	↑	0,1549	3,382	0,000719	1,2464 2,287
VIH: No	0,4358	1,5463	↑	0,1082	4,028	5,62e-05	1,2508 1,912
Sexe: Dona	0,2608	1,2980	↑	0,1265	2,062	0,039161	1,0130 1,663

Codi de significació: 0 0.001 0.01 0.05 0.1 1

Likelihood ratio test	30,23	on	4	df,	p=4,385e-06
Wald test	31,02	on	4	df,	p=3,032e-06
Score (logrank) test	31,43	on	4	df,	p=2,496e-06

Taula 6.7 Resultats model del Cox Estratificat definitiu.

Categories de referència:

- Consum de cocaïna: No.
- VIH: Sí.
- Sexe: Dona.

L'expressió del nostre model és la següent:

$$h_j(t|\mathbf{Z}(t)) = h_{0j}(t) \cdot \exp(\beta_1 \text{Consum cocaïna esporàdicament} + \beta_2 \text{Consum cocaïna diàriament} + \beta_3 \text{VIH: No} + \beta_4 \text{Sexe: Dona}) = h_{0j}(t) \cdot \exp(0,1084 \text{Consum cocaïna esporàdicament} + 0,5239 \text{Consum cocaïna diàriament} + 0,40358 \text{VIH: No} + 0,2608 \text{Sexe: Dona})$$

Veiem que l'expressió es presenta diferent a la del model anterior, però això succeeix perquè en aquest cas tenim un model estratificat i el conjunt de les dades observades seria el següent:

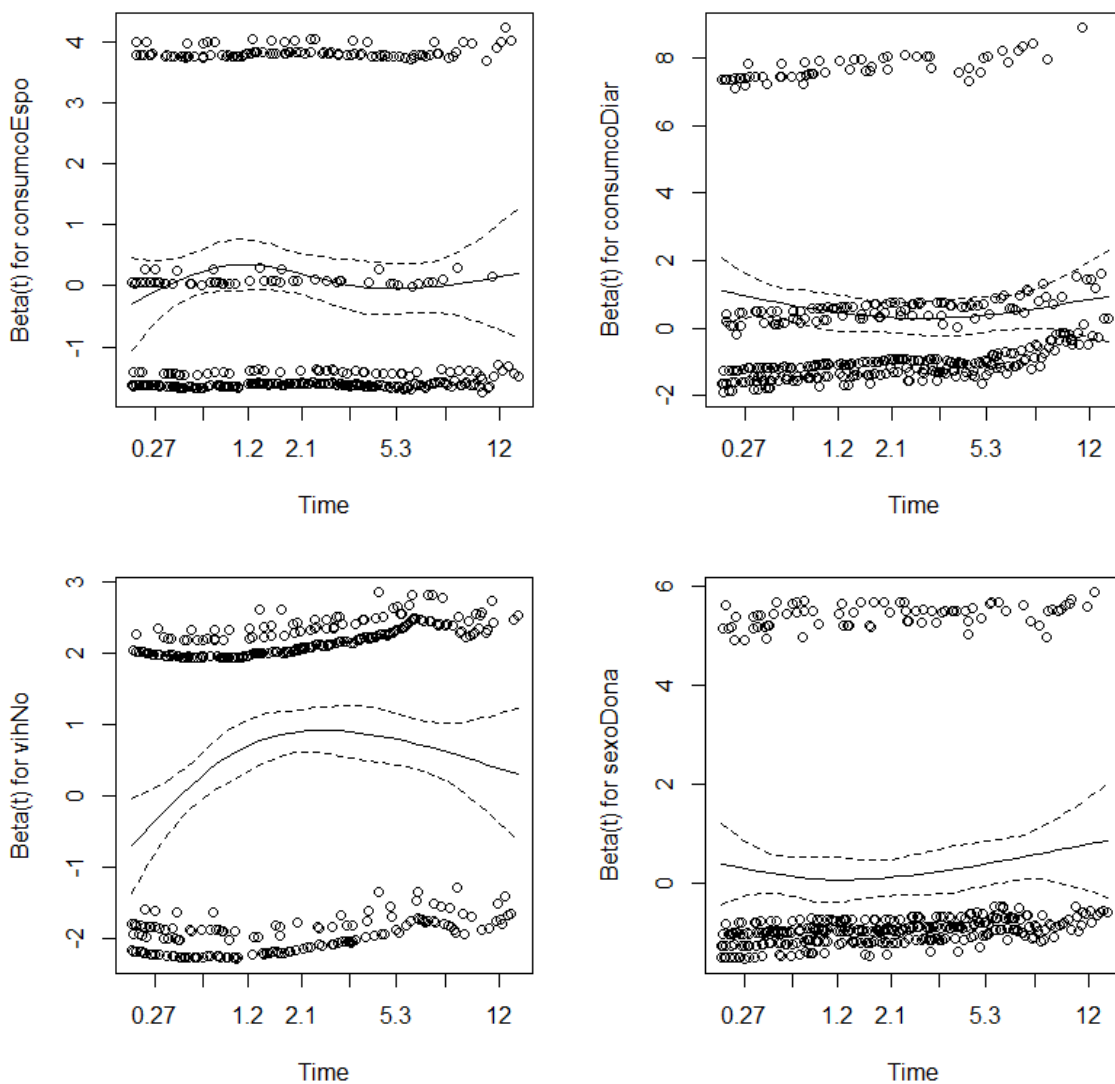
$$D = \{t_i, d_i, e_i, [Z_i(t), 0 \leq t \leq t_i], i = 1, 2 \dots n\}$$

On d i t correspon amb el model normal i e es l'estrat al que pertany cada individu.

Si realitzem un anàlisi ANOVA per veure la significació de les variables al incorporar-les ens donen totes significatives.

### Validació del model

En aquest apartat realitzarem la validació del nostre model, tot comprovant la hipòtesi de proporcionalitat de riscos, per començar farem la gràfica dels residus de Schoenfeld:



Gràfic 6.10 Gràfic de residus de Schoenfeld pel model de Cox Estratificat.

Sembla que, exceptuant la variable VIH, la resta sí que compleixen raonablement bé la hipòtesi de proporcionalitat de riscos.



Per saber si realment la segueixen o no realitzarem el test dels residus:

Variable	Rho	Chisq	P
Consum cocaïna: Esporàdicament	-0,0214	0,188	0,665
Consum cocaïna: Diàriament	0,0090	0,033	0,856
VIH: No	0,0873	3,307	0,069
Sexe: Dona	0,0537	1,176	0,278

Taula 6.8 Taula de Hipotesi de proporcionalitat de riscos.

