



Universitat de Girona

ASPECTES METODOLÒGICS I APLICACIONS DE LA MODELITZACIÓ DEL TEMPS DE SUPERVIVÈNCIA MULTIVARIANT MITJANÇANT MODELS MIXTES

Gemma RENART I VICENS

ISBN: 978-84-692-516-45

Dipòsit legal: GI. 905-2012

<http://hdl.handle.net/10803/7705>

ADVERTIMENT. L'accés als continguts d'aquesta tesi doctoral i la seva utilització ha de respectar els drets de la persona autora. Pot ser utilitzada per a consulta o estudi personal, així com en activitats o materials d'investigació i docència en els termes establerts a l'art. 32 del Text Refós de la Llei de Propietat Intel·lectual (RDL 1/1996). Per altres utilitzacions es requereix l'autorització prèvia i expressa de la persona autora. En qualsevol cas, en la utilització dels seus continguts caldrà indicar de forma clara el nom i cognoms de la persona autora i el títol de la tesi doctoral. No s'autoritza la seva reproducció o altres formes d'explotació efectuades amb finalitats de lucre ni la seva comunicació pública des d'un lloc aliè al servei TDX. Tampoc s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant als continguts de la tesi com als seus resums i índexs.

ADVERTENCIA. El acceso a los contenidos de esta tesis doctoral y su utilización debe respetar los derechos de la persona autora. Puede ser utilizada para consulta o estudio personal, así como en actividades o materiales de investigación y docencia en los términos establecidos en el art. 32 del Texto Refundido de la Ley de Propiedad Intelectual (RDL 1/1996). Para otros usos se requiere la autorización previa y expresa de la persona autora. En cualquier caso, en la utilización de sus contenidos se deberá indicar de forma clara el nombre y apellidos de la persona autora y el título de la tesis doctoral. No se autoriza su reproducción u otras formas de explotación efectuadas con fines lucrativos ni su comunicación pública desde un sitio ajeno al servicio TDR. Tampoco se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al contenido de la tesis como a sus resúmenes e índices.

WARNING. Access to the contents of this doctoral thesis and its use must respect the rights of the author. It can be used for reference or private study, as well as research and learning activities or materials in the terms established by the 32nd article of the Spanish Consolidated Copyright Act (RDL 1/1996). Express and previous authorization of the author is required for any other uses. In any case, when using its content, full name of the author and title of the thesis must be clearly indicated. Reproduction or other forms of for profit use or public communication from outside TDX service is not allowed. Presentation of its content in a window or frame external to TDX (framing) is not authorized either. These rights affect both the content of the thesis and its abstracts and indexes.

ASPECTES METODOLÒGICS I APLICACIONS DE LA MODELITZACIÓ DEL TEMPS DE SUPERVIVÈNCIA MULTIVARIANT MITJANÇANT MODELS MIXTES

GEMMA RENART I VICENS

Grup de Recerca en Estadística, Economia Aplicada i Salut (GRECS),
Universitat de Girona

Febrer de 2009

ASPECTES METODOLÒGICS I APLICACIONS DE LA MODELITZACIÓ DEL TEMPS DE SUPERVIVÈNCIA MULTIVARIANT MITJANÇANT MODELS MIXTES

GEMMA RENART I VICENS

Grup de Recerca en Estadística, Economia Aplicada i Salut (GRECS),
Universitat de Girona

Directors

Dra. Maria Antònia Barceló
GRECS, UdG

Dr. Marc Saez
GRECS, UdG

La doctora **Maria Antònia Barceló Rado**, Titular d'Universitat, i el doctor **Marc Saez Zafra**, Catedràtic d'Universitat, ambdós del Departament d'Economia de la Universitat de Girona,

AUTORITZEN:

La presentació de la Tesis Doctoral amb títol '**Aspectes metodològics i aplicacions de la modelització del temps de supervivència multivariant mitjançant models mixtes**', realitzada per la Sra. **Gemma Renart Vicens**, sota la nostra immediata direcció i supervisió i que presenta per obtenir el grau de Doctor per la Universitat de Girona.

Girona, 22 de desembre de 2008

Dra. Maria Antònia Barceló Rado
Directora

Dr. Marc Saez Zafra
Director

Gemma Renart Vicens
Doctorand

Als meus pares.

Agraïments

Al llarg de tots aquests anys ha estat habitual el fet de pensar en el moment d'omplir aquesta pàgina, però sempre ho veia tan lluny...i ara que finalment ha arribat se'm fa molt difícil trobar les paraules adequades per expressar el meu agraïment a totes les persones, que d'alguna manera, han estat al meu costat i m'han ajudat durant aquest període de la meva vida.

Crec que en primer lloc he d'esmentar els meus directors, el Dr. Marc Saez i la Dra. M^a Antònia Barceló. Els dos han posat a la meva disposició els seus coneixements i la seva experiència i en tot aquest temps no m'ha mancat mai el seu inestimable suport. Gràcies pels vostres suggeriments, per les vostres crítiques, pel temps que m'heu dedicat...són tantes coses que us he d'agrair!!. Marc, gràcies per confiar en mi en el seu moment i donar-me l'oportunitat d'iniciar un nou camí en el món de la recerca. La teva proximitat, disposició i implicació en aquest treball ha fet que tot fos una mica més planer. Antònia, gràcies per creure en mi i formar part d'aquest treball. La teva dedicació, esperit crític i sàvies decisions han estat fonamentals en el seu desenvolupament. Treballar al vostre costat ha estat un plaer i espero ho continuï essent. Sincerament gràcies, perquè heu estat més que els meus directors.

També vull dedicar unes ratlles d'aquests agraïments a la Dra. Carme Saurina. Durant aquests últims anys t'ha tocat compartir assignatura amb mi i has hagut d'aguantar-me en els moments més baixos (sobretot en el tram final), però sempre has tingut paraules per animar-me a continuar. Treballar amb tu és molt enriquidor. Gràcies Carme per ser com ets!!

I a la resta de membres del grup de recerca GRECS: Germà, Lluís, Laura, Àngels, Sònia i Aïna, dir-vos que em sento afortunada d'haver coincidit amb vosaltres i formar part d'aquest grup. Sempre he pogut comptar amb vosaltres i sé que sempre hi podré comptar. A vosaltres, Laura i Àngels, també agrair-vos les bones estones que em passat fent el cafè i compartint despatx; m'heu fet somriure quan més ho necessitava, m'heu aconsellat quan tenia algun dubte...Crec que us puc dir: gràcies per la vostra amistat.

Durant aquesta etapa de la meva vida, també he anat trobant diferents persones que m'han ajudat d'alguna manera i amb les que he compartit molts bons moments. Amb

algunes d'elles segueixo mantenint el contacte, d'altres, s'han desviat del meu camí, però sempre estaran al meu record. Gràcies.

També vull agrair a la Meia Faixedas, responsable d'Epidemiologia i Programes, Regió Sanitària de Girona, CatSalut i Delfí Faixedas, Servei d'Atensió Primària Girona, Institut Català de la Salut (ICS) per proporcionar-me les dades utilitzades en l'aplicació pràctica d'aquest treball.

A la Universitat de Girona per permetre'm gaudir d'un agraït rebaix docent en l'últim tram de l'elaboració d'aquest treball.

I per últim, però sempre els primers en els meus pensaments, vull agrair el suport rebut per part de tota la meva família. Als meus pares, Ramon i Carme, perquè sóc qui sóc i sóc com sóc gràcies a vosaltres, perquè sempre heu estat al meu costat i m'heu recolzat en totes les meves decisions. Al meu germà Jaume perquè també t'ha tocat aguantar-me en alguns dels meus moments baixos. A tu Albert, gràcies per saber entendre que aquesta és una feina diferent, pel teu suport i paciència mostrada durant tots aquests anys, i a tu Nora, ets la "personeta" que sense saber-ho m'has donat més ànims per acabar. A tots, us estimo MOLT.

Moltes gràcies a tots i als que possiblement m'hagi deixat, accepteu també el meu agraïment. Això que teniu a les mans ha estat possible gràcies a tots vosaltres. Moltes gràcies per tot, que no ha estat poc!!!

1. Introducció i objectius	15
Objectius general i específics:	20
2. Anàlisi de la supervivència multivariant.	21
2.1. Aproximació d'Andersen i Gill al model de regressió de Cox.	24
2.2. Models condicionals i models marginals.	29
2.2.1. Models marginals.....	30
2.2.1.1. Estimació robusta de la matriu de covariàncies en l'aproximació AG.....	30
2.2.1.2. El mètode Wei, Lin i Weissfeld.	31
2.2.1.3. El mètode Prentice, Williams i Peterson.	34
2.2.1.4. Cas pràctic.....	35
2.2.2. Models condicionals.	38
2.2.2.1 Models de fragilitat.	38
Referències	41
3. Models lineals generalitzats i Models lineals generalitzat mixtes.	43
3.1. Models lineals generalitzats.	45
3.2. Models lineals generalitzats mixtes.	51
Referències	63
4. Mètodes	65
4.1. El mètode d'Andersen	67
4.1.1. Metodologia.	67
4.1.2. Adaptació del model pel cas de dades de supervivència multivariant.	69
4.1.3. Adaptació al cas d'estudi: reingressos a l'hospital degut a la malaltia pulmonar obstructiva crònica (MPOC).....	70
4.2. El mètode de Goldstein.	71
4.2.1. Metodologia.	72
4.2.2. Adaptació al cas d'estudi: reingressos a l'hospital deguts a MPOC.	75
Referències	78
5. Epidemiologia espacial	81

5.1. Els mapes de malalties.....	83
5.2. Mètodes de suavització de les RME	85
5.3. Variabilitat geogràfica dels reingressos hospitalaris deguts a la malaltia pulmonar obstructiva crònica (MPOC).....	86
5.4. Estimació del model.....	87
5.4.1. Aproximació bayesiana vs. aproximació freqüentista.....	87
5.4.2. Estimació bayesiana.....	88
5.4.3. Convergència i eficiència de les estimacions.....	91
Referències.....	93
6. Resultats de l'estudi.	97
6.1. Descriptius de la base de dades utilitzada.....	99
6.1.1. Anàlisi global de les dades.....	102
6.1.2. Resum de l'anàlisi dels individus segons el nombre d'ingressos.....	112
6.1.3. Anàlisi del nombre de visites entre diferents ingressos	114
6.2. Resultats obtinguts en l'aplicació dels mètodes d'Andersen i de Goldstein sense efectes espacials.	116
6.2.1. Resultats pel mètode d'Andersen sense efectes espacials.....	117
6.2.2. Resultats pel mètode de Goldstein sense efectes espacials.	125
6.3. Resultats obtinguts en l'aplicació dels mètodes d'Andersen i de Goldstein amb efectes espacials.	132
6.4. Comparació dels resultats obtinguts pels diferents mètodes aplicats.	141
Referències.....	143
ANNEXES.....	145
ANNEXI: ANÀLISI DE LA SUPERVIVÈNCIA UNIVARIANT.....	1457
A.I.1. Introducció.....	147
A.I.2. Funcions importants dins l'anàlisi de la supervivència	148

A.I.3. Estimació de la funció de supervivència i de la funció de risc.....	149
A.I.3.1. Estimació de la funció de supervivència.....	150
A.I.3.2. Estimació de la funció de risc.....	155
A.I.4. El model de regressió de Cox o model de riscos proporcionals.....	157
A.I.4.1. El model de Cox o model de riscos proporcionals.....	158
A.I.4.2. Anàlisi residual del model de Cox.....	164
A.I.4.3. Formulació del model de Cox a partir de processos comptadors.....	171
A.I.4.3.1. Introducció a la teoria dels processos comptadors.....	171
A.I.4.3.2. El model de Cox estàndard a partir de processos comptadors.....	174
Referències.....	177
<i>ANNEX II: EL MÈTODE JACKKNIFE.....</i>	<i>179</i>
Referències.....	183
<i>ANNEX III. CENTRE D'ATENCIÓ PRIMÀRIA (CAP) I HOSPITALS DE LA REGIÓ SANITÀRIA DE GIRONA.....</i>	<i>185</i>
<i>ANNEX IV. SINTAXIS I MODELS.....</i>	<i>189</i>
A.IV.1. ANÀLISI ANDERSEN.....	189
A.IV.1.1. Sintaxis.....	189
A.IV.1.2. Model Andersen sense efectes espacials.....	192
A.IV.1.3. Model Andersen amb efectes espacials.....	193
A.IV.2. ANÀLISI GOLDSTEIN.....	193
A.IV.2.1. Sintaxis.....	193
A.IV.2.2. Model Goldstein sense efectes espacials.....	196
A.IV.2.3. Model Goldstein amb efectes espacials.....	197
<i>ANNEX V: MODELLING OF MULTIVARIATE SURVIVAL DATA BY MEANS OF GENERALIZED LINEAR MIXED MODELS. THE CASE OF HOSPITAL ADMISSIONS DUE TO CHRONIC OBSTRUCTIVE PULMONARY DISEASE.....</i>	<i>199</i>
<i>ANNEX VI: ANÀLISI DELS INDIVIDUS SEGONS EL NOMBRE D'INGRESSOS QUE HAN GENERAT.....</i>	<i>229</i>
A.VI.1. Anàlisi dels individus que han generat un sol ingrés.....	229
A.VI.2.. Anàlisi dels individus amb dos ingressos.....	235

A.VI.3. Anàlisi dels individus amb tres ingressos.	241
A.VI.4. Anàlisi dels individus amb quatre ingressos	246
A.VI.5. Anàlisi dels individus amb cinc ingressos.....	251
<i>ANNEX VII: RESULTATS OBTINGUTS EN L'APLICACIÓ DEL MÈTODE D'ANDERSEN</i>	257
A.VII.1. MODEL SENSE EFECTES ESPACIALS	257
A.VII.2. MODEL AMB EFECTES ESPACIALS.....	272
<i>ANNEX VIII: RESULTATS OBTINGUTS EN L'APLICACIÓ DEL MÈTODE DE GOLDSTEIN.....</i>	281
A.VII.1. MODEL SENSE EFECTES ESPACIALS	281
A.VII.1. MODEL AMB EFECTES ESPACIALS.....	304

1. Introducció i objectius.

Els estudis de supervivència s'interessen pel temps que passa des de l'inici de l'estudi (diagnòstic de la malaltia, inici del tractament,...) fins que es produeix l'esdeveniment d'interès (mort, curació, millora,...). No obstant això, moltes vegades aquest esdeveniment s'observa més d'una vegada en un mateix individu durant el període de seguiment (dades de supervivència multivariant). En aquest cas, és necessari utilitzar una metodologia diferent a la utilitzada en l'anàlisi de supervivència estàndard. El principal problema que l'estudi d'aquest tipus de dades comporta és que les observacions poden no ser independents.

Fins ara, aquest problema s'ha solucionat de dues maneres diferents en funció de la variable dependent. Si aquesta variable segueix una distribució de la família exponencial s'utilitzen els *models lineals generalitzats mixtes* (GLMM); i si aquesta variable és el temps, variable amb una distribució de probabilitat no pertanyent a aquesta família, s'utilitza *l'anàlisi de supervivència multivariant*.

El que es pretén en aquesta tesi és unificar aquests dos enfocaments, és a dir, utilitzar una variable dependent que sigui el temps amb agrupacions d'individus o d'observacions, a partir d'un GLMM, amb la finalitat d'introduir nous mètodes pel tractament d'aquest tipus de dades.

En els dos primers capítols d'aquesta tesi es presenta una introducció als mètodes utilitzats fins al moment pel tractament de dades de supervivència multivariant. En el **capítol 2** s'introdueix l'aproximació d'Andersen i Gill al model de regressió de Cox així com els seus principals problemes (**apartat 2.1**), i es descriuen breument els principals mètodes dins dels models marginals (estimació robusta de la matriu de covariàncies en l'aproximació AG, el mètode de Wei, Lin i Weissfeld i el mètode de Prentice, Williams i Peterson) i dins dels models condicionals (models de fragilitat) (**apartat 2.2**).

En aquest punt cal esmentar també que en l'**Annex I** s'adjunta una explicació dels conceptes i funcions més importants dins l'anàlisi de supervivència així com la seva estimació. També es descriu el model semiparamètric més utilitzat dins l'anàlisi de supervivència, el model de riscos proporcionals o model de Cox, i la seva formulació a partir de processos comptadors. Tot el que es presenta en aquest annex són conceptes bàsics que poden ser útils per la comprensió del Capítol 2.

En el **capítol 3** es presenta la definició i formulació dels models lineals generalitzats i dels models lineals generalitzats mixtes així com de la seva estimació.

En el següent capítol, **capítol 4**, s'introdueixen els dos principals enfocaments, Andersen (**apartat 4.1**) i Goldstein (**apartat 4.2**), basats en el model GLMM per treballar amb variables resposta temps de supervivència multivariant.

Primer es descriu la metodologia utilitzada per Andersen (**secció 4.1.1**) que modela probabilitats en models multinivell introduint els anomenats pseudo-valors construïts a partir del mètode Jackknife (s'inclou un breu resum del mètode Jackknife a l'**Annex II**). A continuació es presenta l'adaptació d'aquest mètode per el cas de dades de supervivència multivariant (**secció 4.1.2**) i l'aplicació al cas pràctic que s'adjunta en aquest treball, reingressos a l'hospital degut a la malaltia pulmonar obstructiva crònica (**secció 4.1.3**).

En el següent apartat es presenta el segon enfoc estudiat en aquest treball, el mètode Goldstein (**apartat 4.2**), la idea principal del qual es basa en utilitzar un model de regressió logística després de reestructurar la base de dades convertint cada temps de supervivència en una seqüència binària anomenada resposta. A la **secció 4.2.1** es descriu la metodologia bàsica d'aquest mètode i a la **secció 4.2.2**, la seva adaptació al cas d'estudi.

Una de les innovacions que s'introdueixen en els models utilitzats és el fet de tenir en compte la variabilitat geogràfica en la seva definició, introduint efectes espacials. És per això que ha estat necessari dedicar un capítol a l'epidemiologia espacial (**capítol 5**). Una part d'aquest capítol està dedicada a introduir els conceptes bàsics dins del camp de l'epidemiologia espacial (**apartats 5.1 i 5.2**). A l'apartat 5.3 es descriuen els models utilitzats pels dos enfoc, Andersen i Goldstein, amb efectes espacials. I la resta del capítol està dedicada a l'estimació d'aquest tipus de models utilitzant l'aproximació bayesiana (**apartat 5.4**).