La variable VIH : Utilitzant un nivell de significació de 0,1 no podem suposar que la variable compleix la hipòtesi de proporcionalitat de riscos. És per això que estudiem la possibilitat de treure-la del model, fent un model no. Quan fem el comentat anteriorment, veiem que la significació del model baixa bastant i es per això que ens quedarem amb l'anterior, el model estratificat que inclou la variable VIH.

#### *Interpretació dels paràmetres del model*

A continuació presentem la interpretació dels paràmetres de Cox obtinguts del model definitiu (Taula 6.7):

- Variable consum cocaïna: Esporàdicament. La beta obtinguda per aquesta variable és 0,1084, això implica que el risc de deixar el tractament per a una persona que consumeix esporàdicament cocaïna front d'una que no en consumeix, amb els mateixos valors en la resta de variables (fixos), és 1,1145 vegades major. Veiem que el p-valor, d'aquesta és no significatiu, i per tant, l'interval de confiança calculat recull l'1. Definitivament no podem deduir que consumir esporàdicament influeixi en el temps fins que deixen el tractament.
- Variable consum de cocaïna: Diàriament. Aquesta variable sí que és significativa (p-valor= 0,000719) i per tant influeix en el temps que els individus estan en tractament. Veiem que la beta és superior a 0 i per tant augmenta el risc dels individus perquè deixin el tractament, més concretament l'augmenta 1,6885 vegades, respecte els individus que no consumeixen cocaïna amb les mateixes característiques.
- Variable VIH: No. El signe positiu obtingut per a la beta d'aquesta variable ens indica que els pacients que no tenen VIH, amb respecte els que si que en tenen i mantenint constant la resta de variables, tenen més probabilitats de deixar el tractament. En concret els individus sense VIH tenen un risc de deixar el tractament 1,5463 vegades superior als que pateixen la malaltia. Per aquesta variable també obtenim un p-valor significatiu i per tant deduïm que influeix en el temps fins que deixen el tractament.

- Variable Sexe: Dona. Aquesta variable també obté un coeficient beta amb signe positiu i per tant també serà un factor que influirà negativament a la nostre variable resposta. De fet és de 1,298 vegades el risc de deixar el tractament per als homes, que és la categoria de referència, sempre i quan la resta de variables sigui igual per als dos. Aquesta variable també es significativa i per tant concloem que influeix en el temps fins que abandonen el tractament.

## CONCLUSIONS

L'objectiu del present treball era estudiar els factors associats a les interrupcions del tractament amb metadona i explorar el temps fins la interrupció del procés i els components que influeixen en el tall de seguiment.

Per poder estudiar els factors associats a la interrupció del tractament hem realitzat diferents models de regressió logística i finalment ens hem decantat per un i hem obtingut tres variables que influeixen en la nostra variable resposta. L'edat dels individus a l'inici del tractament, factor que influeix de manera positiva, és a dir, a mesura que augmenta fa que disminueixi el risc de deixar el tractament, la variable antecedents penals, que el fet de tenir-ne augmenta el risc de deixar el tractament, i la variable consum per via venosa, que igual que amb la variable anterior el fet de que aquesta sigui positiva incrementa el risc de deixar el tractament.

Per altra banda amb els factors associats al temps fins la interrupció del tractament, hem trobat tres variables estadísticament significatives. Hem detectat temps en mitjana més curts en persones que consumeixen cocaïna a diari, en persones no infectades pel VIH i en dones.

Per tant hem trobat dos models definitius que compleixen amb els objectius marcats al inici de l'estudi.

## LIMITACIONS I FUTURES LÍNIES D'INVESTIGACIÓ

Cada limitació del treball és una futura línia de discussió, per això aquests dos apartats del treball apareixen junts.

Per la part de l'estudi de regressió logística, ens ha semblat interessant la idea de crear noves variables per a incorporar-les al model, entre d'altres una variable on es representés la dosi en el moment de deixar el tractament, o una variable que expliqués si la dosi actual o la de final de tractament és inferior a l'inicial...

Per la part de l'anàlisi semi-paramètrica ens ha semblat interessant l'idea de tractar als individus morts com a no censurats. També ens ha semblat interessant la idea de poder incorporar una variable canviant en el temps, com és la variable dosi prescrita, ja que creiem que aquesta influeix d'una manera notable en el manteniment del tractament.

També es podrien incorporar variables menys objectives, com la satisfacció dels individus amb el tractament, si els és difícil no consumir heroïna... Per suposat també creiem que seria molt interessant incorporar una variable, per a tots aquells individus que deixen el tractament, que fos el motiu pel que el deixen, perquè tot i que la majoria de persones deixar el tractament implica tornar a consumir heroïna, per una petita part d'individus pot ser que és doni aquesta baixa perquè ja no tenen la necessitat de consumir opiàcis.

Per suposat, es podrien realitzar models d'efectes aleatoris, que tinguin en compte la no independències entre els episodis.

Per finalitzar comentem que també es poden obrir noves línies d'investigació amb anàlisis paramètriques.

## BIBLIOGRAFIA

1. Barrio G, de la Fuente L, Camí J. *El consumo de drogas en España y su posición en el contexto europeo*. Med Clin (Barc)1993; 344-55.
2. David Carslaw and Karl Ropkins (2015). openair: Open-source tools for the analysis of air pollution data. R package version 1.1-5.
3. Juan José Fernández Miranda; César Pereiro Gómez (2007). *Guía para el tratamiento de la dependencia de opiáceos*. Guía clínica. Junta Directiva de Socidrogalcohol.
4. Lluís de la Fuente, M Teresa Brugal, Antonio Domingo-Salvany, María J Bravo, Montserrat Neira-León y Gregorio Barrio. *Más de treinta años de drogas ilegales en España: una amarga historia con algunos consejos para el futuro*. Colaboración especial en Rev. Esp Salud Pública 2006;80:505-520.
5. M<sup>a</sup> Carmen Vidal Casero. *Deshabitación de opiáceos(i)*; Toxicología, de la Vol21 Núm 1 Enero 2002.
6. Observatorio Europeo de las Drogas y las Toxicomanías. *El problema de la drogodependencia en Europa*. Informe anual 2011. ISSN 1609-607X.
7. Terry M. Therneau and Patricia M. Grambsch (2000). *Modeling Survival Data: Extending the Cox Model*. Springer, New York. ISBN 0-387-98784-3.
8. Venables, W. N. & Ripley, B. D. (2002) *Modern Applied Statistics with S*. Fourth Edition. Springer, New York. ISBN 0-387-95457-0
9. Viquipèdia. Opiaci. Última modificació: 8 set 2014. [Consulta: 9 de juny de 2015] Disponible a: <http://ca.wikipedia.org/wiki/Opiaci>.