Finalment, el **capítol 6** és el que recull tots els resultats obtinguts en l'aplicació dels mètodes estudiats en el cas de reingressos a l'hospital deguts a la malaltia pulmonar obstructiva crònica. Es comença presentant uns descriptius globals de la base de dades utilitzada (**apartat 6.1**). A l'**Annex VI** s'adjunta una descripció d'aquesta base de dades més detallada. Després, en els **apartats 6.2 i 6.3** es mostren els resultats obtinguts per les dues aproximacions, Andersen i Goldstein, amb i sense la introducció

d'efectes espacials al model, acompanyats de la seva corresponent explicació. S'acaba aquest capítol amb una comparació dels mètodes estudiats (**apartat 6.4**).

Per acabar esmentar que en l'**Annex IV** s'adjunta tota la sintaxis utilitzada en la programació dels mètodes utilitzats, així com la descripció dels models.

Objectiu general:

Analitzar dades de supervivència multivariant mitjançant models lineals generalitzats mixtes.

Objectius específics:

Descriure els principals models de supervivència multivariant, aproximació d'Andersen i Gill al model de Cox de riscos proporcionals, models marginals i models condicionals; així com dels seus principals problemes (capítol 2).

Descriure la metodologia usada en els models lineals generalitzats i en els models lineals generalitzat mixtes (capítol 3).

Presentar els dos mètodes estudiats per tal de treballar amb dades de supervivència multivariant utilitzant models lineals generalitzats mixtes i adaptar aquests enfocaments al cas d'estudi (capítol 4).

Introduir els conceptes bàsics de l'epidemiologia espacial així com l'estimació de models amb efectes espacials utilitzant l'aproximació bayesiana (capítol 5).

Presentar i discutir els resultats obtinguts en l'aplicació dels mètodes presentats (capítol 6).

2. Anàlisi de la supervivència multivariant.

Quan la variable que es vol estudiar és el temps que passa fins que es produeix un determinat esdeveniment, el conjunt de tècniques estadístiques que s'utilitzen per analitzar aquest tipus de variables s'anomena anàlisi de la supervivència.

A l'Annex I es presenten els mètodes més importants quan l'esdeveniment d'interès només pot aparèixer com a màxim una vegada durant el període d'observació (anàlisi de la supervivència univariant). En aquest context, el mètode més utilitzat per estudiar la relació entre diferents variables registrades per cada individu i el temps de supervivència és l'anomenat model de Cox de riscos proporcionals.

Però sovint pot passar que els individus presentin diversos esdeveniments de diferent tipus o bé recurrències d'un mateix tipus d'esdeveniment (anàlisi de la supervivència multivariant) com per exemple quan s'estudien els múltiples atacs d'epilèpsia, la reaparició de certs tipus de càncers, etc.

En aquest context, els temps de supervivència d'un mateix individu no seran independents degut a l'existència d'heterogeneïtat individual i/o a la dependència "serial", i a més, els efectes de les variables explicatives poden ser diferents entre el primer esdeveniment (o recurrència) i els següents.

El fet que les observacions no siguin independents dóna lloc a l'incompliment de la hipòtesis de riscos proporcionals. Amb el propòsit de solucionar aquests problemes, el model de Cox de riscos proporcionals (Cox, 1972) i els mètodes estadístics associats basats en la versemblança parcial (Cox, 1975) es van generalitzar i situar en el context dels models d'intensitat multiplicatius per processos comptadors multivariants (Aalen, 1978; Andersen i Gill, 1982; Andersen i Borgan, 1985) (veure Annex I). Aquesta nova formulació no només ha fet possible ampliar el rang d'aplicació del model de Cox, sinó que també ha proporcionat una construcció molt més rigorosa i una justificació per la funció de versemblança parcial (Jacod, 1975; Arjas i Haara, 1984) i ha estimulat el desenvolupament d'una teoria asimptòtica per els estimadors parcials màxim versemblants (Andersen i Gill, 1982; Prentice i Self, 1983).

A partir d'aquesta nova formulació apareixen altres mètodes diferents, entre els quals, els més comuns són el mètode d'Andersen-Gill, els models marginals i els models condicionals, que es passen a explicar en els següents apartats.

2.1. Aproximació d'Andersen i Gill al model de regressió de Cox.

L'aproximació d'Andersen i Gill al model de riscos proporcionals, model AG (Andersen i Gill, 1982; Gill, 1984; Andersen i Borgan, 1985; Fleming i Harrington, 1991 i Andersen *et al.*, 1993) és una generalització del model de Cox en el context dels processos comptadors.

Tal i com ja s'explica a l'Annex I, els processos comptadors han resultat ser molt útils tant per la formulació com per l'anàlisi estadístic de determinats models de supervivència, així com per una demostració rigorosa d'algunes propietats matemàtiques d'aquests models. En particular, han desenvolupat el problema de les mesures repetides d'un mateix esdeveniment o diferents esdeveniments per un mateix individu dins d'un marc més natural.

Seguint Barceló (2001) es passa a donar algunes definicions bàsiques de la teoria de processos comptadors (veure Annex I).

Fins ara, les dades de supervivència es representaven com:

$$(\tilde{T}_1, D_1), \dots, (\tilde{T}_n, D_n)$$

on $\tilde{T}_i = \min\{T_i, C_i\}$ essent T_i el temps de supervivència de l'individu i , i C_i el temps de censura; $D_i = 1$ si s'ha produït l'esdeveniment i $D_i = 0$ si s'ha produït una censura (status).

En el model AG cada individu es representa com una sèrie d'observacions en els intervals (temps d'entrada, primer esdeveniment], (primer esdeveniment, segon esdeveniment], ..., (esdeveniment n -èssim, última observació]; i les dades de supervivència es representen com n parelles de processos estocàstics:

$$(N_1(u), Y_1(u)), \dots, (N_n(u), Y_n(u)) \quad u \geq 0$$

on $N_i(u) = I(\tilde{T}_i \leq u, D_i = 1)$ $i = 1, \dots, n$. Aquest procés comptador pren el valor 0 si $\tilde{T}_i > u$ o si $D_i = 0$ i 1 si $\tilde{T}_i \leq u$ i $D_i = 1$; és a dir, compta 1 si l'individu i falla en el període d'observació.

Es defineix també el procés $Y_i(u)$ com

$$Y_i(u) = I(\tilde{T}_i \geq u) \quad i = 1, \dots, n$$

és a dir, $Y_i(u)$ pren el valor 1 si l'individu i sobreviu fins al temps d'observació i 0 altrament.

Llavors, la filtració $F_t^{(n)}$ es defineix de la següent manera:

$$F_t^{(n)} = \{N_i(u), Y_i(u), X_i; i=1, \dots, n, 0 \leq u \leq t\}$$

Formalment, una filtració $\{F_t; t \geq 0\}$ és una família creixent de sub-sigma-àlgebres, i intuïtivament, resumeix la història natural d'un procés comptador i representa l'adquisició progressiva d'informació respecte al temps.

Un procés estocàstic $\{X(t); t \geq 0\}$ es diu que està adaptat a la filtració $\{F_t; t \geq 0\}$ si per tot $t \geq 0$, $X(t)$ és F_t -mesurable.

Finalment, un procés comptador multivariant

$$N = \{N_i(t); 0 \leq t < \infty; i=1, \dots, n\}$$

és un procés estocàstic adaptat a una filtració $\{F_t; t \geq 0\}$ tal que $N_i(0)=0$ per tot i , i $N(t) < \infty$. Dit d'una altra manera, les trajectòries de $N(t)$ són, amb probabilitat 1, contínues per la dreta, constants a trossos i amb discontinuïtats de salt d'alçada 1. El procés compta el nombre d'esdeveniments que han passat en cada un dels n individus en el moment t . Si l'esdeveniment és la mort de l'individu, llavors, $N_i(t) = \{0, 1\}$ ja que els individus només es poden morir una vegada.

Associat a aquest procés comptador existeix un procés d'intensitat acumulat Λ definit per:

$$\Lambda(t) = \int_0^t I\{\tilde{T} \geq u\} \alpha(u) du$$

on α representa la funció de risc. Aquest procés dóna per cada instant t el nombre de salts a l'interval $(0, t)$.

Algunes referències importants a la teoria dels processos comptadors i martingales són Brémaud i Jacod (1977), Jacod (1979) i Brémaud (1981).

Un cop vistes aquestes definicions, es passa a introduir el model AG o aproximació d'Andersen i Gill al model de Cox.

Es considera un procés comptador, $N_{ik}(t)$, on el subíndex $i=1,\dots,n$ es refereix a individus i $k=1,\dots,K$ a diferents tipus d'esdeveniments que un individu pot experimentar al llarg de la seva vida o bé, recurrències d'un mateix tipus d'esdeveniment. $N_{ik}(t)$ representa el nombre d'esdeveniments per un individu i en l'interval $[0,t]$. A cada $N_{ik}(t)$ li correspon un procés d'intensitat $\lambda_{ik}(t)$ suposant que es compleixen certes condicions de regularitat.

Llavors, el procés d'intensitat de $N_{ik}(t)$ respecte la filtració $F_t^{(n)}$ ve donat per la següent expressió:

$$\lambda_{ik}(t) = Y_{ik}(t)\alpha_0(t)e^{\beta'X_{ik}(t)} \quad (2.1)$$

essent α la funció de risc, $X_{ik}(t)$ un vector de variables explicatives per l'individu i -èssim i la recurrència k -èssim i β són paràmetres desconeguts.

El procés de martingala associat ve donat per:

$$M_{ik}(t) = N_{ik}(t) - \int_0^t Y_{ik}(u) e^{\beta'X_{ik}(u)} \lambda_0(u) du \quad (2.2)$$

Amb tot això s'ha aconseguit que en el model AG el risc que es produeixi un determinat esdeveniment o recurrència per un individu, segueixi la hipòtesis de riscos proporcionals i que no estigui afectat per recurrències anteriors. A més, un temps censurat no es considera una dada incompleta com en el model de Cox estàndard, sinó un individu que té el comptador de l'esdeveniment d'interès zero.

A la pràctica, si es disposa d'un individu que ha presentat diverses fallades, el model AG divideix aquest individu en diverses parts i cada part es tracta com si fossin individus diferents.

A continuació s'il·lustra el mètode explicat mitjançant un cas pràctic en el qual es parteix dels següents individus:

Individu 1: ha presentat l'esdeveniment d'interès una vegada entre els dies 0 i 52 i s'ha seguit fins el dia 114.

Individu 2: ha estat seguit fins el dia 114 i no ha presentat cap vegada l'esdeveniment.

Individu 3: ha presentat una primera vegada l'esdeveniment d'interès entre els dies 0 i 32; llavors, una segona entre els dies 32 i 78 i s'ha seguit fins el dia 114.

El model AG tractaria aquests tres individus tal com es mostra a la taula següent:

Taula 1. Exemple model AG.

Individu	Tentrada	Tsortida	Status
0001	0	52	1
0001	52	114	0
0002	0	114	0
0003	0	32	1
0003	32	78	1
0003	78	114	0

A les figures que es presenten a continuació es pot observar el disseny de la situació de partida (figura 1) d'aquest exemple i un disseny de l'aplicació del model AG (figura 2)

Figura 1.

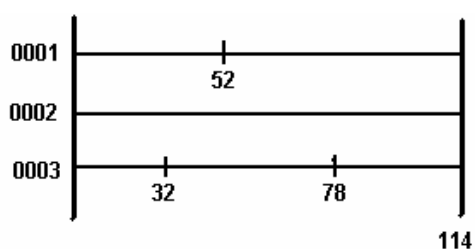
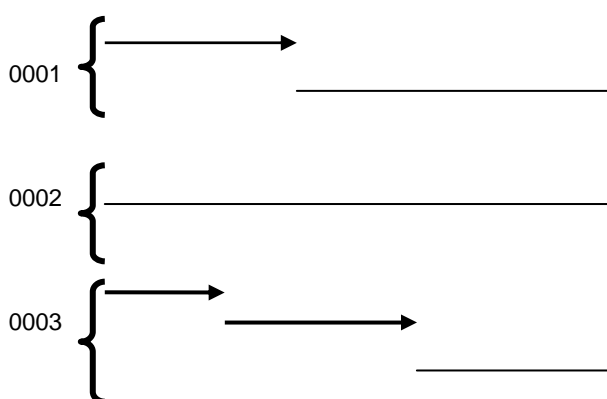


Figura 2.



Els individus es trenquen en diferents registres que es corresponen amb diversos intervals d'observació i cada registre és tractat com si fossin individus diferents. Els intervals es construeixen en funció de si l'individu presenta o no l'esdeveniment d'interès i de possibles canvis en les variables explicatives. En aquest disseny les variables explicatives són fixes dins de cada episodi, però no dins de cada individu. Així, doncs, és el risc de cada esdeveniment concret per un individu determinat, el que compleix la hipòtesis de riscos proporcionals i aquest risc no està afectat per esdeveniments anteriors del mateix individu.

El model AG és un model condicional respecte al conjunt de risc ja que un individu no està en risc fins que s'ha acabat l'esdeveniment anterior, és a dir, l'individu està en risc condicionat a esdeveniments anteriors.

L'aproximació d'AG és senzilla però les seves hipòtesis són molt restrictives i pot ser que no es puguin complir. La hipòtesi clau d'aquest model suposa que les múltiples observacions d'un mateix individu són independents (hipòtesis anomenada d'increments independents). Dit d'una altra manera, per un determinat individu el risc de fallar no està afectat per les fallades anteriors del mateix individu. A la pràctica, aquesta limitació pot donar lloc a estimadors esbiaixats i ineficients. A més, es sol produir una sobreestimació de la precisió dels estimadors ja que les observacions d'un mateix individu, normalment, estan correlacionades positivament (Barceló, 2001).

Tot això implica l'aparició dels models marginals i els models condicionals als quals s'ha fet esment a l'inici d'aquest apartat. La principal diferència entre aquests models és el tractament que donen a la dependència. Els primers estimen el model ignorant la dependència entre observacions, i corregint-la més tard; i els models condicionals l'estimen especificant explícitament la seva distribució de probabilitat i incorporant-la al model.

Tots els models que s'han vist fins al moment són models semiparamètrics (no necessiten especificar cap tipus de distribució de probabilitat pels temps de supervivència); però també existeixen aproximacions paramètriques que es basen en assumir una determinada funció de distribució per aquests temps. Les inferències basades en la parametrització d'aquesta distribució seran més precises o eficients; i els errors estàndard dels estimadors més petits. A més, aquestes aproximacions permeten realitzar inferències poblacionals i no limitar-se només a la mostra

analitzada. Tot això és possible sempre i quan la forma funcional escollida per els temps de supervivència sigui l'adequada, i això no sempre és fàcil.

2.2. Models condicionals i models marginals.

Tal com s'ha dit anteriorment, quan es vol analitzar mesures repetides, el principal problema del model de Cox és que les diferents observacions no són independents, la qual cosa comporta l'incompliment de la hipòtesis de riscos proporcionals.

L'aproximació d'Andersen-Gill vista a l'apartat anterior intenta solucionar aquest problema tractant cada individu com un procés comptador amb esdeveniments múltiples i amb els temps entre esdeveniments successius independents. O sigui, amb el model d'AG el risc d'una determinada recurrència per un individu segueix la hipòtesis de riscos proporcionals i els seus esdeveniments anteriors no afecten. Però, a la pràctica, no és raonable suposar que les diferents observacions d'un mateix individu no estiguin correlacionades, i aquesta limitació pot donar lloc a estimadors esbiaixats i ineficients (Barceló, 2001).

Per això, a finals dels anys vuitanta, apareixen els models marginals i els models condicionals en l'anàlisi de supervivència multivariant.

La principal diferència entre els models marginals i els condicionals està en com tracten la dependència entre les diverses observacions d'un mateix individu. Els models marginals estimen el model sense tenir en compte la dependència entre les observacions, però posteriorment la corregeixen mitjançant estimadors jackknife, bootstrap o "sàndwich". En canvi, els models condicionals estimen la dependència especificant explícitament la seva distribució de probabilitat i incorporant-la al model.

En aquest apartat, dins dels models marginals es parla de l'estimació robusta de la matriu de covariàncies en l'aproximació d'AG; del mètode WLW (Wei, Lin i Weissfeld, 1989) i del mètode PWP (Prentice, Williams i Peterson, 1981). Quan als models condicionals, es parlarà dels models de fragilitat.

2.2.1. Models marginals.

2.2.1.1. Estimació robusta de la matriu de covariàncies en l'aproximació AG.

La primera aproximació marginal que es tracta és l'estimació robusta de la matriu de covariàncies. Es suposa que l'aproximació AG proporciona estimadors consistents dels paràmetres associats a les variables explicatives, però en realitat, si existeix dependència, aquest mètode proporciona variàncies esbiaixades. Es tracta, doncs, d'estimar de forma robusta la matriu de variàncies i covariàncies, és a dir, corregir-la utilitzant estimadors adequats com per exemple l'estimador jackknife en el cas que es pensés en l'existència d'algun error d'especificació (Therneau i Hamilton, 1997).

L'estimador de Jackknife (veure Annex II) opera eliminant un a un els individus de la mostra i tornant a estimar el model amb els individus restants, és a dir, primer s'elimina un individu, s'estima el model sense ell, i llavors es torna a introduir; llavors s'elimina el següent individu i es torna a estimar el model sense ell, i així amb tots els individus del conjunt de dades. Els estimadors finals s'obtenen ponderant els obtinguts després de cada una d'aquestes eliminacions.

Formalment, l'estimador Jackknife es defineix com $(J - \bar{J})'(J - \bar{J})$, on J és la matriu d'elements J_{ij} , cada un d'ells igual a la variació de l'estimació del paràmetre β_j que resulta quan s'elimina l'observació i del conjunt de dades, i \bar{J} és una matriu que conté les mitjanes de les columnes de J (veure Annex II).

El problema d'aquesta aproximació marginal és que es basa en la consistència de les estimacions; no obstant això, les estimacions dels paràmetres seran consistents sempre que el model marginal (pel qual es compleix la hipòtesis de riscos proporcionals per cada una de les observacions, no necessàriament per cada un dels individus) estigui especificat correctament.

Tot i això, els estimadors que proporciona aquesta aproximació són menys esbiaixats que els que proporcionen les altres aproximacions marginals (Therneau i Hamilton, 1997; Barceló, 2001).