## ANNEX

### i. Models no presentats al informe

#### Models de regressió logística

##### a) Model m.

```
Call:
glm(formula = pacienta ~ 1, family = binomial, data = dades)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.8547  0.6282  0.6282  0.6282  0.6282

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.5226     0.1038   14.66  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 593.14  on 630  degrees of freedom
Residual deviance: 593.14  on 630  degrees of freedom
AIC: 595.14

Number of Fisher Scoring iterations: 4
```

##### b) Model m1.

```
Call:
glm(formula = pacienta ~ sexo, family = binomial, data = dades)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.8608  0.6243  0.6243  0.6243  0.6454

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.53651     0.11556   13.30  <2e-16 ***
sexoDona    -0.07368     0.26331   -0.28    0.78
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 593.14  on 630  degrees of freedom
Residual deviance: 593.06  on 629  degrees of freedom
AIC: 597.06

Number of Fisher Scoring iterations: 4
```

c) *Model m1b.*

Call:  
glm(formula = pacienta ~ consumco, family = binomial, data = dades)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.0759	0.4965	0.5657	0.6929	0.6929

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	1.3045	0.1348	9.680	<2e-16 ***
consumcoEspo	0.4468	0.2721	1.642	0.1006
consumcoDiar	0.7270	0.3994	1.820	0.0688 .

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 514.12 on 538 degrees of freedom  
Residual deviance: 508.61 on 536 degrees of freedom  
(92 observations deleted due to missingness)  
AIC: 514.61

Number of Fisher Scoring iterations: 4

d) *Model m1c.*

Call:  
glm(formula = pacienta ~ edadinic, family = binomial, data = dades)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.0001	0.5614	0.5962	0.6327	0.8891

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	2.37675	0.42389	5.607	2.06e-08 ***
edadinic	-0.04349	0.02045	-2.127	0.0334 *

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 591.95 on 627 degrees of freedom  
Residual deviance: 587.56 on 626 degrees of freedom  
(3 observations deleted due to missingness)  
AIC: 591.56

Number of Fisher Scoring iterations: 4

*e) Model m2.*

```
Call:
glm(formula = paciente ~ edadinic + edat, family = binomial,
    data = dades)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.3441	0.3627	0.4985	0.6373	1.5696

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	4.83251	0.60036	8.049	8.32e-16 ***
edadinic	0.01991	0.02350	0.847	0.397
edat	-0.10706	0.01624	-6.592	4.33e-11 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 591.95 on 627 degrees of freedom  
Residual deviance: 540.11 on 625 degrees of freedom  
(3 observations deleted due to missingness)  
AIC: 546.11

Number of Fisher Scoring iterations: 4

*f) Model m2b.*

```
Call:
glm(formula = paciente ~ edat + nombre.d, family = binomial,
    data = dades)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.4066	0.3240	0.4618	0.6118	1.2798

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	5.5109352	1.2248465	4.499	6.82e-06 ***
edat	-0.1151631	0.0310447	-3.710	0.000208 ***
nombre.d	0.0004356	0.0007019	0.621	0.534916

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 181.28 on 205 degrees of freedom  
Residual deviance: 164.13 on 203 degrees of freedom  
(425 observations deleted due to missingness)  
AIC: 170.13

Number of Fisher Scoring iterations: 5



*g) Model m2c.*

Call:  
glm(formula = pacienta ~ edat + vih, family = binomial, data = dades)

Deviance Residuals:  
Min 1Q Median 3Q Max  
-2.2860 0.3716 0.5122 0.6571 1.5165

Coefficients:  
Estimate Std. Error z value Pr(>|z|)  
(Intercept) 5.16955 0.61385 8.422 < 2e-16 \*\*\*  
edat -0.10126 0.01616 -6.265 3.72e-10 \*\*\*  
vihNo -0.36939 0.23402 -1.578 0.114

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 538.40 on 554 degrees of freedom  
Residual deviance: 493.33 on 552 degrees of freedom  
(76 observations deleted due to missingness)  
AIC: 499.33

Number of Fisher Scoring iterations: 4

*h) Model m2d.*

Call:  
glm(formula = pacienta ~ edat + penales, family = binomial, data = dades)

Deviance Residuals:  
Min 1Q Median 3Q Max  
-2.3029 0.3436 0.4995 0.6526 1.6575

Coefficients:  
Estimate Std. Error z value Pr(>|z|)  
(Intercept) 5.00775 0.59345 8.438 < 2e-16 \*\*\*  
edat -0.11072 0.01664 -6.652 2.9e-11 \*\*\*  
penales 0.44935 0.23444 1.917 0.0553 .

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 532.55 on 547 degrees of freedom  
Residual deviance: 481.72 on 545 degrees of freedom  
(83 observations deleted due to missingness)  
AIC: 487.72

Number of Fisher Scoring iterations: 4

*i) Model m2e.*

```
Call:
glm(formula = paciente ~ edat + penales + laboral, family = binomial,
     data = dades)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.4690	0.3283	0.4780	0.6499	1.5657

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	4.83061	0.64740	7.462	8.55e-14	***
edat	-0.10572	0.01726	-6.124	9.14e-10	***
penalesi	0.52937	0.24487	2.162	0.0306	*
laboralsi	0.38807	0.30628	1.267	0.2051	
laboralmi	-0.48828	0.56127	-0.870	0.3843	
laboralar	-0.18200	0.45695	-0.398	0.6904	
laboralanr	-0.83053	0.38464	-2.159	0.0308	*
laboralin	-0.29016	0.41467	-0.700	0.4841	

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)  
Null deviance: 525.04 on 543 degrees of freedom  
Residual deviance: 465.09 on 536 degrees of freedom  
(87 observations deleted due to missingness)  
AIC: 481.09

Number of Fisher Scoring iterations: 5

*j) Model m2f.*

```
Call:
glm(formula = paciente ~ edat + penales + estudios, family = binomial,
     data = dades)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.3195	0.3201	0.5010	0.6571	1.6357

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	4.86263	1.03039	4.719	2.37e-06	***
edat	-0.10877	0.01701	-6.394	1.62e-10	***
penalesi	0.49883	0.24019	2.077	0.0378	*
estudios1Pri	0.08619	0.84386	0.102	0.9186	
estudios2FP	-0.08597	0.86020	-0.100	0.9204	
estudios3Bat	0.24298	1.07751	0.226	0.8216	
estudios4Sup	15.42011	821.34970	0.019	0.9850	
estudios5CT	-0.54384	1.65345	-0.329	0.7422	

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)  
Null deviance: 525.46 on 544 degrees of freedom  
Residual deviance: 473.38 on 537 degrees of freedom  
(86 observations deleted due to missingness)  
AIC: 489.38

Number of Fisher Scoring iterations: 15

*k) Model m2g*

```
Call:
glm(formula = pacienta ~ edat + penales + viavenos, family = binomial,
     data = dades)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.3689	0.3356	0.4791	0.6422	1.6409

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	5.06067	0.60172	8.410	< 2e-16 ***
edat	-0.10513	0.01739	-6.044	1.51e-09 ***
penalesSi	0.41606	0.23805	1.748	0.0805 .
viavenosNo	-0.42858	0.24247	-1.768	0.0771 .

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 524.21 on 541 degrees of freedom  
Residual deviance: 468.37 on 538 degrees of freedom  
(89 observations deleted due to missingness)  
AIC: 476.37

Number of Fisher Scoring iterations: 5

*l) Model m3*

```
Call:
glm(formula = pacienta ~ edat + penales + viavenos + viaveno1,
     family = binomial, data = dades)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.3806	0.3308	0.4840	0.6591	1.5851

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	5.03246	0.60287	8.347	< 2e-16 ***
edat	-0.10525	0.01742	-6.041	1.53e-09 ***
penalesSi	0.47711	0.25143	1.898	0.0578 .
viavenosNo	-0.52941	0.27540	-1.922	0.0546 .
viaveno1No	0.25927	0.34433	0.753	0.4515

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 524.21 on 541 degrees of freedom  
Residual deviance: 467.80 on 537 degrees of freedom  
(89 observations deleted due to missingness)  
AIC: 477.8

Number of Fisher Scoring iterations: 5

*m) Model m3a.*

Call:

```
glm(formula = paciente ~ edat + penales + viavenos + programa,  
     family = binomial, data = dades)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.4177	0.3294	0.4857	0.6492	1.6326

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	5.17962	0.62647	8.268	< 2e-16	***
edat	-0.10422	0.01747	-5.967	2.41e-09	***
penalesSi	0.39748	0.23865	1.666	0.0958	.
viavenosNo	-0.28773	0.28286	-1.017	0.3090	
programaNo	-0.29060	0.34792	-0.835	0.4036	

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 521.69 on 535 degrees of freedom  
Residual deviance: 465.75 on 531 degrees of freedom  
(95 observations deleted due to missingness)  
AIC: 475.75

Number of Fisher Scoring iterations: 5

*n) Model m3b.*

Call:

```
glm(formula = paciente ~ edat + penales + viavenos + relacdin,  
     family = binomial, data = dades)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.4484	0.3372	0.4804	0.6403	1.6273

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	5.54510	0.94407	5.874	4.26e-09	***
edat	-0.10412	0.01766	-5.897	3.71e-09	***
penalesSi	0.42069	0.24218	1.737	0.0824	.
viavenosNo	-0.40845	0.24704	-1.653	0.0983	.
relacdinNo	-0.52871	0.70600	-0.749	0.4539	

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 508.29 on 531 degrees of freedom  
Residual deviance: 455.11 on 527 degrees of freedom  
(99 observations deleted due to missingness)  
AIC: 465.11

Number of Fisher Scoring iterations: 5

*o) Model m3c.*

Call:

```
glm(formula = paciente ~ edat + penales + viavenos + preserva,  
     family = binomial, data = dades)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.3598	0.3252	0.4636	0.6613	1.5506

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	5.20510	0.65750	7.917	2.44e-15	***
edat	-0.10952	0.01816	-6.030	1.64e-09	***
penalesSi	0.47605	0.24570	1.938	0.0527	.
viavenosNo	-0.47226	0.25208	-1.873	0.0610	.
preserva75Pre	-0.87517	0.58047	-1.508	0.1316	
preserva50Pre	-0.28769	0.46422	-0.620	0.5354	
preserva25Pre	0.19037	0.82170	0.232	0.8168	
preserva0Pre	0.33639	0.30951	1.087	0.2771	
preservaNoRel	-0.11297	0.32371	-0.349	0.7271	

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 508.29 on 531 degrees of freedom  
Residual deviance: 449.92 on 523 degrees of freedom  
(99 observations deleted due to missingness)  
AIC: 467.92

Number of Fisher Scoring iterations: 5

*p) Model m4d.*

Call:

```
glm(formula = paciente ~ edat + viavenos + penales + gonococ,  
     family = binomial, data = dades)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.3907	0.3439	0.4863	0.6446	1.5824

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	4.24974	0.82613	5.144	2.69e-07	***
edat	-0.10155	0.01864	-5.448	5.10e-08	***
viavenosNo	-0.35003	0.25201	-1.389	0.1649	
penalesSi	0.52021	0.24685	2.107	0.0351	*
gonococNo	0.66893	0.46388	1.442	0.1493	

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 496.48 on 524 degrees of freedom  
Residual deviance: 444.76 on 520 degrees of freedom  
(106 observations deleted due to missingness)  
AIC: 454.76

Number of Fisher Scoring iterations: 5

## Models de Cox

### a) *Model coxmod2*

Call:

```
coxph(formula = Surv(sup, estat) ~ consumco + vih + sexo, data = episodis)
```

n= 720, number of events= 417  
(160 observations deleted due to missingness)

	coef	exp(coef)	se(coef)	z	Pr(> z )	
consumcoEspo	0.1046	1.1103	0.1140	0.917	0.358957	
consumcoDiar	0.5532	1.7388	0.1470	3.764	0.000167	***
vihNo	0.5176	1.6780	0.1004	5.153	2.57e-07	***
sexoDona	0.2613	1.2986	0.1242	2.103	0.035428	*