2.2.1.2. El mètode Wei, Lin i Weissfeld.

Un altre model marginal és el mètode Wei, Lin i Weissfeld, mètode WLW, (Wei *et al.*, 1989). En aquest mètode cada individu amb $k=1,2,\dots,K$ recurrències es divideix en submostres, la primera des de l'inici de l'estudi fins a la primera recurrència, la segona des de l'inici de l'estudi fins a la segona recurrència, etc. Així, cada recurrència de l'esdeveniment (o tipus diferent d'esdeveniment) es modelitza com un estrat diferent utilitzant el model de riscos proporcionals de Cox i els paràmetres associats a les variables explicatives, β_k , s'estimen per màxima versemblança parcial. El resultat final és que tots els individus apareixen en tots els estrats.

Wei *et al.*, (1989) demostren que la distribució conjunta dels β_k ($k=1,\dots,K$) és aproximadament normal amb una covariància que es pot estimar consistentment. Els paràmetres per cada una de les variables explicatives es poden obtenir combinant linealment els diferents β_k .

És a dir, per cada una de les k recurrències ($k=1,\dots,K$) que pot tenir un individu i ($i=1,\dots,n$), la funció de risc pot expressar-se com:

$$\lambda_{ik}(t) = \lambda_{0k}(t)e^{\beta_k' X_{ik}(t)} \quad t \geq 0 \quad (2.3)$$

on $\lambda_{0k}(t)$ és una funció de risc subjacent no especificada i $\beta_k=(\beta_{1k},\dots,\beta_{pk})'$ un vector d'ordre $p \times 1$ de paràmetres desconeguts de la regressió. Es tracta doncs, d'estimar β_k per màxima versemblança parcial per cada recurrència k .

Així es defineix $R_k(t)=\{i: T_{ik} \geq t\}$ com el conjunt d'individus que estan a risc just abans del moment t respecte la k -èssima recurrència i la funció de màxima versemblança parcial per la k -èssima recurrència com:

$$L_k(\beta) = \prod_{i=1}^n \left[\frac{\exp\{\beta' X_{ik}(T_{ik})\}}{\sum_{i \in R_k(X_{ik})} \exp\{\beta' X_{ik}(T_{ik})\}} \right]^{D_{ik}} \quad (2.4)$$

Així, l'estimador de màxima versemblança $\hat{\beta}_k$ dels paràmetres del model s'obté solucionant la següent equació:

$$\frac{\partial \log L_k(\beta)}{\partial \beta} = 0$$

Es pot veure que l'estimador $\hat{\beta}_k$ és consistent per β_k si el model (2.3) està especificat correctament. A més, Wei *et al.* (1989) demostren que el vector d'estimadors $(\hat{\beta}_1, \dots, \hat{\beta}_k)'$ on el subíndex indica la recurrència és aproximadament normal amb mitjana $(\beta_1, \dots, \beta_k)'$ i matriu de covariàncies Q que es pot estimar consistentment.

L'expressió de l'estimador de la matriu de covariàncies Q proporcionada per Wei *et al.* (1989) és molt complexa, en concret:

$$\hat{Q} = \frac{1}{n} \begin{bmatrix} \hat{D}_{11}(\hat{\beta}_1, \hat{\beta}_1) & \dots & \hat{D}_{1k}(\hat{\beta}_1, \hat{\beta}_k) \\ \dots & \dots & \dots \\ \hat{D}_{k1}(\hat{\beta}_k, \hat{\beta}_1) & \dots & \hat{D}_{kk}(\hat{\beta}_k, \hat{\beta}_k) \end{bmatrix} \quad (2.5)$$

essent,

$$\hat{D}_{kl}(\hat{\beta}_k, \hat{\beta}_l) = \hat{A}_k^{-1}(\hat{\beta}_k) \hat{B}_{kl}(\hat{\beta}_k, \hat{\beta}_l) \hat{A}_l^{-1}(\hat{\beta}_l) \quad (2.6)$$

on:

$$\hat{A}_k(\beta) = \frac{1}{n} \sum_{j=1}^n D_{jk} \left[\frac{\sum_{i=1}^n Y_{ik}(\tilde{T}_{jk}) X_{ik}(\tilde{T}_{jk})^{\otimes 2} e^{\beta' X_{ik}(\tilde{T}_{jk})}}{\sum_{i=1}^n Y_{ik}(\tilde{T}_{jk}) e^{\beta' X_{ik}(\tilde{T}_{jk})}} \right] \left[\frac{\sum_{i=1}^n Y_{ik}(\tilde{T}_{jk}) X_{ik}(\tilde{T}_{jk}) e^{\beta' X_{ik}(\tilde{T}_{jk})}}{\sum_{i=1}^n Y_{ik}(\tilde{T}_{jk}) e^{\beta' X_{ik}(\tilde{T}_{jk})}} \right]^{\otimes 2} \quad (2.7)$$

en la que $a^{\otimes 2}$ denota la matriu aa' construïda a partir del vector columna a .

$$\hat{B}_{kl}(\hat{\beta}_k, \hat{\beta}_l) = \frac{1}{n} \sum_{j=1}^n W_{jk}(\hat{\beta}_k) W_{jl}(\hat{\beta}_l)' \quad (2.8)$$

on,

$$W_k(\beta_k) = D_{kj} \left\{ Z_{kj}(X_{kj}) - \frac{S_k^{(1)}(\beta_k; X_{kj})}{S_k^{(0)}(\beta_k; X_{kj})} \right\} - \sum_{m=1}^n \frac{D_{km} Y_{kj}(X_{km}) e^{\beta Z_{kj}(X_{km})}}{n S_k^{(0)}(\beta_k; X_{km})} \times \left\{ Z_{kj}(X_{km}) - \frac{S_k^{(1)}(\beta_k; X_{km})}{S_k^{(0)}(\beta_k; X_{km})} \right\} \quad (2.9)$$

i,

$$S_k^{(1)}(\beta; t) = \frac{1}{n} \sum_{i=1}^n Y_{ik}(t) X_{ik}(t) e^{\beta X_{ik}(t)} \quad S_k^{(0)}(\beta; t) = \frac{1}{n} \sum_{i=1}^n Y_{ik}(t) e^{\beta X_{ik}(t)} \quad (2.10)$$

Therneau i Hamilton (1997), demostren que l'expressió d'aquesta matriu és totalment equivalent a la matriu:

$$Q = \begin{pmatrix} D'_1 D_1 & D'_1 D_2 & \dots & D'_1 D_k \\ D'_2 D_1 & D'_2 D_2 & \dots & D'_2 D_k \\ \dots & \dots & \dots & \dots \\ D'_k D_1 & D'_k D_2 & \dots & D'_k D_k \end{pmatrix} \quad (2.11)$$

on D_1 és la matriu dels residus β corresponents a l'ajust del model en la primera recurrència, D_2 els de l'ajust a la segona, etc.

Es pot concloure, doncs, que el mètode WLW és un model marginal respecte al conjunt de risc ja que l'individu està a risc des de l'inici de l'estudi i, per tant, no depèn dels esdeveniments anteriors. A més, és un mètode bastant bo per realitzar inferències poblacionals, és a dir, quan el que interessa són els efectes promig de les covariables sobre la variable resposta.

La limitació més important que presenta aquest mètode és que suposa que en el moment t tots els individus observats formen el conjunt de risc per la recurrència corresponent, independentment del nombre de recurrències prèvies que hagin patit (Pepe i Cai, 1993; Cook i Lawless, 1997; Lipschutz i Snapinn, 1997). Això no presentaria cap problema si es tractés de diferents esdeveniments; però si tenim recurrències del mateix esdeveniment podria passar que es dugués a terme alguna acció per reduir el nombre de recurrències i que, realment, les reduís. Com a conseqüència augmentaria la correlació entre els estimadors dels paràmetres

(Barceló, 2001). Aquesta limitació es pot superar, en part, utilitzant l'aproximació de Prentice, Williams i Peterson (Prentice *et al.*, 1981).

2.2.1.3. El mètode Prentice, Williams i Peterson.

El model Prentice, Williams i Peterson, mètode PWP (Prentice *et al.*, 1981), és un model marginal respecte l'estimació dels paràmetres, però condicional respecte la construcció del conjunt d'individus a risc. Aquest model permet que el risc basal variï entre diferents recurrències. Es tracta, doncs, d'un model de riscos proporcionals amb estrats dependents del temps, on la dependència entre recurrències es controla estratificant pel nombre de vegades que l'esdeveniment d'interès ha passat anteriorment.

Així, a diferència del mètode WLW, en el que tots els individus estan a risc en tot moment, el model PWP inclou en el conjunt de risc per la recurrència k-èsima, únicament els individus que han experimentat k-1 recurrències.

Prentice *et al.* (1981) proposen dos models semiparamètrics de riscos proporcionals:

- temps des de l'inici de l'estudi, on la funció de risc subjacent s'inclou com una funció del temps des de l'inici de l'estudi.

$$\lambda_{ik}(t) = \lambda_{0k}(t)e^{\beta_k X_{ik}(t)} \quad (2.12)$$

- temps des de la última recurrència, on la funció de risc subjacent és una funció del temps des de la recurrència anterior al temps de fallada.

$$\lambda_{ik}(t) = \lambda_{0k}(t - t_{n(t)})e^{\beta_k X_{ik}(t)} \quad (2.13)$$

S'observa que, en definitiva, es tracta simplement de models AG estratificats, degut a que tenen un conjunt de risc condicional i el risc basal és específic per cada recurrència (o esdeveniment diferenciat). Aquest model es diferencia de AG en dos aspectes (Barceló, 2001):

- el conjunt a risc per la (k+1)-èsima recurrència es limita a individus que han experimentat les k primeres recurrències.
- la funció de risc subjacent i els paràmetres de la regressió poden variar entre les diferents recurrències

El principal problema del model PWP es troba en la forma de construir els conjunts a risc, ja que comporta una pèrdua d'aleatorietat, donat que els individus amb més risc van sortint de l'estudi. Aquesta pèrdua serà més gran a mesura que ens anem movent a estrats superiors (i com més estrats es considerin). En aquests casos, la consistència de les estimacions depèn de la inclusió de totes les variables rellevants.

2.2.1.4. Cas pràctic.

A continuació s'exposa un senzill exemple per tal d'aclarir els mètodes exposats. Es suposen tres individus, un home que ha patit un episodi d'infecció; una dona que no n'ha presentat cap i un altre home que ha patit dos episodis d'infecció.

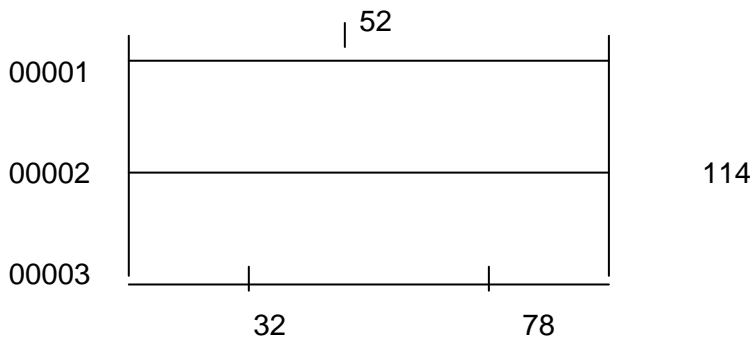
En el quadre i esquema següents es mostra el tracte que dona cada mètode a aquest conjunt de dades:

Taula 2. Exemple d'aplicació dels diferents mètodes.

<i>Model</i>	<i>Individu</i>	<i>T.entrada</i>	<i>T.sortida</i>	<i>Estatus</i>	<i>Numero</i>	<i>Sexe</i>
Andersen-	00001	0	52	1		Home
Gill	00001	52	114	0		Home
	00002	0	114	0		Dona
	00003	0	32	1		Home
	00003	32	78	1		Home
	00003	78	114	0		Home
Mètode	00001	0	52	1	1	Home
WLW	00001	0	114	0	2	Home
	00001	0	114	0	3	Home
	00002	0	114	0	1	Dona
	00002	0	114	0	2	Dona
	00002	0	114	0	3	Dona
	00003	0	32	1	1	Home
	00003	0	78	1	2	Home
	00003	0	114	0	3	Home
	00003	0	114	0	3	Home
Mètode PWP	00001	0	52	1	1	Home
	00001	52	114	0	2	Home
	00002	0	114	0	1	Dona
	00003	0	32	1	1	Home
	00003	32	78	1	2	Home
	00003	78	114	0	3	Home

T.entrada i T.sortida: dies que passen. Estatus: 0, sense infecció, 1, infecció; Numero: número de recurrència.

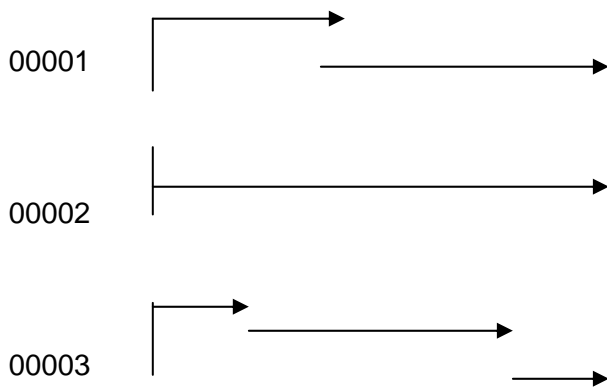
Figura 3. Situació inicial



Aquest disseny de dades és tractat inicialment respecte l'origen i les recurrències de forma diferenciada segons cadascun dels mètodes:

a) Model d'Andersen-Gill

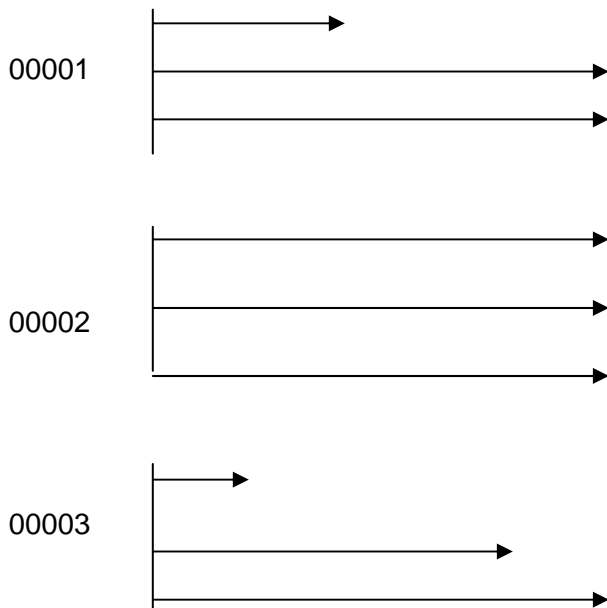
Figura 4. Tractament dels individus segons AG.



En aquest model els individus es divideixen en diferents registres segons els intervals d'observació i cada registre es tracta com si fos un individu diferent. Les variables explicatives són fixes dins de cada episodi però no dins de cada individu; així el risc que es compleixi la hipòtesi de riscos proporcionals és el risc de cada esdeveniment o recurrència per un individu determinat i aquest risc no està afectat per esdeveniments o recurrències anteriors del mateix individu.

b) Model de Wei, Lin i Weissfeld

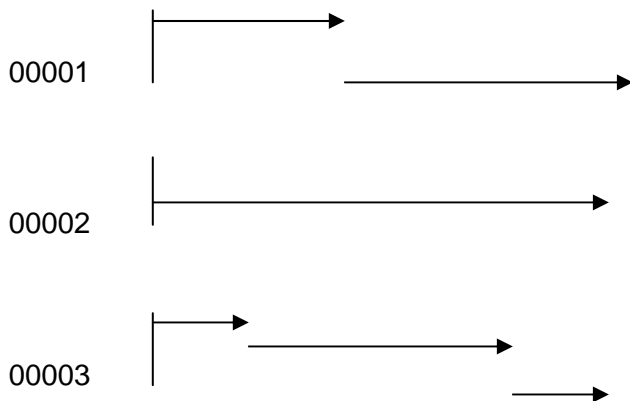
Figura 5. Tractament dels individus segons WLW.



En aquest model cada interval de temps comença en zero i així, tots els individus apareixen en tots els estrats i estan en risc des de l'inici de l'estudi, i per tant, no depenen d'esdeveniments previs.

c) Model de Prentice, Williams i Peterson

Figura 6. Tractament dels individus segons PWP.



Aquest model inclou en el conjunt a risc per la recurrència k només els individus que han experimentat $k-1$ recurrències. S'observa que es tracta de models AG estratificats.

2.2.2. Models condicionals.

En alguns casos, no només interessa l'efecte marginal de les covariables sobre la funció de risc dels temps de fallada dels individus, sinó també com els esdeveniments anteriors poden influir en el risc de desenvolupar futures fallades. Per respondre a aquestes qüestions necessitarem recórrer als models condicionals.

La hipòtesi d'increments independents en el model AG pot relaxar-se si s'inclou en el model variables dependents del temps, com el nombre de recurrències prèvies, per tal de capturar la possible estructura de dependència entre les recurrències. En concret, es podria treballar amb el model autoregressiu o markovià. Aquest model serà apropiat quan la dependència entre observacions sigui, principalment, "serial". No obstant això, tot i ser un model molt general, suposa que la dependència entre recurrències decau exponencialment i que la dependència és idèntica per tots els individus, la qual cosa pot representar una limitació.

Això es pot solucionar mitjançant els models de fragilitat. Es tracta d'introduir una (models de fragilitat univariant) o més (models de fragilitat multivariant) variables aleatòries comunes a tots els individus amb l'objectiu de permetre que els temps de supervivència multivariants, és a dir, les observacions repetides d'un mateix individu, estiguin correlacionades.

2.2.2.1 Models de fragilitat.

En els darrers anys ha esdevingut un fet important en la recerca la inclusió d'efectes aleatoris en els models de supervivència. En aquest sentit, un efecte aleatori és una variable continua que descriu l'excés de risc o fragilitat per diferents categories, com individus o grups d'individus. La idea és que els individus poden tenir diferent fragilitat, i que els individus amb menys fragilitat fallen abans.

Així, doncs, els models de fragilitat introdueixen una (*models de fragilitat univariant*) o més (*models de fragilitat multivariant*) variables aleatòries comunes a tots els individus amb l'objectiu de capturar dues fonts de variació diferents però relacionades. Una d'aquestes fonts sorgeix de covariables comunes a tots els individus, no observades, les quals, quan no es tenen en compte, generen dependència entre esdeveniments. L'altra font de variació prové de covariables individuals no observades i per tant, no incloses en l'estudi (ja sigui per circumstàncies pràctiques, o per no estar associades a

cap factor de risc de la supervivència de l'individu). Els efectes d'aquesta fonts serien totalment equivalents a la dependència serial citada en la introducció d'aquest apartat.

En els models de fragilitat univariants es pretén modelitzar la dependència entre els diferents temps de recurrència introduint un únic efecte aleatori multiplicatiu, comú a totes les observacions d'un mateix individu (i , per tant, invariant en el temps), i diferent per individus diferents. Així, la fragilitat es pot entendre com una variable aleatòria no observable i , per tant, no inclosa en el model, la qual és la causant de la dependència entre les recurrències.

S'han estudiat moltes alternatives per la distribució de probabilitat de la variable de fragilitat, com per exemple la distribució gamma, la distribució positiva estable i la distribució gaussiana inversa, entre d'altres. Però, el més utilitzat és un model de fragilitat amb distribució de probabilitat gamma de mitjana unitària i variança desconeguda.

En particular, el model AG de fragilitat gamma es pot expressar, en el seu cas més simple, de la següent manera:

$$\lambda_i(t) = \tilde{w}_i Y_i(t) \alpha_0(t) e^{\beta X_i(t)} \quad (2.14)$$

on $N=(N_i; i=1, \dots, n)$ és un procés comptador multivariant que compta el nombre d'esdeveniments que ha produït cada un dels n individus en el temps t . Associat a aquest procés comptador, existeix un procés d'intensitat λ amb components λ_i . El procés Y amb components Y_i és un procés no negatiu, observable i previsible, és a dir, el seu valor en qualsevol moment t és conegut just abans de t , indicant si un individu i s'observa en el conjunt a risc d'experimentar un determinat esdeveniment en temps t , λ_0 denota una funció de risc basal desconeguda i \tilde{w}_i (fragilitat) és un conjunt de variables aleatòries independents i idènticament distribuïdes segons una distribució de probabilitat gamma.

El model descrit és un model molt simple, ja que assumeix que la fragilitat és constant al llarg del temps, és a dir, la fragilitat mostraria en aquest cas diferències individuals presents al començament de l'estudi. A més a més, aquest model suposa que la fragilitat actua proporcionalment sobre la funció basal comú a tots els individus. Amb això es suposa que tots els individus segueixen un patró uniforme.

S'han suggerit diversos mètodes per estimar els paràmetres del model de regressió (β) i el paràmetre de fragilitat. El més important des d'un punt de vista pràctic és l'algoritme EM (Theerneau i Grambsch, 1998), que es basa en només dos passos.

Resumint, el primer pas consisteix en calcular l'esperança del logaritme de la funció de versemblança condicionada al conjunt de dades observades i a la fragilitat; i en el segon pas es maximitza l'expressió obtinguda en el pas anterior respecte la fragilitat.

Però, existeixen un conjunt de limitacions compartides per tots els models de fragilitat univariants, entre les quals cal destacar que suposen que la dependència entre observacions és sempre positiva, que no permeten la interacció entre les covariables observables i la fragilitat, i més important encara, que només permeten la presència d'un únic efecte aleatori, que a més és constant al llarg del temps.

Els models de fragilitat multivariants intenten solucionar algunes d'aquestes limitacions. En particular, aquests models permeten incorporar més d'un efecte aleatori al model introduint el terme $\tilde{w}^{(i)} = A_i^T \tilde{w}$ que és la fragilitat per a l'individu i . És a dir,

$(\tilde{w}^{(1)}, \dots, \tilde{w}^{(n)})^T = A\tilde{w}$, on A és la matriu de disseny per tot el grup.

Per estimar aquests models, i seguint a Petersen (1998), existeixen quatre aproximacions diferents començant per la més simple, que és la màxima versemblança paramètrica, i acabant per la més complexa numèricament, l'aproximació semiparamètrica utilitzant l'algoritme EM.

Barceló i Saez (2001) proposen un mètode d'especificació i estimació original que consisteix, de manera molt resumida, en introduir un nou pas en l'algoritme EM, modelitzant la variància de la fragilitat directament. Mitjançant l'algoritme així modificat solucionen part de les limitacions citades anteriorment; en particular, permeten que la fragilitat no sigui constant entre individus.

Referències

Aalen O. (1978): Non-parametric inference for a family of counting processes. *The Annals of Statistics*, 6 (4), 701-726.

Arjas E. i Haara PA. (1984): A marked point process approach to censored failure data with complicated covariates. *Scandinavian journal of statistics*, 11, 193-209.

Andersen PK. i Gill RD. (1982): Cox's Regression Model for Counting Processes: A Large Sample Study. *Annals of Statistics*, 10(4), 1100-1120.

Andersen PK. i Borgan O. (1985): Counting process models for life history data (with discussion). *Scandinavian journal of statistics*, 12: 97-158.

Andersen PK., Borgan O., Gill RD., i Keiding N. (1993). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York.

Barceló MA. (2001): *Modelos de fragilidad en el análisis multivariante de supervivencia* [tesis doctoral]. Girona: Universitat de Girona.

Barceló MA. i Sáez M. (2001): A modifications of the EM algorithm to estimate an Andersen-Gill gamma frailty model for multivariate failure time data. Mimeo, 2001.

Bremaud P. i Jacod J. (1977): Processus ponctuels et martingales: résultats récents sur la modélisation et filtrage. *Advances in Applied Probability*, 9(2), 362-416.

Bremaud P. (1981). *Point Processes and Queues, Martingale Dynamics*. Springer-Verlag, New York.

Cook RJ. i Lawless JF. (1997): Marginal analysis of recurrent events and a terminating event. *Statistics in Medicine*, 16(88), 911-924.

Cox DR. (1972): Regression models and life tables. *Journal of the Royal Statistical Society, Series B (Methodological)*, 34 (2), 187-220.

Cox DR. (1975): Partial likelihood. *Biometrika*, 62 (2), 269-276.

Fleming TR. i Harrington DP. (1991): *Counting Processes and Survival Analysis*. New York, John Wiley & Sons, Inc.

Gill RD. (1984): Understanding Cox's regression model: a martingale approach. *Journal of the American Statistical Association*, 79, 441-447.

Jacod J. (1975): Multivariate point processes: predictable projection, Radon-Nikodym derivatives, representation of martingales. *Probability Theory and Related Fields*, 31,235-253.

Jacod J. (1979): *Calcul Stochastique et Problèmes de Martingales*. Springer-Verlag, Berlin.

Lipschutz KH. i Snapinn SM. (1997): Discussion of paper by Wei and Glidden. *Statistics in Medicine* 16, 846-848.

Pepe M. i Cai J. (1993): Some graphical displays and marginal regression analysis of recurrent failure times and time dependent covariates. *Journal of the American Statistical Association* , 88, 811-820.

Petersen JH. (1998): An additive frailty model for correlated life times. *Biometrics*, 54(2), 646-661.

Prentice RL., Williams BJ. i Peterson AL. (1981): On the regression analysis of multivariate failure time data. *Biometrika*, 68(2), 373-379.

Prentice RL. i Self SG. (1983): Asymptotic distribution theory for Cox-type regression models with general relative risk form. *The Annals of Statistics*, 11(3),804-813.

Therneau TM. i Hamilton SA. (1997): rhDNase as an example of recurrent event analysis. *Statistics in Medicine*, 16(18), 2029-2047.

Therneau TM. i Grambsch PM. (1998): *Penalized Cox Models*. Mayo Clinic.

Wei LJ., Lin DY. i Weissfeld L. (1989): Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *JASA*, 84(408),1065-1073.

3. Models lineals generalitzats i Models lineals generalitzats mixtes.

3.1. Models lineals generalitzats.

Com indica el seu nom, els models lineals generalitzats, GLM (McCullagh i Nelder, 1989), generalitzen el model de regressió lineal i permeten utilitzar el mateix tipus de modelització, especificació, estimació i diagnòstic per variables dependents no normals, com per exemple: variables comptadores (distribució de Poisson); variables dicotòmiques (distribució binomial), etc (veure Saez i Saurina, 2007).

L'objectiu dels GLM és descriure la dependència de la variable resposta (o variable dependent), Y , respecte les variables explicatives.

$$E(Y|x) = \mu$$

Les variables resposta dels models GLM pertanyen a la família exponencial amb funció de densitat de la forma:

$$f(y) = e^{\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right]}$$

on θ és un paràmetre de localització (no necessàriament la mitjana) i ϕ un paràmetre de dispersió (només apareix en distribucions que es caracteritzen per dos paràmetres, com per exemple la normal).

La forma de les funcions $a()$, $b()$ i $c()$ és diferent segons la distribució considerada.

Taula 3. Forma de les funcions $a()$, $b()$ i $c()$ segons tipus de distribució.

Distribució	$a(\phi)$	$b(\theta)$	$c(y)$
Bernouilli	ϕ	$\ln(1 + e^\theta)$	1
Binomial	ϕ/n	$\ln(1 + e^\theta)$	$\ln[n!/((ny)!(n-ny)!)]$
Poisson	ϕ	e^θ	$-\ln(y!)$
Poisson amb offset	ϕ/n	e^θ	$yn \ln(yn) - \ln(yn!)$

El logaritme de la funció de versemblança s'expressa com:

$$l(\theta; \Phi; y) = \log f_y(y; \theta, \Phi) = [y\theta - b(\theta)]/a(\Phi) + c(y, \Phi)$$

Es verifica que:

$$\frac{\partial l}{\partial \theta} = [y - b'(\theta)]/a(\Phi) \quad \frac{\partial^2 l}{\partial \theta^2} = -b''(\theta)/a(\Phi)$$

i, segons la teoria de la versemblança,

$1) E_{\theta} \left(\frac{\partial l}{\partial \theta} \right) = 0 \quad 2) E_{\theta} \left(\frac{\partial^2 l}{\partial \theta^2} \right) + \text{var}_{\theta} \left(\frac{\partial l}{\partial \theta} \right) = 0$ <p style="text-align: center;">on $\text{var}_{\theta} \left(\frac{\partial l}{\partial \theta} \right) = E_{\theta} \left(\frac{\partial l}{\partial \theta} \right)^2$</p>

Així, doncs,

$$1) E_{\theta} \left(\frac{\partial l}{\partial \theta} \right) = [\mu - b'(\theta)]/a(\Phi) = 0 \Rightarrow E(Y) = \mu = b'(\theta)$$

2)

$$-\frac{b''(\theta)}{a(\Phi)} + E \left[\frac{Y - b'(\theta)}{a(\Phi)} \right]^2 = -\frac{b''(\theta)}{a(\Phi)} + \frac{E[Y - E(Y)]^2}{a^2(\Phi)} = -\frac{b''(\theta)}{a(\Phi)} + \frac{\text{Var}(Y)}{a^2(\Phi)} = 0 \Rightarrow \text{Var}(Y) = b''(\theta)a(\Phi)$$

Per tant, la mitjana, μ , no és més que la primera derivada de la funció $b(\theta)$ ($\mu=b'(\theta)$), mentre que la variància de Y és proporcional a la segona derivada de $b(\theta)$ ($\text{var}(Y)=a(\phi)b''(\theta)$); ϕ s'anomena paràmetre de dispersió.

Els resultats obtinguts per la mitjana i la variància de les diferents distribucions prenen la forma:

Taula 4. Forma mitjana i variància de les diferents distribucions.

Distribució	Mitjana $\mu=b'(\theta)$	Variància en funció de $\theta, a(\phi)b(\theta)$	Variància en funció de μ
Bernouilli	$\frac{1}{1+e^{-\theta}}$	$\phi \frac{e^{\theta}}{(1+e^{\theta})^2}$	$\phi\mu(1-\mu)$
Binomial	$\frac{1}{1+e^{-\theta}}$	$\frac{\phi}{n} \frac{e^{\theta}}{(1+e^{\theta})^2}$	$\phi\mu(1-\mu)/n$
Poisson	e^{θ}	ϕe^{θ}	$\phi\mu$
Poisson con offset	e^{θ}	$\frac{\phi}{n} e^{\theta}$	$\phi\mu/n$

Així, doncs, l'especificació del model GLM pot expressar-se com:

$$\begin{aligned}
 E(Y| x) &= \mu \\
 g(\mu) &= \beta_0 + \sum_i \beta_i X_i \quad (i=1, \dots, n) \\
 \text{Var}(Y| x) &= \phi v(\mu)
 \end{aligned}
 \tag{3.1}$$

La funció $g()$ s'anomena 'vinçle' (link), i la funció $v()$ s'anomena 'variància' i ϕ és un paràmetre d'escala (o de dispersió).

L'aportació dels models GLM s'entén millor quan es comparen amb els models de regressió lineal clàssics.

En el quadre resum que es mostra a continuació s'observa com la funció vinçle manté l'additivitat¹ dels efectes, aplicant un canvi d'escala quan és necessari.

¹ L'additivitat és una propietat de les respostes esperades $E[Y|X]$ ($\hat{\mu} = \sum x_j \hat{\beta}_j$).

Taula 5. Comparació model lineal clàssic i model lineal generalitzat

Model lineal clàssic	Model lineal generalitzat
<p>Part sistemàtica</p> $E(y_i) = \mu_i = \sum x_{ij} \beta_j \quad i = 1, \dots, n$ <p>$\mu = X\beta$ $\mu = n \times 1$ $X = n \times p$ $\beta = p \times 1$</p> <p>X.- matriu model</p> <p>β.- vector de paràmetres</p> <p>Part aleatòria</p> <p>Es suposa independència i variància constant dels errors</p> $\varepsilon \sim N(0, \sigma^2)$ <p>Resum</p> $Y \sim N(0, \sigma^2)$ <p>$\mu = E(Y) = X\beta$</p> <p>Y independents</p>	<p>1.-Part sistemàtica</p> $\eta_i = \sum x_{ij} \beta_j \quad i = 1, \dots, n$ <p>2.-Part aleatòria</p> $Y \sim N(0, \sigma^2) \quad Y \text{ independents}$ <p>3.-Predictor lineal η $\mu = \eta$</p> <p>Generalització</p> <p>La distribució normal suposada a 2) per Y pot ser qualsevol distribució de la família exponencial</p> <p>Si s'escriu $\eta_i = g(\mu_i)$ es té $g(\cdot) =$ funció link, que pot ser qualsevol funció monòtona diferenciable.</p>

Font: Saez i Saurina, 2007.

Per exemple, quan Y és una variable comptadora i, per tant, està generada per una distribució de Poisson, la funció vincle és el logaritme, $g(\mu) = \ln(\mu)$; la variància, $v(\cdot)$, és la funció identitat i el paràmetre de dispersió, ϕ , és igual a 1.

$$E(Y | x) = \mu$$

$$\ln(\mu) = \beta_0 + \sum \beta_i X_i \quad (i=1, \dots, n)$$

$$\text{Var}(Y | x) = \mu$$

Sovint, que la mitjana sigui igual a la variància ($\phi=1$) és una hipòtesi massa restrictiva. Per això, es permet que existeixi sobre-dispersió ($\phi>1$) o, menys habitual, sub-dispersió ($\phi<1$). En aquest cas, la variància s'especificarà com:

$$\text{Var}(Y | x) = \phi\mu$$

Per exemple, en el cas d'una resposta dicotòmica, distribució binomial, el més habitual és que existeixi una sub-dispersió:

$$E(Y | x) = \mu$$

$$\text{Ln}\left(\frac{\mu}{1-\mu}\right) = \beta_0 + \sum_i \beta_i X_i \quad (i = 1, \dots, n)$$

$$\text{Var}(Y | x) = \phi\mu(1-\mu)$$

Els model GLM es poden estimar per màxima versemblança utilitzant la forma general de la funció de densitat per les distribucions més habituals, corresponents a una distribució exponencial d'un únic paràmetre (en aquest cas a és una constant):

$$f(Y; \theta) = e^{\left[\frac{y\theta - b(\theta)}{a} + c(Y) \right]}$$

El logaritme de la funció de versemblança per un conjunt de n observacions es pot expressar com (i denota la observació):

$$\log L = \frac{\sum_i (y_i \theta_i - b(\theta_i))}{a_i} + k$$

o en notació matricial:

$$\log L = Y' A^{-1} \theta - b(\theta)^{1/2} A^{-1} b(\theta)^{1/2} + k$$

on $A = \text{diag}(a_i)$, $\theta = (\theta_1, \theta_2, \dots, \theta_n)'$; $b(\theta) = (b(\theta_1), b(\theta_2), \dots, b(\theta_n))'$; k una constant; i $\theta = X\beta$.

Però, degut a que la derivada de $\log L$ respecte β no és lineal, el logaritme de la funció de versemblança no es pot maximitzar directament. Alternativament, es pot utilitzar una de les dues aproximacions següents:

1.- A partir de l'algoritme de Newton-Raphson. L'estimador del vector de paràmetres β a la iteració $(s+1)$ es pot expressar com:

$$\beta_{s+1} = \beta_s + I^{-1}(\beta_s) U(\beta_s)$$

on $U(\beta)$ és el vector de les primeres derivades o de scores, és a dir, $U(\beta)=\partial\log L/\partial\beta$ i $I(\beta)$ la matriu $I(\beta)=-H(\beta)$; on $H(\beta)$ és el Hessià o matriu de segones derivades, $H(\beta)=\partial^2\log L/\partial\beta\partial\beta'$.

En cada iteració, $s+1=1,2,\dots$, tant el vector de scores com la matriu d'informació s'avaluen en s , o sigui, en la iteració anterior. Aquest procés iteratiu acaba quan el canvi en el logaritme de la funció de versemblança és prou petit, o bé, quan ho sigui el més gran dels canvis relatius en els estimadors dels paràmetres. La variància dels paràmetres es pot calcular en la última iteració, $\text{Var}(\beta)=(B'X'V^{-1}XB)^{-1}$, on $B=\text{diag}(g'(1(\mu)))$ i $V=\text{Var}(Y|x)$.

2.- Alternativament es pot utilitzar el mètode dels mínims quadrats iteratius (IRLS) (Nelder i Wedderburn, 1972).

Partint de la funció vinkle $\eta=g(\mu) = X\beta$, es defineix una pseudo-variable, z , com el desenvolupament de primer ordre de Taylor de $g(Y)$ entorn μ tal que:

$$g(Y)\approx g(\mu)+(y - \mu) g'(\mu) = \eta + (y-\mu) d \eta/d\mu$$

$$z_0 = g(\mu_0)+(Y - \mu_0) g'(\mu_0) = g(\mu_0) + (Y - \mu_0)B^{-1} \quad B=\text{diag}(g''(\mu))$$

z és la descomposició en sèries de Taylor de primer ordre de $g(Y)$ al voltant de μ .