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
consumcoEspo	1.110	0.9007	0.8879	1.388
consumcoDiar	1.739	0.5751	1.3036	2.319
vihNo	1.678	0.5960	1.3781	2.043
sexoDona	1.299	0.7700	1.0180	1.657

Concordance= 0.585 (se = 0.015 )  
Rsquare= 0.056 (max possible= 0.999 )  
Likelihood ratio test= 41.59 on 4 df, p=2.025e-08  
Wald test = 42.47 on 4 df, p=1.331e-08  
Score (logrank) test = 43.17 on 4 df, p=9.547e-09

### b) *Model coxsinvih*

Call:

```
coxph(formula = Surv(sup, estat) ~ consumco + strata(laboral) +  
sexo, data = episodis)
```

n= 743, number of events= 435  
(137 observations deleted due to missingness)

	coef	exp(coef)	se(coef)	z	Pr(> z )	
consumcoEspo	0.1331	1.1424	0.1113	1.196	0.23156	
consumcoDiar	0.4961	1.6423	0.1509	3.287	0.00101	**
sexoDona	0.2678	1.3070	0.1222	2.191	0.02842	*

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
consumcoEspo	1.142	0.8754	0.9186	1.421
consumcoDiar	1.642	0.6089	1.2217	2.208
sexoDona	1.307	0.7651	1.0287	1.661

Concordance= 0.557 (se = 0.025 )  
Rsquare= 0.02 (max possible= 0.995 )  
Likelihood ratio test= 15.01 on 3 df, p=0.00181  
Wald test = 16.24 on 3 df, p=0.001011  
Score (logrank) test = 16.47 on 3 df, p=0.0009093

## ii. Llibreries utilitzades

- `library("ROCR")`
- `library(AER)`
- `library(car)`
- `library(descr)`
- `library(effects)`
- `library(Epi)`
- `library(epicalc)`
- `library(FactoMineR)`
- `library(fmsb)`
- `library(foreign)`
- `library(Hmisc)`
- `library(lattice)`
- `library(lmtest)`
- `library(lubridate)`
- `library(MASS)`
- `library(openair)`
- `library(plyr)`
- `library(RColorBrewer)`
- `library(survival)`
- `library(vcdExtra)`

### iii. Scripts d'R

- Preprocessament de les dades

```
## Lectura de les dades ----
```

```
#Primera base de dades
```

```
episodis <- spss.get("metadona_episodis.sav", datevars=c('datainic', 'datafina', 'fechanac',  
'data.mort2013', 'ult.viu2013'))
```

```
#extracció d'una petita mostra de la base de dades
```

```
head(episodis)
```

```
#resum presentació de les variables
```

```
str(episodis)
```

```
#Resum numèric de les variables
```

```
summary(episodis)
```

```
#Segona base de dades
```

```
dosis <- spss.get("dosis_prescrites.sav", datevars='data')
```

```
#extracció d'una petita mostra de la base de dades
```

```
head(dosis)
```

```
#resum presentació de les variables
```

```
str(dosis)
```

```
#resum numèric de les variables
```

```
summary(dosis)
```

```
#Preprocessing----
```

```
##Episodis
```

```
#recodificació de les variables
```

```
episodis$mort.2013<-factor(episodis$mort.2013,labels=c("No", "Sí"))
```

```
levels (episodis$consumco) <- c("No","Espo","Diar")
```

```
levels (episodis$preserva)<- c("100Pre","75Pre","50Pre","25Pre","0Pre","NoRel")
```

```
levels (episodis$laboral) <- c("SiF","SiO","Mai","AR","AnR","In")
```

```
levels (episodis$estudios) <- c("0An","1Pri","2FP","3Bat","4Sup","5CT")
```

```
summary(episodis)
```

```
episodis$edat<-trunc(episodis$edat)
```

```
##dosis
```

```
colnames(dosis) <- c("nummetad","date","dosiprec")
```

```
summary(dosis)
```

```
attach(dosis)
```

```
#individu 414 dades de l'any 1011 les canviem per 2011
```

```
selectByDate(dosis, year=1011)
```

```
year(dosis$date) <- ifelse(year(dosis$date) == 1011, 2011, year(dosis$date))
```

```
selectByDate(dosis, year=1011)
```



```

#Eliminem dades a partir de l'any 2014 i les de l'any 0199
dosis <- subset(dosis, year(date) < 2014 & year(date) != 0199)

summary(dosis)

#recodificació variable datafina
episodis$datafina <- with(episodis, as.Date(ifelse(mort.2013=="Sí" & datafina > data.mort2013,
data.mort2013, episodis$datafina), origin="1970-1-1"))

summary(episodis)

#creació de la variable estat
episodis$estat <- with(episodis, ifelse(mort.2013=="Sí" | is.na(datafina), 0, 1))

#declaració data de tall
a <- as.Date("2013-12-31")
#creacio variable sup
episodis$sup <- with(episodis, round(ifelse(estat == 1, datafina-datainic, ifelse(mort.2013=="Sí",
data.mort2013-datainic, a-datainic))/365.25, 2))

#Guardem les noves bases de dades
#save(dosis, file='dosis.RData')
#save(episodis, file='episodis.RData')

#Bases de dades a excel ----
write.csv(dosis, 'dosis.csv')
write.csv(episodis, 'episodis.csv')

```

- Descriptiva

```

load("dosis.RData")
load("episodis.RData")

#Episodis
summary(episodis)
attach(episodis)

#Calcul CV
100*sd(episodis$dosi.ini, na.rm = T)/mean(episodis$dosi.ini)
100*sd(episodis$nombre.d, na.rm = T)/mean(episodis$nombre.d, na.rm = T)
100*sd(episodis$edat, na.rm = T)/mean(episodis$edat)
100*sd(episodis$edadinic, na.rm = T)/mean(episodis$edadinic, na.rm = T)
100*sd(episodis$edadini1, na.rm = T)/mean(episodis$edadini1, na.rm = T)