Llavors, es tracta d'estimar β i la seva variància, $\text{Var}(\beta)$, en el model de regressió:

$$z = X\beta + e$$

Com que $\eta=g(\mu) = X\beta$, i η i μ són valors fixats i coneguts, la variància de z és igual a:

$$Vz = \text{Var}\left[\eta + (y - \mu) \frac{\partial\eta}{\partial\mu}\right] = \left(\frac{\partial\eta}{\partial\mu}\right)^2 \text{Var}y = \text{Var}(X\beta) + B^{-1}\text{Var}(Y-\mu)B^{-1} = 0 + B^{-1}ABB^{-1} = AB^{-1}$$

L'algoritme IRLS consisteix en:

1. Prendre Y com a valors inicials de z .

2. En la primera iteració $\beta=(X'V_z^{-1}X)^{-1}X'V_z^{-1}z$.
3. El valor de β s'utilitza per calcular nous valors de z i de V_z .
4. Repetir el procés d'iteració fins que convergeixi.

La variància asimptòtica de β es pot obtenir en la última iteració com,

$$\text{Var}(\beta)=(X' V_z X)^{-1}$$

Per els links canònics ($\eta=\theta$), el valor esperat i l'actual a la matriu hessiana coincideixen i, per tant, el mètode desenvolupat per IRLS (Fisher scoring method) coincideix amb l'algoritme de Newton Raphson.

3.2. Models lineals generalitzats mixtes.

En l'anàlisi economètric aplicat són importants els dissenys mixtes caracteritzats per considerar a la vegada dues o més dimensions d'anàlisi (Saez 2001). Aquests dissenys comprenen els dissenys multinivell o jeràrquics i els models longitudinals o de mesures repetides. Els dissenys multinivell tenen una estructura jeràrquica amb agrupacions de dades en grups o clústers. Per altra banda, els estudis longitudinals es caracteritzen per mesurar repetidament al llarg del temps l'individu d'estudi. Per tant, es pot considerar l'estructura com mixta, amb les observacions (repetides) agrupades dins de cada individu i el temps es pot considerar com una variable explicativa més dins de cada grup (Saez, 2001).

L'avantatge més important dels estudis longitudinals respecte altres dissenys, és que permet distingir la variació entre individus (variació interindividual) de la variació dins d'un mateix individu (variació intraindividual).

Començant amb una resposta lineal, els estudis longitudinals es poden especificar com:

$$Y_{ik} = \beta_1 X_{1ik} + \beta_2 X_{2ik} + \dots + \beta_p X_{pik} + \varepsilon_{ik} \quad (3.2)$$

El subíndex $i=1,\dots,n$ denota els individus i el subíndex $k=1,\dots,K$ les diferents observacions (repetides) per cada individu (observem que el nombre de períodes, K , no té perquè ser el mateix per a tots els individus).

Es denota per Y_{ik} les observacions de la variable dependent, o resposta. Es considera que la resposta de cada individu i està determinada per p variables explicatives observades en el moment k , X_{jik} . Algunes de les variables explicatives són fixes ($X_{jik} = X_{ji} \forall k$) i altres variables amb el temps. Els β són p paràmetres desconeguts. Finalment, ε_{ik} és un terme de pertorbació amb mitjana nul·la.

Cal suposar que:

$$\begin{aligned} E(\varepsilon_{ik}) &= 0 \\ E(\varepsilon_{ik}^2) &= \sigma_{\varepsilon}^2 \text{ constant} \\ E(\varepsilon_{ik}\varepsilon_{js}) &= 0 \quad \forall k \neq s \text{ y } \forall i \neq j \end{aligned} \quad (3.3)$$

De forma simplificada, es considera cinc individus $i=1,2,3,4,5$; tres períodes $k=1,2,3$; i dues variables explicatives ($p=2$), X_1 (fixa) i X_2 (dependent del temps). Llavors, el model s'especificaria com:

$$Y_{ik} = \beta_0 + \beta_1 X_{1ik} + \beta_2 X_{2ik} + \varepsilon_{ik} \quad (3.4)$$

En notació matricial s'escriuria:

a) Per cada individu:

$$Y_i = X_i \beta + \varepsilon_i$$

$$\text{on } Y_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \end{bmatrix}, i=1,2,\dots,5; X_i = \begin{bmatrix} 1 & X_{1i1} & X_{2i1} \\ 1 & X_{1i2} & X_{2i2} \\ 1 & X_{1i3} & X_{2i3} \end{bmatrix}; \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}; \text{ i } \varepsilon_i = \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \end{bmatrix}.$$

La mitjana i la variància de la variable dependent es poden formular com:

$$\begin{aligned} E(y_{ik}) &= \mu_i & E(Y_i) &= \mu_i \\ \text{Var}(y_{ik}) &= \sigma_{\varepsilon_i}^2 & \text{Var}(Y_i) &= V_i \end{aligned}$$

$$\text{on } \sigma_{\varepsilon_i}^2 \text{ és una constant per cada individu } i; \text{ i (en aquest cas) } V_i = \begin{bmatrix} \sigma_{\varepsilon_i}^2 & 0 & 0 \\ 0 & \sigma_{\varepsilon_i}^2 & 0 \\ 0 & 0 & \sigma_{\varepsilon_i}^2 \end{bmatrix}.$$

b) Per tota la mostra:

$$Y = X\beta + \varepsilon$$

$$\text{on } Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{bmatrix}; X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{bmatrix}; \text{ i } \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{bmatrix}.$$

Ara:

$$E(Y) = \mu \\ \text{Var}(Y) = V$$

$$\text{essent } V = \begin{bmatrix} V_1 & 0 & 0 & 0 & 0 \\ 0 & V_2 & 0 & 0 & 0 \\ 0 & 0 & V_3 & 0 & 0 \\ 0 & 0 & 0 & V_4 & 0 \\ 0 & 0 & 0 & 0 & V_5 \end{bmatrix}.$$

És important assenyalar que la matriu V ha de ser sempre diagonal en blocs, és a dir, que s'assumeix que els individus (o grups) són independents entre si.

Segons quin sigui l'objectiu, els estudis longitudinals es poden aproximar marginal o condicionalment. S'utilitza l'enfoc marginal quan es vol realitzar inferències poblacionals, és a dir, si es vol explicar la relació entre la variable dependent i les variables explicatives amb independència de la variabilitat intraindividual.

Per exemple, si es recupera l'expressió (3.4);

$$Y_{ik} = \beta_0 + \beta_1 X_{1ik} + \beta_2 X_{2ik} + \varepsilon_{ik}$$

l'ordenada a l'origen (β_0) i els coeficients associats a les variables explicatives (β_1 i β_2) són comuns a tots els individus. No existeix heterogeneïtat individual o, d'una altra forma, tots els efectes (de les variables explicatives, inclosa l'ordenada a l'origen) són fixes. Es tracta d'estimar els paràmetres corresponents a la mitjana, β . Poques

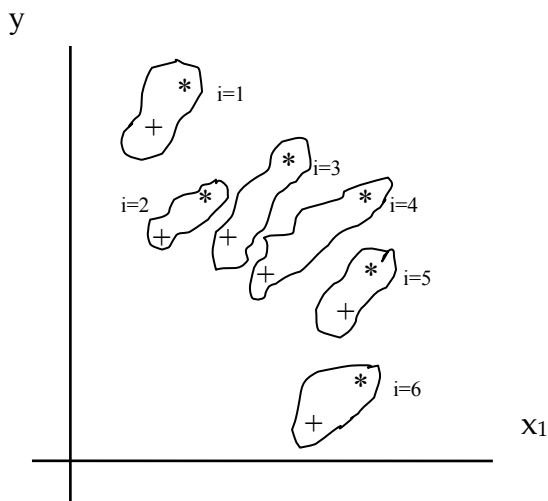
vegades són d'interès els paràmetres de la covariància (la variància i , en el cas de no ser nul·la, les covariàncies de ϵ). De fet aquests es solen tractar com una molèstia, controlats, però no estimats en l'enfoc marginal.

L'enfoc condicional pretén realitzar inferències individuals. Per això, modelitza simultàniament, la mitjana de la variable dependent (la variabilitat interindividual), i l'estructura de covariàncies o correlacions (la variabilitat intraindividual). En aquest enfoc, els paràmetres que defineixen la correlació tenen el mateix, i a vegades fins i tot més, interès que els corresponents a la mitjana.

Els efectes aleatoris és l'enfoc més conegut dins dels condicionals. Suposa que la dependència que pot existir entre diferents observacions de la variable dependent, és a dir, entre les respostes repetides d'un mateix individu, és deguda a que els coeficients de la regressió (de la mitjana) no són els mateixos per a tots els individus. De fet, la interpretació més senzilla, és la que suposa que l'heterogeneïtat individual és deguda a factors no observables (o variables omeses) fixes en el temps però variables entre individus. Aquests factors són els que provoquen la dependència entre les diferents observacions de la variable dependent.

Així, per exemple, es suposa que es tenen sis individus $i=1, \dots, 6$ i dos períodes per cada individu $k=1, 2$ (amb una $+$ es representa el primer i amb $*$ el segon),

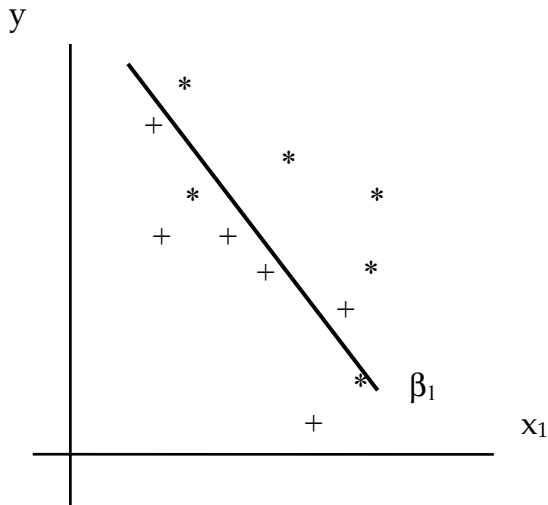
Gràfic 1. Exemple.



Un model d'efectes fixes (enfoc marginal) s'especificaria com:

$$Y_{ik} = \beta_0 + \beta_1 X_{ik} + \varepsilon_{ik} \quad (3.5)$$

Gràfic 2. Exemple.

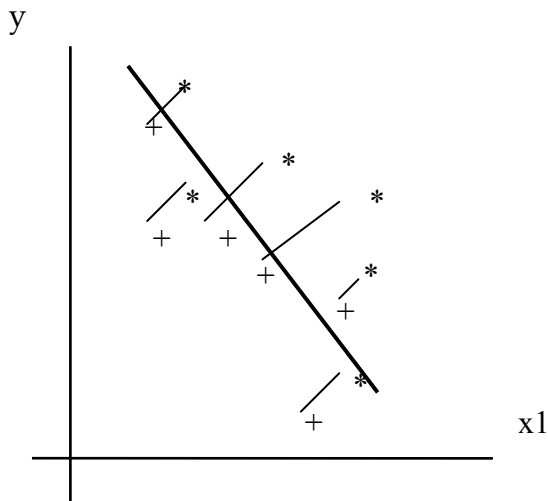


Un model d'efectes aleatoris (enfoc condicional) s'especificaria com:

$$Y_{ik} = \beta_{0i} + \beta_1 X_{ik} + \varepsilon_{ik} \quad (3.6)$$

L'ordenada a l'origen, β_{0i} , és diferent per a cada individu. No obstant això, l'efecte de X_1 és fix, aproximadament el mateix, per a tots els individus.

Gràfic 3. Exemple.



En el següent model, X_1 és l'efecte aleatori:

$$Y_{ik} = \beta_0 + \beta_{pi} X_{ik} + \varepsilon_{ik} \quad (3.7)$$

Aquest tipus de model, en el qual el paràmetre està associat a una variable continua s'anomena també de coeficient aleatoris. Naturalment, el model pot combinar efectes i coeficients aleatoris.

Resumint, els models d'efectes aleatoris intenten modelitzar l'esperança condicional de Y_{ik} , donats els coeficients específics dels individus, β_i

$$Y_{ik} = \beta_i x_{ik} + \varepsilon_{ik}$$

El problema és que, normalment, per a cada individu existeixen molt poques observacions per estimar β_i a partir de (y_i, x_i) . Per això, es suposa que els β_i només són realitzacions independents d'alguna distribució de probabilitat amb mitjana β :

$$\begin{aligned} \beta_i &= \beta + u_i \\ Y_{ik} &= \beta X_{ik} + (\varepsilon_{ik} + x_{ik} u_i) \end{aligned} \quad (3.8)$$

Així, a més de:

$$\begin{aligned} E(\varepsilon_{ik}) &= 0 \\ E(\varepsilon_{ik}^2) &= \sigma_\varepsilon^2 \text{ constant} \\ E(\varepsilon_{ik} \varepsilon_{js}) &= 0 \quad \forall k \neq s \text{ y } \forall i \neq j \end{aligned}$$

cal suposar també:

$$\begin{aligned} E(u_i) &= 0 \\ E(u_{ij}^2) &= \sigma_u^2 \text{ constant} \\ E(\varepsilon_{ik} u_j) &= 0 \quad \forall i, k, j \\ E(u_i u_j) &= 0 \quad \forall i \neq j \end{aligned}$$

Les dues últimes hipòtesis són particularment importants. Els efectes aleatoris han de ser independents dels fixes i els individus han de ser independents entre sí.

Així, doncs, generalitzant, i seguint amb la terminologia anterior, un model lineal generalitzat mixt, GLMM, es pot expressar com:

$$\begin{aligned}
E(Y_{ik} | X_{ik}, u_{ik}) &= \mu_{ik} \\
g(\mu_{ik}) &= X_{ik}\alpha + Z_{ik}'\beta_i \\
\text{Var}(Y_{ik} | X_{ik}, u_{ik}) &= \phi v(\mu_{ik})
\end{aligned}
\tag{3.9}$$

on α són els efectes fixes (x les variables associades) i β_i els efectes aleatoris (z les variables associades); essent habitual que els efectes aleatoris es distribueixin normalment (multivariant) amb mitjana nul·la i matriu de covariàncies G, és a dir, $\beta_i \sim N(0, G)$.

En termes matricials:

$$g(\mu) = X\alpha + Z\beta$$

la variància $V = \text{Var}(Y)$,

$$V = \text{Var}(\mu) + R$$

En particular, quan la variable dependent és continua, el model mixta es pot especificar en termes matricials com:

$$\begin{aligned}
\mu &= X\alpha + Z\beta \\
\text{Var}(\mu) &= \varphi
\end{aligned}$$

L'estructura de G, la matriu de covariàncies dels efectes aleatoris pot ser:

a) Identitat o uniforme:

$$\begin{bmatrix}
\sigma_c^2 & 0 & 0 \\
0 & \sigma_c^2 & 0 \\
0 & 0 & \sigma_c^2
\end{bmatrix}$$

b) No estructurada o matriu general definida positiva:

$$\begin{bmatrix}
\sigma_1^2 & \sigma_{12} & \sigma_{13} \\
\sigma_{12} & \sigma_2^2 & \sigma_{23} \\
\sigma_{13} & \sigma_{23} & \sigma_3^2
\end{bmatrix}$$

c) Diagonal:

$$\begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{bmatrix}$$

d) Simetria composta:

$$\begin{bmatrix} \sigma^2 & \theta & \theta \\ \theta & \sigma^2 & \theta \\ \theta & \theta & \sigma^2 \end{bmatrix}$$

e) Autoregressiva de primer ordre :

$$\begin{bmatrix} \sigma^2 & \rho & \rho^2 \\ \rho & \sigma^2 & \rho \\ \rho^2 & \rho & \sigma^2 \end{bmatrix}$$

f) Toeplitz (AR(p) con $p > 1$):

$$\begin{bmatrix} \sigma^2 & \sigma_1 & \sigma_2 \\ \sigma_1 & \sigma^2 & \sigma_1 \\ \sigma_2 & \sigma_1 & \sigma^2 \end{bmatrix}$$

g) Espacial:

$$\begin{bmatrix} \sigma^2 & \rho^{d_{12}} & \rho^{d_{13}} \\ \rho^{d_{12}} & \sigma^2 & \rho^{d_{23}} \\ \rho^{d_{13}} & \rho^{d_{23}} & \sigma^2 \end{bmatrix}$$

on $d_{ij} = |t_i - t_j|$ i t_i, t_j són els períodes associats a i i j , respectivament.

La matriu de covariàncies residual, R , pot adoptar també diverses estructures: identitat, autoregressiva, exponencial (autoregressiva general en el sentit que els intervals de temps no necessiten ser iguals), espacial exponencial, espacial gaussiana, espacial lineal, de simetria composta, Toeplitz.

Mitjana mòbil:

$$MA(1) \text{ per exemple } \begin{bmatrix} \sigma^2 & \gamma & 0 \\ \gamma & \sigma^2 & \gamma \\ 0 & \gamma & \sigma^2 \end{bmatrix}$$

ARMA:

$$ARMA(1,1) \text{ per exemple } \begin{bmatrix} \sigma^2 & \gamma_1 & \gamma_2 \\ \gamma_1 & \sigma^2 & \gamma_1 \\ \gamma_2 & \gamma_1 & \sigma^2 \end{bmatrix}$$

Com era d'esperar, la variància de la variable dependent, la dels efectes aleatoris i la variància residual estan relacionades (essent X_i i Z_i no aleatòries i β i ε_i independents):

$$V_i = \text{Var}(Y_i) = \text{Var}(X_i\alpha) + \text{Var}(Z_i\beta) + \text{Var}(\varepsilon_i) = 0 + Z_i\text{Var}(\beta)Z_i' + \text{Var}(\varepsilon_i) = Z_iGZ_i' + R$$

Amb la qual cosa, a partir de G i/o R no resulta difícil derivar l'estructura de la matriu V_i . Finalment, tant l'estructura de G com la de R (i per tant la de V_i) poden estar organitzades per blocs d'individus i/o de períodes de temps.

Però, quan la variable dependent no és continua (és a dir, és comptadora, dicotòmica, etc.), μ no és una funció lineal de β , i per tant, V no es deriva tan fàcilment a partir de G i R , com en el cas continu. En aquest cas, es pot utilitzar una aproximació basada en sèries de Taylor de primer ordre.

$$V \approx BZGZ'B + R$$

on B és una matriu diagonal de variàncies determinades per la funció vincle. Per exemple en una Poisson, $B = \text{diag}(\mu_i)$ ($B = \text{diag}(\mu_i/t_i)$ en una Poisson amb offset), en una binomial, $B = \text{diag}(\mu_i(1 - \mu_i))$.

La matriu R pot ser la identitat (no existeix autocorrelació) i llavors $R=AB$; o bé, pot existir autocorrelació i llavors $R=A^{1/2}B^{1/2}R_sB^{1/2}A^{1/2}$, on R_s és una de les matrius R definides més amunt.

Només existeix una funció de versemblança completa quan $R=I$ (és a dir, quan no existeix autocorrelació), existeixin o no efectes aleatoris. Llavors, es pot escriure:

$$L(\alpha, \beta, \gamma_G; Y) = L(\alpha; Y | \beta) L(\gamma_G; \beta)$$

on γ_G representen els coeficients de G .

Suposant una distribució normal per els efectes, $\beta \sim N(0, G)$,

$$L(\alpha, \beta, \gamma_G; Y) \propto L(\alpha, Y | \beta) |G|^{-1/2} e^{-\frac{1}{2} \beta' G^{-1} \beta}$$

Com que Y i β són independents ($R=I$), podem utilitzar l'expressió del GLM estàndard,

$$L(\alpha; Y | \beta) = e^{Y' A^{-1} \theta - b(\theta)^{1/2} A^{-1} b(\theta)^{1/2} + K} \propto e^{Y' A^{-1} \theta - b(\theta)^{1/2} A^{-1} b(\theta)^{1/2}}$$

on $\theta = X\alpha + Z\beta$, $A = \text{diag}(a_i)$ i K és una constant.

La versemblança completa es pot expressar com:

$$L(\alpha, \beta, \gamma_G; Y) \propto e^{Y' A^{-1} \theta - b(\theta)^{1/2} A^{-1} b(\theta)^{1/2}} e^{-\frac{1}{2} |G|^{-1/2} \beta' G^{-1} \beta}$$

i el logaritme de la funció de versemblança:

$$\text{Ln} L(\alpha, \beta, \gamma_G; Y) \propto Y' A^{-1} \theta - b(\theta)^{1/2} A^{-1} b(\theta)^{1/2} - \frac{1}{2} |G|^{-1/2} \beta' G^{-1} \beta + K$$

En una Poisson, $a_i=1$ i $b(\theta)=\exp(\theta)$ (en una Poisson amb offset, $a_i=1/t_i$ i $b(\theta)=\exp(\theta)$).

En una binomial, $a_i=1/n_i$ i $b(\theta)=\text{Ln}(1+\exp(\theta))$.

Quan R no és la matriu identitat, és a dir, existeix autocorrelació, no existeix la versemblança completa. No obstant això, es pot definir la quasi-versemblança:

$$QL(\alpha, \beta, \gamma; Y) \propto \text{Ln}(f(\alpha, \gamma; Y | \beta)) - \frac{1}{2} |G|^{-1/2} \beta' G^{-1} \beta + K$$

Essent $\gamma=(\gamma_G, \gamma_R)$ i $f(\alpha, \gamma; R; Y)$ una funció amb propietats similars al logaritme de la funció de versemblança, i es defineix com,

$$\frac{\partial \ln(f(\alpha, \gamma; R; Y))}{\partial \alpha} = X' B R^{-1} (Y - \mu)$$

que té exactament la mateixa forma que la derivada parcial de la verdadera versemblança.

La maximització de la quasi-versemblança és molt complexa. Es pot recórrer a l'aplicació de mètodes bayesians tal i com s'explica a l'annex III, però per això cal realitzar l'avaluació total de la quasi-versemblança, cosa que comporta un ús de tècniques computacionals molt intens.

Alternativament, es pot utilitzar la pseudo-versemblança o pseudo-versemblança penalitzada, PQL (Breslow i Clayton, 1993). L'adjectiu "pseudo" fa referència al fet que a cada iteració es maximitza la funció de versemblança d'una pseudo-variable, diferent a la corresponent a les dades originals. S'utilitza la mateixa pseudo-variable basada en l'extensió de sèries de Taylor de primer ordre que s'ha definit anteriorment:

$$z = g(\mu) + (y - \mu) B^{-1} = X\alpha + Z\beta + (Y - X\alpha - Z\beta) B^{-1}$$

La variància de z es defineix,

$$V_z = \text{Var}(X\alpha + Z\beta) + B^{-1} \text{Var}(Y - \mu) B^{-1} = ZGZ' + B^{-1}R B^{-1}$$

La matriu R es pot reescriure com $R = A^{1/2} B^{1/2} R_s B^{1/2} A^{1/2}$

$$V_z = ZGZ' + A^{1/2} B^{1/2} R_s B^{1/2} A^{1/2}$$

Aquesta matriu es força semblant a la de les dades originals, $V = BZGZ'B + A^{1/2} B^{1/2} R_s B^{1/2} A^{1/2}$

Wolfinger i O'Connell (1993) mostren que $z|\beta$ es distribueix normalment,

$$z|\beta \sim N(X\alpha + Z\beta, A^{1/2} B^{1/2} R_s B^{1/2} A^{1/2})$$

cosa que permet ser estimat ponderadament com un model mixt normal amb matriu residual R_s i matriu de ponderacions diagonal, $A^{-1}B$.

Degut a que z i B depenen dels estimadors de α i β , el model s'ha d'estimar iterativament com:

0.- Utilitzar com a valors inicials de z les dades originals Y (o bé $g(Y)$ evitant possibles valors infinits). La matriu de ponderacions inicial pot ser la identitat.

1.- Estimar un model mixt mitjançant REML tal i com es va veure més amunt.

2.- Calcular nous valors per z i B utilitzant els estimadors de α i β obtinguts en 1. Calcular la matriu de ponderacions $A^{-1}B$.

3.- Tornar al pas 1.

El nombre d'iteracions pot ser elevat ja que dins de cada etapa s'itera per tal d'obtenir els estimadors REML. Un cop el model ha convergit, es calcula la variància dels estimadors dels efectes fixes:

$$Var(\hat{\alpha}) = (X'V_z^{-1}X)^{-1}$$

Donat que V_z s'estima, aquesta variància pot estar lleugerament esbiaixada.

Referències

Breslow N. i Clayton DG. (1993): Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88(421), 9-25.

McCullagh P. i Nelder JA. (1989): *Generalized Linear Models*. Chapman & Hall, Londres.

Nelder JA. i Wedderburn RWM. (1972): Generalized Linear Models. *Journal of the Royal Statistical Association*, 135(3), 370-384.

Saez M. (2001): El problema de las medidas repetidas. Análisis longitudinal en epidemiología. *Gaceta Sanitaria*, 15(4), 347-352.

Saez M. i Saurina C. (2007): *Estadística y Epidemiología Espacial*. Girona: Edicions a Petició.

Wolfinger R. i O'Connell M. (1993): Generalized linear mixed models: a pseudo-likelihood approach. *Journal of Statistical Computation and Simulation*, 48, 223-243.

4. Mètodes

Fins ara s'han vist dues aproximacions al problema del tractament de mesures repetides: anàlisi de la supervivència multivariant i GLMM. És veritat que aquests últims només molt recentment han estat utilitzats per tractar variables resposta temps de supervivència i només aplicat en el camp de riscos competitiu. En aquest capítol s'adaptaran els dos principals enfocaments, Andersen i Goldstein, basats en els models GLMM per treballar amb variables resposta temps de supervivència multivariant en general, i no només per riscos competitiu.

Com es veurà a continuació, en l'aproximació d'Andersen la variable dependent és discreta mentre que en la de Goldstein és dicotòmica. El que es farà és aplicar els dos mètodes a un problema particular (reingressos deguts a la malaltia pulmonar obstructiva crònica) i comparar-los.

4.1. El mètode d'Andersen

En qualsevol estudi de supervivència, els individus es segueixen durant tot l'interval de temps i es van observant les diferents ocurrencies del(s) esdeveniment(s) d'interès. Sovint s'utilitzen els models multinivell per analitzar aquest tipus de dades on els esdeveniments es representen per transicions entre els diferents estats del model. En aquest context, els models multinivell han estat discutits per diversos autors, entre ells Andersen *et al.* (1993); i la majoria formulen els models de regressió per un procés multinivell incloent les covariables X_{ji} explícitament en un model per les intensitats de transició.

Andersen *et al.* (2003) mostren com modelar probabilitats en models multinivell introduint els anomenats pseudo-valors construïts a partir del mètode Jackknife. Llavors especifiquen la regressió a partir d'un GLM i així obtenen les estimacions dels paràmetres del model (veure també Andersen *et al.*, 2002; Klein i Andersen, 2005; Andersen i Klein, 2007 i Andrei i Murray, 2007). Més endavant, Klein *et al.* (2007) apliquen aquest mètode per l'anàlisi de la supervivència univariant (veure també Pohar i Andersen, 2008).

4.1.1. Metodologia.

Es comença aquest apartat de metodologia introduint una definició de les pseudo-observacions per passar, seguidament a definir el model lineal generalitzat i la seva

estimació. Després s'especificarà el model utilitzat per el cas de dades de supervivència multivariant.

Les pseudo-observacions s'obtenen a partir del mètode de Jackknife (Annex II) i la seva idea original consisteix en trobar un estimador $\hat{\theta}$ d'un paràmetre θ a partir de n estimacions de θ obtingudes en eliminar els individus de la mostra un a un i estimar cada vegada el paràmetre d'interès.

Seguint Andersen *et al.* (2003) i Andersen *et al.* (2004), les pseudo-observacions es poden definir tal com segueix. Es suposa que T_i ($i=1, \dots, n$) són variables aleatòries independents i idènticament distribuïdes, i θ un paràmetre de la forma

$$\theta = E(f(T_i)) \quad (4.1)$$

i es suposa també, que existeix un estimador no esbiaixat $\hat{\theta}$ d'aquest paràmetre. Per altra banda, si X_i ($i=1, \dots, n$) són covariables independents i idènticament distribuïdes, es pot definir l'esperança condicionada com:

$$\theta_i = E\{f(T_i)|X_i\} \quad (4.2)$$

Llavors, la pseudo-observació i -èssima es construeix com:

$$\hat{\theta}_i = n \cdot \hat{\theta} - (n-1) \cdot \hat{\theta}^{-i} \quad (4.3)$$

on $\hat{\theta}^{-i}$ és un estimador de θ obtingut a partir de les observacions T_j amb $j \neq i$, és a dir, aplicant el mètode de Jackknife (Annex II).

Normalment, els pseudo-valors han estat utilitzats per estudiar el biaix i la precisió de l'estimador $\hat{\theta}$, però Andersen *et al.* (2003) van més lluny i els utilitzen per a realitzar un model de regressió per el paràmetre θ gràcies a la relació definida en (4.2) entre θ_i i X_i ; i això permet definir un model lineal generalitzat amb funció vicle $g(\cdot)$

$$g(\theta_i) = \beta^T X_i \quad (4.4)$$

i per tant, la inversa d'aquesta funció vincle serà:

$$\theta_i = g^{-1}(\beta^T X_i) = \mu(\beta^T X_i) \quad (4.5)$$

Els coeficients del model, β , es poden estimar a partir de l'estimació d'equacions generalitzades

$$U(\beta) = \sum_{i=1}^n U_i(\beta) = \sum_{i=1}^n \left(\frac{\partial}{\partial \beta} g^{-1}(\beta^T X_i) \right) V_i^{-1} (\hat{\theta}_i - g^{-1}(\beta^T X_i)) = 0, \quad (4.6)$$

on V_i representa la matriu de covariàncies per $\hat{\theta}_i$ (Liang i Zeger, 1986; Zeger i Liang, 1986).

Andersen *et al.* (2003) mostren que a partir de l'expressió (4.6) s'obtenen estimadors consistents de β .

4.1.2. Adaptació del model pel cas de dades de supervivència multivariant.

Inicialment es suposa que $T = \{T_i\}$ ($i=1, \dots, n$) és una variable aleatòria que informa dels temps de supervivència amb funció de supervivència associada $S(t) = P(T \geq t)$. Aquesta funció de supervivència es denota per θ i l'estimador de Kaplan-Meier corresponent $\hat{\theta}$ serà un estimador no esbiaixat de θ (Annex I).

Per una mostra de mida n els pseudo-valors es poden definir de la següent manera (Pohar i Andersen, 2008):

$$\hat{\theta}_i = n \cdot \hat{\theta} - (n-1) \cdot \hat{\theta}^{-i} = n \cdot \hat{S}(t) - (n-1) \cdot \hat{S}^{-i}(t) \quad (4.7)$$

on $\hat{\theta}^{-i}$ és un estimador de la funció de supervivència obtingut a partir del mètode Jackknife eliminant les observacions una a una.

Aquests pseudo-valors s'utilitzen per definir un model lineal generalitzat amb funció vincle $g(x) = \text{cloglog}(x)$.

$$\text{cloglog}(\hat{\theta}_i) = \beta_0 + \beta^T X_i \quad (4.8)$$

Per el cas multivariant es vol estudiar K temps de supervivència a la vegada t_1, \dots, t_K , i llavors $\theta = (\theta_1, \dots, \theta_K) = \{S(t_1), \dots, S(t_K)\}$.

Fixat un moment t_k qualsevol es calcula l'estimador de Kaplan-Meier de la funció de supervivència per tot el conjunt d'individus, $\hat{S}(T_k)$. Després s'elimina l'observació i-èssima i es torna a fer la mateixa estimació, $\hat{S}_{-i}(T_k)$. A partir d'aquestes dues estimacions es defineix el pseudo-valor per l'individu i-èssim en el moment t_k com:

$$\hat{\theta}_{ik} = n\hat{S}(T_k) - (n-1)\hat{S}_{-i}(T_k) \quad i = 1, \dots, n \quad k = 1, \dots, K$$

Això es repeteix per a cada individu i per a cada moment del temps, i obtenim un vector de pseudo-valors $\hat{\theta}_i = (\hat{\theta}_{ik}, k = 1, \dots, K)$ per a cada individu.

Tal com es mostra en Andersen *et al.* (2003), aquests pseudo-valors es poden utilitzar per definir un model lineal generalitzat que permeti modelar l'efecte de les covariables sobre els temps de supervivència.

Existeixen diferents opcions per la funció vicle $g(\cdot)$ d'aquest GLM; però en el cas d'estudi s'utilitza la funció log-log complementària. Així, es defineix el GLM següent:

$$g(\theta_{ik}) = \beta_0 + \beta^T X_{jik} \quad (4.9)$$

$$g(x) = \log(-\log(x)) = \text{cloglog}(x)$$

on el subíndex i fa referència a l'individu ($i=1, \dots, n$), j fa referència a la covariable ($j=1, \dots, p$) i k a la recurrència ($k=1, \dots, K$).

Un cop definit el model lineal generalitzat, interessa estimar els paràmetres β del model; per això Andersen *et al.* (1993) utilitza l'aproximació de Liang i Zeger (1986).

4.1.3. Adaptació al cas d'estudi: reingressos a l'hospital degut a la malaltia pulmonar obstructiva crònica (MPOC).

La base de dades que s'utilitza per l'aplicació pràctica conté tots els individus que han ingressat com a mínim una vegada durant l'any 2003 per MPOC en algun hospital de

la Regió Sanitària de Girona (veure capítol 6). El que es vol modelar és la probabilitat de no tornar a ingressar a l'hospital d'un individu qualsevol.

Seguint el procés explicat en els apartats anteriors, el primer que es fa és contruir les pseudo-observacions a partir dels estimadors de Kaplan-Meier corresponents.

Aquestes pseudo-observacions configuraran la variable resposta del model final, que seguint (4.9), es pot especificar com:

$$c\log\log(\theta_{ik}) = \beta_{0i} + \sum_{j=1}^p \beta_j X_{jik} + u_i \quad (4.10)$$

on el subíndex $i=1,\dots,n$ fa referència a l'individu, $j=1,\dots,p$ fa referència a la covariable i $k=1,\dots,K$ a la recurrència; i els efectes aleatoris associats a l'individu venen representats per u_i . A més, el terme independent β_{0i} és un efecte aleatori específic per cada individu per tal de recollir la possible heterogeneïtat entre ells.

4.2. El mètode de Goldstein.

Tal i com ja s'ha dit, quan un esdeveniment pot passar més d'una vegada per cada individu, la durada entre recurrències pot estar correlacionada degut a la presència de factors individuals no observats. Aquest tipus d'esdeveniments normalment es tracten introduint efectes aleatoris individuals en el model, passant a tenir un model multivariant. Els efectes aleatoris representen la "fragilitat" individual i la variància d'aquests efectes mesura l'heterogeneïtat no observada entre individus (Steele *et al.*, 2004).

El mètode que es proposa en aquest apartat és un model general per dades multivariants. Aquest mètode és un cas especial del model no lineal multivariant proposat per Goldstein (1991) (de fet, un GLMM), i desenvolupat per Steele *et al.* (1996) per treballar amb models de riscos competitius i pel mateix Goldstein *et al.*, (2002) per treballar models de riscos competitius amb múltiples estats d'origen. Finalment, Steele, Goldstein i Browne (2004) van combinar totes aquestes extensions per obtenir un model de riscos competitius multinivell més general (veure també, Goldstein *et al.*, 2004 i Steele, 2008).

Com ja s'ha dit abans, aquest model general per l'anàlisi de dades multivariants basat en GLMM, on es poden tenir diversos estats d'origen i diverses destinacions des d'aquests estats, s'adapta al cas particular de temps de supervivència multivariant. Multivariant aquí vol dir diferents temps de fallida, en particular, diferents recurrències d'un mateix esdeveniment.

La idea principal d'aquest model es basa en utilitzar un model de regressió logística després de reestructurar la base de dades, convertint cada temps de supervivència en una seqüència binària anomenada resposta.

4.2.1. Metodologia.

Es comença amb una base de dades estructurada segons dos nivells (individus i episodis); i es suposa el cas en què no es té cap individu censurat. Més endavant s'explica el cas amb censura.

Com que la distribució de la durada és generalment asimètrica positiva (és impossible trobar una durada negativa), es treballa amb el logaritme de la durada (t) com a variable resposta. Així, es denota com a resposta y_{ik} on el subíndex i fa referència a l'individu i k a l'episodi.