```

```

#variables numèriques
hist(episodis$dosi.ini)
hist(episodis$nombre.d)
hist(episodis$edat)
hist(episodis$edadinic)
hist(episodis$edadini1)

freq(episodis$dosi.ini)
freq(episodis$nombre.d)
freq(episodis$edat)
freq(episodis$edadinic)
freq(episodis$edadini1)

summary(episodis)
sd(episodis$dosi.ini, na.rm = T)
sd(episodis$nombre.d, na.rm = T)
sd(episodis$edat, na.rm = T)
sd(episodis$edadinic, na.rm = T)
sd(episodis$edadini1, na.rm = T)
#variables categoriques

#Gràfics circulars

pie(c(170,419,80,73), labels=c( "0","1","2", ">2"), shadow=FALSE, edges=200, radius=0.8,
col=brewer.pal(4,"Blues"),
  main=("Número de parelles en els darrers 6 mesos"))
pie(c(466,191,92), labels=c( "No","Esporàdicament","Diàriament"), shadow=FALSE, edges=200,
radius=0.8, col=brewer.pal(3,"Blues"),
  main=("Consum de cocaïna"))
pie(c(158,323,39,66,84,88), labels=c( "Sí, fixe",
                                     "Sí, ocasionalment",
                                     "Mai",
                                     "Atur retribuït",
                                     "Atur no retribuït",
                                     "Incapacitat"),
  shadow=FALSE, edges=200, radius=0.8, col=brewer.pal(6,"Blues"), main=("Situació laboral"))
pie(c(631,249), labels=c( "No","Sí"), shadow=FALSE, edges=200, radius=0.8,
col=brewer.pal(2,"Blues"),
  main=("Mort (tallant el seguiment a 31/12/2013)"))

par(las=1)
plot(episodis$estudios,col=brewer.pal(6,"Blues"),main=("Estudios"))

plot(sexo~pacienta,data=episodis,col=brewer.pal(3,"Blues"),main=("Sexe vs
Pacienta"),xlab=("Pacient actiu"),ylab=("Sexe"))

```

```
plot(consumco~penales,data=episodis,col=brewer.pal(3,"Blues"),main=("Viavenos vs Penales"),xlab=("Antecedents penals"),ylab=("Consum de cocaïna"))
```

```
par(mfrow=c(1,2),las=1)
mosaicplot(vih~factor(preserva),data=episodis,col=brewer.pal(6,"Blues"),main=("Vih vs Preserva"),ylab=("Ús del preservatiu en les relacions"),xlab=("Infecció per VIH"))
mosaicplot(vih~factor(preserva),data=subset(episodis, preservava != 'NoRel'),col=brewer.pal(6,"Blues"),main=("Vih vs Preserva"),ylab=("Ús del preservatiu en les relacions"),xlab=("Infecció per VIH"))
```

```
mosaicplot(sexo~factor(preserva),data=episodis,col=brewer.pal(6,"Blues"),main=("Sexe vs Preserva"),xlab=("Sexe"),ylab=("Ús del preservatiu en les relacions"))
mosaicplot(sexo~factor(preserva),data=subset(episodis, preservava != 'NoRel'),col=brewer.pal(6,"Blues"),main=("Sexe vs Preserva"),xlab=("Sexe"),ylab=("Ús del preservatiu en les relacions"))
par(mfrow=c(1,1))
```

```
plot(vih~programa,data=episodis,col=brewer.pal(6,"Blues"),main=("Vih vs Preserva"),ylab=("Infecció per VIH"),xlab=("Participat en programes d'intercanvi de xeringues"))
```

```
#tassa incidència casos en tractament
```

```
any <- 1995
```

```
a <- as.Date("1996-1-1")
```

```
b <- as.Date("1995-1-1")
```

```
for(i in 1:19)
```

```
{
```

```
  res <- sum(episodis$datainic<a & episodis$datafina>b, na.rm=T)
```

```
  print(any)
```

```
  print(res)
```

```
  any <- any+1;
```

```
  year(a) <- year(a)+1;
```

```
  year(b) <- year(b)+1;
```

```
}
```

## *Descriptiva dosis*

```
## Les dades
load("dosis.RData")

## Dades ordenats cronològicament
dosis <- with(dosis, dosis[order(nummetad, date), ])

## Noves variables
library(plyr)

# Temps entre un canvi i un altre
dosis <- ddply(dosis, ~nummetad, transform, tiempentre = c(0, diff(date)))
dosis <- ddply(dosis, ~nummetad, transform, tiemptotal = cumsum(tiempentre))
head(dosis)

save(dosis, file='dosis.RData')

## El gràfic
library(lattice)

dosis$totemps <- dosis$tiemptotal/365.25
pdf('Todos5Anyos3.pdf', width=16, height=10)
xyplot(dosiprec~totemps | as.factor(nummetad), data=dosis, layout=c(11, 9),
  par.strip.text=list(cex=.85, font=2),
  main="Dosi de metadona",
  strip = strip.custom(var.name='Id', strip.names = T, strip.levels=T, sep=' ', bg="wheat1", style=1),
  xlab=list('Anys', font=4),
  ylab=list('Dosi', font=4),
  scales=list(font=4, tck=c(0.7, 0.5), x=list(alternating=T, at=seq(0, 5, 1), limits=c(0, 5)),
y=list(alternating=T)),
  panel=function(x, y){
    panel.grid(h=-1, v=-1)
    panel.xyplot(x, y, type='b', col=1, pch=18, cex=1.1, lwd=2)
    ## Alternativa (¡Sólo para pacientes con más de una dosis!)
    # panel.xyplot(x, y, type='s', col=1, pch=18, cex=1.1, lwd=2)
  })
dev.off()
```