Quan a les covariables, es poden classificar en dos grups. El primer fa referència a característiques dels individus que no varien amb el temps (com per exemple el sexe). Per aquest tipus de covariables s'utilitza la notació x_{ji} ($j=1,2,\dots,p$) ja que aquesta variable serà la mateixa per cada episodi dins de cada individu. El segon grup engloba les variables "temps-dependents" que varien amb el temps (per exemple l'edat). En aquest cas s'utilitza la notació x_{jik} ($j=1,2,\dots,p$) per indicar que la covariable pot variar el seu valor d'un episodi a l'altre dins de cada individu.

A més a més, cada individu tindrà els seus propis efectes aleatoris, els quals es poden entendre com un descriptor de la durada mitjana de cada individu; i habitualment s'assumeix que es distribueixen com una normal a través dels individus. Per la seva simplicitat s'acostuma a agafar una distribució normal encara que de fet els resultats són robustos a la distribució aleatòria dels efectes aleatoris. La notació utilitzada per descriure aquests efectes aleatoris és u_i .

Finalment, també s'assumeix que, degut a les característiques individuals, a alguns efectes temps-dependents i als efectes aleatoris individuals, la durada per cada estat es distribueix de forma independent; és a dir, dins de cada individu la durada mitjana d'un estat qualsevol i dels successius es distribueix de forma independent. Això es denota com e_{ik} .

El model doncs que s'utilitza en aquest cas és el que segueix:

$$y_{ik} = \beta_0 + \sum_{j=1}^p \beta_j x_{jik} + u_i + e_{ik} \quad (4.11)$$

$$y_{ik} = \log(t_{ik}), u_i \sim N(0, \sigma_u^2), e_{ik} \sim N(0, \sigma_e^2)$$

on σ_u^2 i σ_e^2 fan referència a la variància entre individus i entre episodis respectivament.

Per algunes situacions el model (4.11) és perfectament adequat, però apareixen problemes quan es tenen dades censurades. En aquest cas el supòsit de normalitat de e_{ik} és molt important i diferents supòsits donaran lloc a diferents estimacions. En molts casos no és fàcil establir la distribució més apropiada per e_{ik} , i per aquesta raó la majoria de models utilitzats intenten evitar utilitzar supòsits sobre la distribució tan forts. El procediment que es descriu evita l'ús d'aquests supòsits i permet incorporar una estructura multinivell completa.

Així doncs, situats en el cas de censura, la primera cosa és construir una variable resposta binària que pren el valor 1 si s'ha produït un canvi d'estat i 0 en cas contrari. L'objectiu és construir un model per la probabilitat de que es produeixi un canvi d'estat en un interval qualsevol. Es suposa que aquesta probabilitat depèn de l'estat al qual es troba l'individu en el moment d'estudi, el temps de permanència en aquest estat, d'algunes covariables i dels efectes aleatoris. Aquesta probabilitat (λ) és el que es coneix com a funció risc i en aquest cas es defineix com:

$$\lambda_{hik(t)} = P(y_{hik(t)} = 1 | y_{hik(t-1)} = 0) \quad (4.12)$$

on el subíndex i fa referència a l'individu, k a l'episodi i h a l'estat.

Finalment, per predir aquesta probabilitat es pot utilitzar un simple model logit treballant així amb el logaritme de l'odds ratio del canvi d'estat, és a dir, el ratio de la probabilitat de canvi d'estat respecte la probabilitat de no canvi. La variable resposta actual, la qual s'ha definit com una variable binària, es suposa que es distribueix com una Binomial, $y_{hik(t)} \sim \text{Binomial}(1, \lambda_{hik(t)})$.

El model general que s'utilitza és que el es mostra a continuació:

$$\log\left(\frac{\lambda_{hik(t)}}{1-\lambda_{hik(t)}}\right) = \text{logit}(\lambda_{hik(t)}) = \beta_0 + \sum_{j=1}^p \beta_j x_{jhik(t)} + f(t) + u_i + e_{ik} \quad (4.13)$$

on els termes u_i i e_{ik} fan referència a la variació individual i dels episodis respectivament, $x_{jhik(t)}$ és la variable dummy per l'estat actual i $f(t)$ és una funció del temps de permanència a l'estat actual. La part dreta de (4.13) segueix l'estructura estandard d'un model lineal multinivell; i seguint aquesta terminologia, la component

$\beta_0 + \sum_{j=1}^p \beta_j x_{jhik(t)} + f(t)$ representa la part fixa del model i la component $u_i + e_{ik}$ la part variable.

Per la funció suavitzadora $f(t)$ es podria utilitzar un conjunt de variables dummy ('blocking factors') pels diferents intervals $1, 2, \dots, n$, on n és el nombre màxim d'intervals per qualsevol estat, però apareix un problema quan es tenen pocs episodis de durada llarga, ja que això implicaria necessitar moltes variables dummy i llavors algunes d'elles podrien tenir una estimació molt pobra. Enlloc d'això, normalment s'utilitza una funció polinòmica del temps, $f(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \dots$

En general, utilitzant com a funció $f(t)$ el 'blocking factors', el model es pot especificar com:

$$\text{logit}(\lambda_{hik(t)}) = \beta_0 + \sum_{j=1}^p \beta_j x_{jhik(t)} + \sum_{r=1}^m \alpha_r^* z_{rht} + u_i^{(h)} + e_{ik} \quad (4.14)$$

$$y_{hik(t)} \sim \text{Binomial}(1, \lambda_{hik(t)})$$

on z_{rht}^* representa el conjunt de variables dummy per l'interval de temps t ; $u_i^{(h)}$ els efectes aleatoris per l'individu i a l'estat h ; i e_{ik} els efectes aleatoris associats a l'episodi k -èsim de l'individu i -èsim.

A la pràctica, Goldstein *et al.* (2004) exclou del model els efectes aleatoris associats a l'episodi degut a que no es mostren diferències en cap d'ells; i a més, utilitza una funció polinòmica de grau quatre enlloc del conjunt de variables dummy per la funció $f(t)$. Així, el model que presenta finalment és:

$$\text{logit}(\lambda_{hik(t)}) = \beta_0 + \sum_{j=1}^p \beta_j x_{jhik(t)} + \sum_{r=1}^4 \alpha_r t^r + u_i^{(h)} \quad (4.15)$$

on el polinomi $\sum_{r=1}^4 \alpha_r t^r$ substitueix el conjunt de variables dummy $\sum_{r=1}^m \alpha_r^* z_{rht}^*$.

4.2.2. Adaptació al cas d'estudi: reingressos a l'hospital deguts a MPOC.

En aquest apartat s'aplica el model descrit a una base de dades d'ingressos hospitalaris per malaltia pulmonar obstructiva crònica (MPOC) on la variable que es vol modelar és el risc que té un individu de tornar a ingressar a l'hospital degut a la MPOC, per tant, es tracta d'estudiar diferents recurrències d'un mateix esdeveniment (veure capítol 6).

Aquesta base de dades està estructurada en tres nivells: l'individu, que es representa amb el subíndex i ; l'episodi o recurrència, que correspon al subíndex k i l'estat del pacient que pot ser ingressat o no i està representat amb el subíndex h .

L'objectiu és construir un model per la probabilitat de que es produeixi un canvi d'estat en un interval qualsevol, en aquest cas passar d'estar ingressat a no estar-ho o al revés; i aquesta probabilitat (λ) és el que es coneix com a funció risc i tal com s'ha dit anteriorment es defineix com:

$$\lambda_{hik(t)} = P(y_{hik(t)} = 1 | y_{hik(t-1)} = 0) \quad (4.16)$$

on, tal com s'ha dit, el subíndex i fa referència a l'individu, k a l'episodi i h a l'estat.

Per tal de predir aquesta probabilitat s'ha utilitzat el model general definit a (4.15) i s'ha intentat modelar la funció suavitzadora $f(t)$ utilitzant *cubic splines* (funcions matemàtiques contínues definides a trossos per polinomis) (Remontet *et al.*, 2007) perquè després de realitzar un anàlisi de sensibilitat per compara-ho amb el polinomi de grau 4 que proposava Goldstein *et al.* (2004) es van obtenir millors resultats. Però, degut a que algunes vegades aquestes funcions es poden comportar malament a les cues, s'ha acabat aproximant aquesta funció a partir de *restricted cubic splines* (Harrell *et al.*, 1988). Aquest tipus de splines són similars als cubics splines amb la diferència que la part anterior al primer punt de tall i després de l'últim s'imposa que sigui lineal, és a dir, un polinomi de primer grau.

En general, si es defineixen r punts de tall, t_1, t_2, \dots, t_r , els restricted cubic splines es poden ajustar creant $r-2$ variables derivades (Durrleman *et al.*, 1989; Harrell *et al.*, 1988) i la seva representació és:

$$f(t) = \beta_0 + \beta_1 t + \sum_{j=1}^{k-2} \beta_{2j} \left[(t - t_j)^3 + \frac{t_k - t_j}{t_k - t_{k-1}} (t - t_{k-1})^3 + \frac{t_{k-1} - t_j}{t_k - t_{k-1}} (t - t_k)^3 \right] \quad (4.17)$$

on $\theta = (\beta_0, \beta_1, \beta_{21}, \dots, \beta_{2k-2})$ és el vector de coeficients.

En el cas que s'estudia s'han utilitzat quatre punts de tall corresponents als quartils del conjunt de dades.

Finalment, doncs, el model que s'ha utilitzat és:

$$\text{logit}(\lambda_{hik(t)}) = \beta_{0i} + \sum_{j=1}^p \beta_j x_{jhik(t)} + f(t) + u_i^{(h)} \quad (4.18)$$

on $u_i^{(h)}$ representa els efectes aleatoris per l'individu i a l'estat h .

S'ha exclòs del model els efectes aleatoris associats a l'episodi (e_{ik}) degut a que tal com passava en Goldstein *et al.* (2004), no es mostren diferències en cap d'ells. No obstant, el terme independent β_{0i} és un efecte aleatori específic per cada individu per tal de recollir la possible heterogeneïtat entre individus.

Per més informació sobre l'estudi realitzat s'adjunta a l'annex V l'article "*Modelling of multivariate survival data by means of Generalized Linear Mixed Models. The case of hospital admissions due to chronic obstructive pulmonary disease.*" on s'explica tot el procés amb més detall.

Referències

Andersen PK., Borgan O., Gill RD. i Keiding N. (1993): *Statistical models based on counting processes*. Springer-Verlag, New York.

Andersen PK., Abildstrom SZ. i Rosthoj S. (2002): Competing risks as a multi-state model. *Statistical Methods in Medical Research*, 11(2), 203-215.

Andersen PK., Klein JP. i Rosthoj S. (2003): Generalised linear models for correlated pseudo-observations, with applications to multi-state models. *Biometrika*, 90(1), 15-27.

Andersen PK., Hansen MG. i Klein JP. (2004): Regression analysis of restricted mean survival time based on pseudo-observations. *Lifetime Data Analysis*, 10(4), 335-350.

Andersen PK., Klein JP. (2007): Regression analysis for multistate models based on a pseudo-value approach, with applications to bone marrow transplantation studies. *Scandinavian Journal of Statistics*, 34(1), 3-16.

Andrei AC. i Murray S. (2007): Regression models for the mean of the quality-of-life-adjusted restricted survival time using pseudo-observations. *Biometrics*, 63, 398-404.

Goldstein H. (1991): Nonlinear multilevel Models, with an application to discrete response data. *Biometrika*, 78(1), 45-51.

Goldstein H., Browne W. i Rasbash J. (2002): Multilevel modelling of medical data. *Statistics in Medicine*, 21(21), 3291-3315.

Goldstein H., Pan H. i Bynner J. (2004): A flexible procedure for analysing longitudinal event histories using a multilevel model. *Understanding Statistics*, 3(2), 85-99.

Harrell FE., Lee KL. i Pollock BG. (1988): Regression models in clinical studies: determining relationships between predictors and response. *Journal of National Cancer Institutes*, 80(15), 1198-1202.

Klein JP. i Andersen PK. (2005): Regression modeling of competing risks data based on pseudo-values of the cumulative incidence function. *Biometrics*, 61(1), 223-229.

Klein JP., Logan B., Harhoff M. i Andersen PK. (2007): Analysing survival curves at a fixed point in time. *Statistics in Medicine*, 26, 4505-4519.

Liang KY. i Zeger SL. (1986): Longitudinal data analysis using generalized linear models. *Biometrika*, 78(1), 13-22.

Pohar M. i Andersen PK. (2008): Checking hazard regression models using pseudo-observations. *Statistics in medicine*, 27, 5309-5328.

Remontet L., Bossard N., Belot A. i Esteve J. (2007): An overall strategy based on regression models to estimate relative survival and model the effects of pronostic factors in cancer survival studies. *Statistics in Medicine*, 26(10):2214-2228.

Steele F., Diamond I. i Amin S. (1996): Immunization uptake in rural blangladesh: a multilevel analysis. *Journal of the Royal Statistical Society*, 159(2),289-299.

Steele F., Goldstein H. i Browne W. (2004): A General Multilevel Multistate Competing Risks Model for Event History Data, with an Application to a Study of Contraceptive Use Dynamics. *Statistical Modelling*, 4(2), 145-159.

Steele F. (2008): Multilevel models for longitudinal data. *Journal of the Royal Statistical Society*, 171,5-19.

Zeger SL. i Liang KY. (1986): Longitudinal data analysis for discrete and continuos outcomes. *Biometrics*, 42, 121-130.

5. Epidemiología espacial

El terme epidemiologia espacial o geogràfica s'utilitza per a descriure estudis sobre les causes i la prevenció de malalties a partir de l'ús de diferents perspectives d'anàlisi en les quals la localització geogràfica dels esdeveniments és una de les components fonamentals (Thomas, 1990; Saez i Saurina, 2007; Barceló *et al.*, 2008). Un dels seus principals objectius és mostrar quina part de la variació espacial de la distribució de la freqüència d'una malaltia no està explicada per la distribució espacial de factors explicatius, ni es pot atribuir a l'atzar. De fet, sovint interessa trobar pistes sobre algun factor de risc desconegut d'aquesta malaltia.

En concret en aquest context, l'enfoc de l'anàlisi s'emmarca en les dues primeres perspectives que segons Elliott i Watenberg (2004) es podria dividir l'epidemiologia espacial: mapes de malalties i estudis d'associació geogràfica.

5.1. Els mapes de malalties

Els mapes de malalties proporcionen un resum visual d'informació geogràfica complexa i permeten identificar patrons geogràfics de la malaltia que d'una altra manera podrien passar desapercebuts en les presentacions tabulars (Elliott i Watenberg, 2004; Quinn, 2007; Barcelo *et al.*, 2008). De fet, s'utilitzen amb propòsits descriptius; amb l'objectiu de generar hipòtesis etiològiques; per la vigilància de la salut pública, amb la finalitat de detectar àrees amb un risc aparentment major; i com ajuda en la definició de polítiques de salut i d'assignació de recursos.

També són útils per localitzar conglomerats (clusters) o agrupacions específiques de malalties, així com per estudiar els resultats d'altres estudis de focus contaminants (Wilkinson *et al.*, 1997; Saez i Saurina, 2007; Quinn, 2007).

Els mapes de malalties més simples són aquells en els quals es representen els casos observats en cada àrea geogràfica. No obstant, qualsevol interpretació de l'estructura geogràfica dels casos està limitada per la falta d'informació sobre la distribució espacial de la població que podria estar "en risc" i que, conseqüentment, ha generat aquests casos observats. Per tant, és preferible la representació de taxes que permetin incorporar l'efecte de la població en risc enlloc de representar els casos bruts. No obstant això, l'ús directe de les taxes brutes no permet la comparació entre diferents àrees, ja que les diferències observades entre elles poden ser degudes a factors de risc que no s'hagin tingut en compte com per exemple l'edat.

A continuació es començarà explicant els mapes per la mortalitat i després es particularitzarà al cas de reingressos hospitalaris deguts a la malaltia pulmonar obstructiva crònica (MPOC).

Una mesura de la mortalitat que té en compte l'estructura de l'edat és la taxa estandarditzada per edat. Existeixen dos mètodes per l'estandardització per edat que es coneixen com estandardització directa i indirecta. En la representació dels mapes de mortalitat es prefereix utilitzar el mètode indirecte que consisteix en la comparació de les morts observades en una àrea amb les esperades si els riscos per cada grup d'edat fossin els mateixos que els d'una certa població de referència. La raó observat/esperat s'anomena *raó de mortalitat estandarditzada* (RME o SMR en anglès), que no és més que un estimador del risc relatiu de l'àrea, és a dir del risc de mort en relació al grup considerat de referència.

Tot i que les RME han estat molt utilitzades tenen algunes limitacions (Clayton i Kaldor, 1987; Böhning, 1999). Depenen en gran mesura de la mida poblacional, ja que la varianza de les RME és inversament proporcional als valors esperats i així àrees amb poca població representaran estimadors amb gran variabilitat. En aquest sentit, les RME extremes i, per tant, dominants en l'aparent patró geogràfic, són les estimades amb menys precisió en les àrees que tenen pocs casos.

A més, la variabilitat dels casos observats sol ser bastant més gran que la dels esperats en una distribució de Poisson, produint-se el que s'anomena extra-variabilitat. De fet, quan es disposa de dades espacials és important distingir dues fonts d'extra-variació (Saez i Saurina, 2007; Barceló et al., 2008). En primer lloc, la font més important sol ser l'anomenada dependència espacial i és conseqüència de la correlació de la unitat espacial amb unitats espacials veïnes, generalment les contigües geogràficament. D'aquesta manera, les RME d'àrees contigües, o properes, són més similars que les RME d'àrees allunyades espacialment. Part d'aquesta dependència no és realment una dependència estructural, sinó que és deguda, principalment, a l'existència de variables no controlades, és a dir no incloses a l'ànàlisi. Respecte a la segona font, cal assumir l'existència d'extra-variació independent i incorrelacionada espacialment, anomenada heterogeneïtat incorrelacionada o no espacial, deguda a variables no observades sense estructura espacial que podrien influir en el risc (Lawson et al., 2003; Lawson, 2006).

5.2. Mètodes de suavització de les RME

Amb l'objectiu de solucionar els problemes derivats de la utilització directa de les RME s'han proposat diverses alternatives per suavitzar-les, és a dir, reduir la extra-variació.