- Modelització

```
load("dosis.RData")
load("episodis.RData")
dades <- subset(episodis, mort.2013 == "No" )

plot(density(dades$dosi.ini[dades$pacienta=="No"]),ylab="Densitat",xlab="Dosi inicial")
lines(density(dades$dosi.ini[dades$pacienta=="Si"]), col="blue")
legend("topright",c("Pacients no actius", "Pacients actius"),col=c(1,4),lty=1)
plot(density(dades$edat[dades$pacienta=="No"]),ylab="Densitat",xlab="Edat al inici del tractament")
lines(density(dades$edat[dades$pacienta=="Si"]), col="blue")
legend("topright",c("Pacients no actius", "Pacients actius"),col=c(1,4),lty=1)
plot(density(dades$nombre.d[dades$pacienta=="Si"], na.rm=T),
col="blue",ylab="Densitat",xlab="Nombre de CD4")
lines(density(dades$nombre.d[dades$pacienta=="No"], na.rm=T))
legend("topright",c("Pacients no actius", "Pacients actius"),col=c(1,4),lty=1)

pairs(dades[,c(6,10,11,23,28)])
boxplot(nombre.d~vih, dades, pch=16, xlab="VIH", ylab="Nombre de CD4",col="blue")
boxplot(nombre.d~vih,dades,pch=16, xlab="Pacient actiu", ylab="Nombre de CD4",col="blue")

mosaicplot(penales~pacienta,data=dades)
mosaicplot(viavenos~pacienta,data=dades)

m <- glm(pacienta~1,data=dades,family=binomial)
summary(m)

m1 <- glm(pacienta~sexo,data=dades,family=binomial)
summary(m1)
anova(m,m1,test="Chisq")
#no significatiu

m1b <- glm(pacienta~consumco,data=dades,family=binomial)
summary(m1b)

m1c <- glm(pacienta~edadinic,data=dades,family=binomial)
summary(m1c)

m2 <- glm(pacienta~edadinic+edat,data=dades,family=binomial)
summary(m2)

m2b <- glm(pacienta~edat+nombre.d,data=dades,family=binomial)
summary(m2b)
```

```
m2c <- glm(pacienta~edat+vih,data=dades,family=binomial)
summary(m2c)
```

```
m2d <- glm(pacienta~edat+penales,data=dades,family=binomial)
summary(m2d)
```

```
m2e <- glm(pacienta~edat+penales+laboral,data=dades,family=binomial)
summary(m2e)
```

```
m2f <- glm(pacienta~edat+penales+estudios,data=dades,family=binomial)
summary(m2f)
```

```
m2g <- glm(pacienta~edat+penales+viavenos,data=dades,family=binomial)
summary(m2g)
```

```
m3 <- glm(pacienta~edat+penales+viavenos+viaveno1,data=dades,family=binomial)
summary(m3)
```

```
m3a <- glm(pacienta~edat+penales+viavenos+programa,data=dades,family=binomial)
summary(m3a)
```

```
m3b <- glm(pacienta~edat+penales+viavenos+relacdin,data=dades,family=binomial)
summary(m3b)
```

```
m3c <- glm(pacienta~edat+penales+viavenos+preserva,data=dades,family=binomial)
summary(m3c)
```

```
m4d <- glm(pacienta~edat+viavenos+penales+gonococ,data=dades,family=binomial)
summary(m4d)
```

```
#model definitiu
```

```
dades$viavenos <- relevel(dades$viavenos,"No")
```

```
mde <- glm(pacienta~edat+penales+viavenos,data=dades,family=binomial)
```

```
summary(mde)
```

```
anova(mde, test="Chisq")
```

```
logistic.display(mde)
```

```

#Interpretació en l'escala dels logit
exp(-0.10513)
exp(0.41606)
exp(0.42858)
#Interpretació en l'escala dels odds
exp(-0.10513)*100-100
exp(0.41606)*100-100
exp(0.42858)*100-100
#prob marginal
#Escala probabilitats
pm <- 518/(518+113)
pm*(1-pm)*-0.10513
pm*(1-pm)*0.41606
pm*(1-pm)*0.42858

#prova
mdeb <- glm(pacienta~edat+penales+viavenos+nombre.d,data=dades,family=binomial)
summary(mdeb)
#nombre.d no significatiu però disminueix molt el AIC (perdem moltes observacions 447)
#no s'inclou perquè perdem molta informació i a més a més no es significatiu.

#VALIDACIÓ DEL MODEL----

marginalModelPlots(m4,labels=row.names(dades),id.method=abs(cooks.distance(mde)), id.n=5)
outlierTest(mde)
vif(mde)

## Prova de Hosmer-Lemeshow
HLtest(mde)

#Altres gràfiques
avPlots(mde,id.method=abs(cooks.distance(mde)),id.n=5)
residualPlots(mde)
influenceIndexPlot(mde,label=row.names(dades),vars=c("Cook", "Student", "hat"), id.n=10)
dades[177,]
influenceIndexPlot(mde, vars=c("Cook", "hat"), id.n=3)

# Corba roc
ROC(form = pacienta ~edat+penales+viavenos, data=dades, plot="ROC")

```

## ▪ Supervivència

```
load("dosis.RData")
load("episodis.RData")
attach(episodis)
add<-Surv(sup,estat)

#no paramètric- funció de supervivència ----
surv <- survfit(Surv(sup, estat)~1, episodis)
surv

surv1 <- survfit(Surv(sup, estat)~sexo+vih, episodis)
surv1

summary(surv1)
plot(surv,xlab="Temps (en anys)",ylab="Supervivència",lty=c(1,3))
title("Funció de supervivència")

plot(surv1,xlab="Temps (en anys)",ylab="Supervivència",col=c(1,2,4,5),lty=1)
legend("topright",c("Home-VIH=Sí","Home-VIH=No","Dona-VIH=Sí","Dona-
VIH=No"),col=c(1,2,4,5),lty=1)
title("Funcio de Supervivència per Sexe i VIH")

surv2 <- survfit(Surv(sup, estat)~penales, episodis)
surv2
plot(surv2,xlab="Temps (en anys)",ylab="Supervivència",col=c(1,2),lty=1)
legend("topright",c("No","Sí"),col=c(1,2),lty=1)
title("Funcio de Supervivència per Antecedents Presó")

surv3 <- survfit(Surv(sup, estat)~relacdin, episodis)
surv3
plot(surv3,xlab="Temps (en anys)",ylab="Supervivència",col=c(1,2),lty=1)
legend("topright",c("Sí","No"),col=c(1,2),lty=1)
title("Funcio de Supervivència per Relacions per diners o drogues")

#Comprovació funcions de supervivència ----
# Logrango
survdifff(Surv(sup, estat)~sexo, episodis)
#peto
survdifff(Surv(sup, estat)~sexo,rho=1, episodis)

# Logrango
survdifff(Surv(sup, estat)~consumco, episodis)
#peto
survdifff(Surv(sup, estat)~consumco,rho=1, episodis)
```



```

# Logrango
survdifff(Surv(sup, estat)~penales, episodis)
#peto
survdifff(Surv(sup, estat)~penales,rho=1, episodis)

# Logrango
survdifff(Surv(sup, estat)~laboral, episodis)
#peto
survdifff(Surv(sup, estat)~laboral,rho=1, episodis)

# Logrango
survdifff(Surv(sup, estat)~vih, episodis)
#peto
survdifff(Surv(sup, estat)~vih,rho=1, episodis)

#Funcions de supervivència----
#funció supervivència vih si
add.vih.si<-survfit(add~vih,subset=(episodis$vih=="Si"))
#funció supervivència vih no
add.vih.no<-survfit(add~vih,subset=(episodis$vih=="No"))

par(mfrow = c(1,2))
#Funció de supervivència
plot(add.vih.si,xlab="Temps (anys)",ylab="Supervivència",lty=c(1,3))
title("Individus amb VIH ")
plot(add.vih.no,xlab="Temps (anys)",ylab="Supervivència",lty=c(1,3))
title("Individus sense VIH")
# Tot i no mostrar grans diferències en el test logrango les funcions de supervivència
#Son diferents

# funció supervivència VIH si
add.sexo.dona<-survfit(add~sexo,subset=(episodis$sexo=="Dona"))
# funció supervivència VIH no
add.sexo.home<-survfit(add~sexo,subset=(episodis$sexo=="Home"))

plot(add.sexo.dona,xlab="Temps (anys)",ylab="Supervivència",lty=c(1,3))
title("Dones")
plot(add.sexo.home,xlab="Temps (anys)",ylab="Supervivència",lty=c(1,3))
title("Homes")
#En el cas de les dones sembla que existeix mes variabilitat
par(mfrow = c(1,1))

```

```

#MODEL DE COX----
# anàlisi semi-paramètric
# Només -> no censurats
Y=Surv(episodis$sup, episodis$estat==1)
coxmod <- coxph(Surv(sup, estat)~consumco+laboral+vih+sexo, episodis)
summary(coxmod)
anova(coxmod)

## Comprovar proporcionalitat de riscos

cox.zph(coxmod)
par(mfrow = c(1,2))
(testmodel<- cox.zph(coxmod))
for (i in 6:7)
{plot(testmodel,var=i)}
par(mfrow = c(1,1))
# La variable laboral no compleix la hipòtesi de proporcionalitat de riscos.

coxmod2 <- coxph(Surv(sup, estat)~consumco+vih+sexo, episodis)
summary(coxmod2)
anova(coxmod2)

## Comprovar proporcionalitat de riscos
cox.zph(coxmod2)
testmodel2<- cox.zph(coxmod2)
for (i in 1:4)
{plot(testmodel2,var=i)}

#Var sexe----
#hipòtesis proporcionalitat de riscos gràficament---
add<-Surv(sup,estat)
(add.sexe<-survfit(add~factor(sexo)))

#KM sexe
plot(add.sexe,xlab="Temps (en anys)",ylab="Supervivència",col=c(1,2),lty=c(1,2))
legend("topright",c("Home ", "Dona"),col=c(1,2),lty=c(1,2))
title("Funcio de Supervivència per Sexe")

plot(add.sexe,fun="cloglog",xlab="Log-Temps (en anys)",ylab="log-log
Supervivència",lty=c(1,2),col=c(1,2))
legend("topleft",c("Home ", "Dona"),lty=c(1,2),col=c(1,2))
title("Corba log-log per Sexe")

#Var consumco----
#hipòtesis proporcionalitat de riscos gràficament
(add.consumco<-survfit(add~factor(consumco)))

```

```

#KM consumco
plot(add.consumco,xlab="Temps (en anys)",ylab="Supervivència",col=c(1,2,4))
legend("topright",c("NO ","Esporàdicament","Diariament "),col=c(1,2,4))
title("Funcio de Supervivència per Consum de cocaïna ")

plot(add.consumco,fun="cloglog",xlab="Log-Temps (en anys)",ylab="log-log
Supervivència",col=c(1,2,4),lty=1)
legend("topleft",c("No ","Esporàdicament","Diariament "),col=c(1,2,4),lty=1)
title("Corba log-log per consum de Cocaïna")

#Var laboral----
#hipòtesis proporcionalitat de riscos gràficament
(add.laboral<-survfit(add~factor(laboral)))

#KM laboral
plot(add.laboral,xlab="Temps (en anys)",ylab="Supervivència",col=c(brewer.pal(5,"Set1"),9),lty=1)
legend("topright",c("Sí, Fixe ","Sí, ocasionalment","Mai","Atur Retribuit", "Atur no retribuit",
"Incapacitat"),col=c(brewer.pal(5,"Set1"),9),lty=1)
title("Funcio de Supervivència per Situació Laboral")

plot(add.laboral,fun="cloglog",xlab="Log-Temps (en anys)",ylab="log-log
Supervivència",col=c(brewer.pal(5,"Set1"),9),lty=1)
legend("topleft",c("Sí, Fixe ","Sí, ocasionalment","Mai","Atur Retribuit", "Atur no retribuit",
"Incapacitat"),col=c(brewer.pal(5,"Set1"),9),lty=1)
title("Corba log-log per Situació Laboral")

#Var VIH----
#hipotesis proporcionalitat de riscos graficament
(add.vih<-survfit(add~factor(vih)))

#KM VIH
plot(add.vih,xlab="Temps (en anys)",ylab="Supervivència",col=c(1,2),lty=c(1,2))
legend("topright",c("Sí ","No"),col=c(1,2),lty=c(1,2))
title("Funcio de Supervivència per VIH")

plot(add.vih,fun="cloglog",xlab="Log-Temps en anys",ylab="log-log
Supervivència",lty=c(1,2),col=c(1,2))
legend("topleft",c("Sí ","No"),lty=c(1,2),col=c(1,2))
title("Corba log-log per VIH")

#Model estratificat
Y <- Surv(sup, estat)
coxmodest <- coxph(Surv(sup, estat)~consumco+strata(laboral)+vih+sexo, episodis)

```

```

summary(coxmodest)

# Anem a comprovar la significació
ANOVA (coxmodest)

#Proporcionalitat de riscos
cox.zph(coxmodest)
par(mfrow = c(2,2))
testmodel2<- cox.zph(coxmod2)
for (i in 1:4)
{plot(testmodel2,var=i)}
par(mfrow = c(1,1))

coxsinvih <- coxph(Surv(sup, estat)~consumco+strata(laboral)+sexo, episodis)
summary(coxsinvih)

#Akaike (ens interessa minimitzar)
#AIC
#anova(coxmod)
(AICM <- -2*(-2418.8)+2*4)
#AIC
#anova(coxmod2)
(AICM <- -2*(-2623.4)+2*3)
#AIC
#anova(coxmodest)
(AICM <- -2*(-1831.9)+2*3)
#AIC
#anova(coxsinvih)
(AICM <- -2*(-1954.6)+2*2)

```