Es parteix del supòsit que els casos observats es distribueixen com una Poisson,

$$O_i \sim \text{Poisson}(\mu_i E_i) \quad (5.1)$$

on O_i denota els casos observats a l'àrea i ; E_i són els casos esperats en aquesta àrea; i μ_i és el risc relatiu a l'àrea i .

En aquest cas l'estimador màxim versemblança del risc relatiu de cada àrea, $\hat{\mu}_i$, correspon a la RME (Saez i Saurina, 2007).

Resumint, els mètodes de suavització impliquen la formulació de models complexos on cal estimar els paràmetres corresponents. En general, s'utilitzen models jeràrquics bayesians complets (Besag *et al.*, 1991; Mollie, 1996; Best *et al.*, 2005).

El més habitual per suavitzar les raons és utilitzar el model proposat per Besag, York i Mollie (BYM) (Besag *et al.*, 1991; Mollie, 1996), el qual inclou dos efectes aleatoris: que recullen total la dependència espacial i l'heterogeneïtat o sobre dispersió (no espacial) desestructurada.

El model BYM s'especifica com un model lineal generalitzat mixte (GLMM) amb variable resposta de Poisson i considerant com offset els casos esperats (Saez i Saurina, 2007; Barceló *et al.*, 2008).

$$\begin{aligned} O_i &\sim \text{Poisson}(\mu_i E_i) \\ \text{Ln}(\mu_i) &= \alpha + \text{Ln}(E_i) + v_i + S_i \end{aligned} \quad (5.2)$$

on O_i , E_i i μ_i tenen el mateix significat que (5.1); v_i és l'efecte aleatori no espacial; i S_i l'efecte aleatori espacial.

L'efecte aleatori no espacial, també anomenat heterogeneïtat, es suposa distribuït normalment amb mitjana zero i variança constant. Per l'efecte aleatori, que recull la variabilitat espacial, s'utilitza un model condicional autoregressiu CAR (Besag, 1974;

Clayton *et al.*, 1993). Aquesta aproximació és la més habitual i la més simple en termes computacionals; i aproxima la dependència espacial com un promig de l'efecte espacial de les seves àrees veïnes (Saez i Saurina, 2007; Barceló *et al.*, 2008).

$$S_i | S_j, j \neq i \sim Normal(m_i, s_i^2)$$

$$m_i = \frac{1}{n_i} \sum_{j \in \vartheta_i} S_j \quad s_i^2 = \frac{s^2}{n_i} \quad (5.3)$$

on ϑ_i representa el conjunt d'àrees adjacents a l'àrea i ; i n_i és el número de veïns.

5.3. Variabilitat geogràfica dels reingressos hospitalaris deguts a la malaltia pulmonar obstructiva crònica (MPOC)

La part pràctica d'aquesta tesi s'ha aplicat a l'estudi dels reingressos hospitalaris deguts a la malaltia pulmonar obstructiva crònica (MPOC) (capítol 6). Una de les innovacions que s'introdueixen pels dos mètodes estudiats, Andersen i Goldstein (capítol 4) ha estat la introducció en el model utilitzat d'efectes espacials per tal de recollir la variabilitat geogràfica.

En el cas del mètode de Goldstein, l'especificació del model amb efectes espacials consisteix en:

$$\text{logit}(\lambda_{hik}(t)) = \beta_{0i} + \sum_{j=1}^p \beta_j X_{jhik}(t) + f(t) + S_i(t) + u_i^{(h)} \quad (5.4)$$

on $f(t)$ és una funció suavitzadora que s'ha modelat utilitzant *cubic splines*, $u_i^{(h)}$ representa els efectes aleatoris per l'individu i a l'estat h i s'incorporen els efectes espacials amb la funció S_i , i $S_i | S_j, j \neq i \sim Normal(m_i, s_i^2)$.

Pel mètode d'Andersen l'especificació del model amb efectes espacials consisteix en:

$$\text{cloglog}(\theta_{ik}) = \beta_{0i} + \sum_{j=1}^p \beta_j X_{jik}(t) + S_i(t) + u_i \quad (5.5)$$

on igualment, s'incorporen els efectes espacials amb la funció S_i i $S_i | S_j, j \neq i \sim Normal(m_i, s_i^2)$.

5.4. Estimació del model

5.4.1. Aproximació bayesiana vs. aproximació freqüentista

Els models s'estimen utilitzant l'aproximació bayesiana. De fet, quan es modelitza la variabilitat geogràfica aquesta aproximació és la única factible. El problema de l'aproximació freqüentista és que el fet de treballar amb matrius de dimensió gran la tasca de maximització es converteix en inabordable amb mitjans computacionals (Saez i Saurina, 2007).

No obstant això, encara que no es tingui en compte la variabilitat geogràfica, l'estimació bayesiana presenta moltes més avantatges respecte la seva alternativa freqüentista (Best, 2006).

En primer lloc, mentre que l'aproximació freqüentista no pot proporcionar una estimació on intervinguin probabilitats, aquesta estimació resulta molt senzilla utilitzant l'aproximació bayesiana. Per exemple, un aspecte que cal tenir en compte quan s'utilitzen mètodes de suavització és el grau de suavització que s'aconsegueixi (Elliott i Wartenberg, 2004). Aquest grau determina el compromís entre una alta sensibilitat (àrees d'excés de risc correctament identificades) i una alta especificitat (àrees que no tenen excés de risc correctament identificades) (Richardson *et al.*, 2004). Un resultat sensible però no específic generarà falsos positius, mentre que un resultat específic però no sensible no identificarà aquelles àrees amb risc alt. Richardson *et al.* (2004) proposen estimar, a més a més de les RME, la probabilitat a posteriori, PrP, que les raons siguin majors que la unitat, és a dir $Pr P = Prob(\hat{\mu} > 1)$. Després d'un exhaustiu estudi de simulació aquests autors van mostrar que el mètode de suavització proporcionarà resultats suficientment específics, (mantenint una sensibilitat raonable) quan aquesta probabilitat sigui major del 80% ($PrP > 0.8$), proposant utilitzar aquests valors per identificar àrees d'alt risc. Anàlogament, quan $PrP < 0.2$ l'àrea es podria considerar de baix risc (Barceló *et al.*, 2008). Doncs bé, les PrP només es poden estimar mitjançant l'aproximació bayesiana (Best, 2006).

En segon lloc, l'estimació freqüentista d'una funció de paràmetres $g(\theta)$, necessita la construcció d'un estimador específic de la funció i això no sempre és possible. El risc atribuïble, per exemple, és una funció del risc relatiu i de la probabilitat d'exposició (Best, 2006). Però, encara que es pogués construir l'estimador de la funció, la inferència sobre aquesta pot ser complicada en alguns casos. Per exemple, en funcions no lineals dels paràmetres del model, la variança no es pot obtenir directament, i es fa necessari recórrer, per exemple, al mètode δ (Lambert i Dickman, 2007). Si alternativament, s'utilitza una aproximació bayesiana, llavors s'obtidria directament una estimació d'aquesta variança.

En tercer lloc, la manipulació de dades mancants, de patrons no estàndard de censura i d'errors de mesura en les variables explicatives, resulta molt més fàcil utilitzant una aproximació bayesiana (Best, 2006).

I per últim, l'aproximació bayesiana no es basa en aproximacions asimptòtiques, la qual cosa resulta molt apropiada quan s'analitzen mostres petites i/o quan es pretén estimar les variàncies dels paràmetres d'interès (Best, 2006).

5.4.2. Estimació bayesiana

L'estimació bayesiana es basa en la distribució a posteriori, $p(\theta|x)$, del valor del paràmetre d'interès, θ , la qual expressa la incertesa sobre el paràmetre un cop s'ha tingut en compte les dades (Best, 2006; Saez i Saurina, 2007).

El teorema de Bayes indica com calcular la distribució a posteriori:

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{\int p(\theta)p(x|\theta)d\theta}$$

Com que el denominador és en realitat una constant, l'equació bàsica de l'aproximació bayesiana s'escriuria:

$$p(\theta|x) \propto p(x|\theta)p(\theta) \tag{5.6}$$

on θ és un vector de paràmetres, habitualment de dimensió gran; $p(x|\theta)$ s'anomena versemblança; i $p(\theta)$ és la distribució a priori.

L'estimació bayesiana completa requereix que es proporcionin la versemblança, és a dir, que s'especifiqui el model; i les distribucions a priori dels paràmetres d'interès. En concret, distribucions a priori per els paràmetres del model (per exemple, l'associat a alguna variable explicativa), per els efectes aleatoris, així com per les desviacions típiques d'aquests efectes (Saez i Saurina, 2007; Barceló *et al.*, 2008).

Mentre que $p(x|\theta)$ (la versemblança) i $p(\theta)$ (distribució a priori) són fàcils d'obtenir, $p(\theta|x)$ sol ser analíticament inabordable, especialment quan no s'utilitzen distribucions a priori conjugades.

Quan una distribució a posteriori és de la mateixa família que la distribució prior utilitzada es parla de distribucions conjugades. Utilitzar aquest tipus de distribucions comporta que els paràmetres (a priori) tinguin bones propietats matemàtiques per el càlcul de les distribucions a posteriori. El següent requadre mostra les distribucions a priori conjugades per les distribucions més habituals (normal, binomial i poisson).

Taula 6. Distribucions a priori conjugades

Versemblança	Paràmetre a estimar	Distribucions	
		A priori	A posteriori
Normal	Mitjana	Normal	Normal
Normal	Precisió	Gamma	Gamma
Binomial	Probabilitat èxit	Beta	Beta
Poisson	Mitjana	Gamma	Gamma

Font: Saez i Saurina, 2007.

En qualsevol cas interessa obtenir la distribució marginal a posteriori,

$$p(\theta_i|x) = \int \int \dots \int p(\theta|x) d\theta_{(-i)}$$

on $\theta_{(-i)}$ denota el vector θ excloent la component i ; i ser capaços de calcular algunes propietats de $p(\theta_i|x)$ com la mitjana ($\int \theta_i p(\theta_i|x) d\theta_i$) o determinats percentils ($\int_t^\infty p(\theta_i|x) d\theta_i$).

Quan no és possible obtenir-ho per mètodes analítics cal recórrer a la integració numèrica.

L'objectiu és obtenir una mostra de la distribució conjunta a posteriori $p(\theta|x)$. El mostreig independent sol ser molt complex, però el mostreig dependent a partir de cadenes de Markov amb $p(\theta|x)$ com la seva distribució estacionària (d'equilibri) (Monte Carlo Markov Chain, MCMC) és molt més senzill.

Una seqüència de variables aleatòries $\theta^{(0)}, \theta^{(1)}, \theta^{(2)}, \dots$ forma una cadena de Markov si $\theta^{(i+1)} \longrightarrow p(\theta^{(i+1)}|\theta^{(i)})$, és a dir, condicionat al valor de $\theta^{(i)}$, $\theta^{(i+1)}$ és independent de $\theta^{(i-1)}, \dots, \theta^{(0)}$..

Existeixen diversos algorismes per dissenyar cadenes de Markov amb una distribució estacionària $p(\theta|x)$; un dels més senzills és l'algorisme 'Gibbs sampling'. Aquest algorisme va ser introduït per Geman i Geman (1984) en el context de processament d'imatges, més tard va ser discutit en el context de dades mancants per Tanner i Wong (1987) i aplicat a problemes bayesians per Gelfand i Smith (1990).

L'algorisme Gibbs sampling es pot descriure tal com es fa a continuació (Gilks *et al.*, 1996; Spiegelhalter *et al.*, 2008):

Sigui θ un vector de paràmetres desconeguts tal que $\theta = (\theta_1, \theta_2, \dots, \theta_k)$:

1.- s'escullen valors inicial $\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_k^{(0)}$ per les components.

2.- es mostreja $\theta_1^{(1)}$ a partir de $p(\theta_1|\theta_2^{(0)}, \theta_3^{(0)}, \dots, \theta_k^{(0)}, x)$

es mostreja $\theta_2^{(1)}$ a partir de $p(\theta_2|\theta_1^{(1)}, \theta_3^{(0)}, \dots, \theta_k^{(0)}, x)$

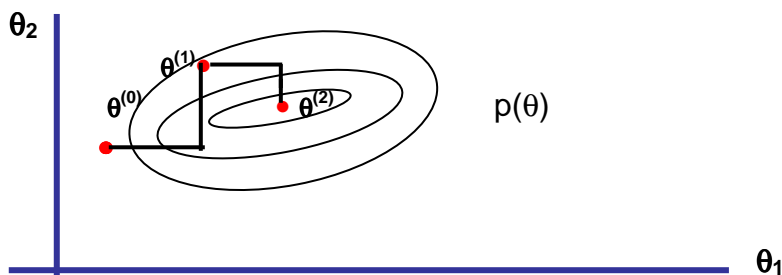
...

es mostreja $\theta_k^{(1)}$ a partir de $p(\theta_k | \theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_{k-1}^{(1)}, x)$.

3.- es repeteix l'etapa 2 moltes vegades (si el número de repeticions és molt gran s'obtindrà una mostra per $p(\theta|x)$).

A la figura 7 es mostra un exemple de Gibbs sampling per $k=2$.

Figura 7.- Gibbs sampling per $k=2$.



- Mostrejar $\theta_1^{(1)}$ a partir de $p(\theta_1 | \theta_2^{(0)}, x)$
- Mostrejar $\theta_2^{(1)}$ a partir de $p(\theta_2 | \theta_1^{(1)}, x)$
- Mostrejar $\theta_1^{(2)}$ a partir de $p(\theta_1 | \theta_2^{(1)}, x)$
- ...

5.4.3. Convergència i eficiència de les estimacions.

Com a conseqüència de la seva forma iterativa, la convergència de les simulacions resulta crucial per validar els resultats obtinguts a partir dels mètodes bayesians complets (Barceló *et al.*, 2008). La convergència es pot definir com a quina velocitat s'aproxima la distribució del paràmetre d'interès a la distribució a posteriori (Best, 2006).

Després d'aconseguir la convergència, la representació gràfica de les simulacions es distribuirà aleatòriament entorn una mitjana constant. Per tant, la pràctica usual és utilitzar diverses cadenes, partint de valors inicials diferents i comprovar la convergència, ja sigui gràficament, mitjançant gràfics anomenats de traça; o bé

analíticament, mitjançant l'estadística de Brooks-Gelman (1998), que s'ha d'aproximar a la unitat.

Quan les simulacions han convergit, cal comprovar la seva eficiència. Per això cal realitzar iteracions addicionals i observar l'error estàndard de Monte Carlo (MC error), que no és més que l'error estàndard de la mitjana de la simulació a posteriori, és a dir, de l'estimador de l'esperança teòrica del paràmetre d'interés (Best, 2006). Com a regla intuïtiva, l'eficiència s'aconseguirà quan el MC error sigui menor del 5% de la desviació típica a posteriori del paràmetre d'interés (Saez i Saurina, 2007; Best, 2006).

Referències

Barceló MA., Saez M., Cano-Serral G., Martínez-Beneito MA., Martínez JM., Borrell C., Ocaña-Riola R., Montoya I., Calvo M., López-Abente G., Rodríguez-Sanz M., Toro S., Alcalá JT., Saurina C., Sánchez-Villegas P. i Figueiras A. (2008): Métodos para la suavización de indicadores de mortalidad: aplicación al análisis de desigualdades en mortalidad en ciudades del estado Español (proyecto MEDEA). *Gaceta Sanitaria*, 22(6): 596-608.

Besag J. (1974): Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of Royal Statistical Society, Series B*, 36, 192–236.

Besag J., York J. i Mollie A. (1991): Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43, 1-59.

Best N., Richardson S. i Thomson A. (2005): A comparison of Bayesian spatial models for disease mapping. *Statistical Methods in Medical Research*, 14, 35-39.

Best N. (2006): *Bayesian approaches to cancer survival analysis*. Short course on Cancer survival: principles, methods and applications. London.

Böhning D. (1999): *Computer-assisted analysis of mixtures and applications: meta-analysis, disease mapping and others*. Boca Raton, FLA: Chapman and Hall.

Brooks SP. i Gelman A. (1998): Alternatives methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7, 434-455.

Clayton D. i Kaldor J. (1987): Empirical Bayes estimates for age-standardized relative risks. *Biometrics*, 43, 671-681.

Clayton D., Bernadinelli L. i Montomoli C. (1993): Spatial correlation in ecological analysis. *International Journal of Epidemiology*, 22, 1193–1202.

Elliott P. i Wartenberg D. (2004): Spatial epidemiology: current approaches and future challenges. *Environmental Health Perspectives*, 112(9), 998-1006.

Gelfand AE. i Smith AFM. (1990): Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398-409.

Geman S. i Geman D. (1984): Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12, 609-628.

Gilks WR., Richardson S. i Spiegelhalter DJ. (1996): *Markov Chain Monte Carlo in Practice*. Chapman & Hall, New York.

Lambert PC. i Dickman PW. (2008): *Statistical models for relative survival*. Barcelona, 22 i 23 de novembre de 2007. [citat el 28 de Agosto de 2008]. Disponible a: <http://www.pauldickman.com/teaching/index.php>

Lawson AB., Browne WJ. i Vidal-Rodeiro CL. (2003): *Disease Mapping with WinBUGS and MLwiN*. Chichester, West Sussex: John Wiley & Sons.

Lawson AB. (2006): *Statistical methods in spatial epidemiology* (second edition). New Jersey: John Wiley & Sons.

Mollie A. (1996): *Bayesian mapping of disease*. En Gilks WR, Richardson S, Spiegelhalter DJ (eds). *Markov Chain Monte Carlo in Practice*. Nueva York: Chapman & Hall, 359-379.

Quinn M. Introduction.(2007): *En Office for National Statistics (ONS). Cancer Atlas of the United Kingdom and Ireland 1991-2000*. Londres: ONS, 2005, pp. 1-5. [citat el 20 de juliol de 2007]. Disponible a:

<http://www.statistics.gov.uk/statbase/Product.asp?vlnk=14059>

Richardson S., Thomson A., Best N. i Elliot P. (2004): Interpreting posterior relative risk estimates in disease-mapping studies. *Environmental Health Perspectives*, 112:1016-1025.

Saez M. i Saurina C. (2007): *Estadística y Epidemiología Espacial*. Girona: Edicions a Petició.

Spiegelhalter D., Thomas A., Best N. i Lunn D. (2008): *WinBUGS User Manual*. Version 1.4. [citat el 28 de Agosto de 2008]. Disponible a: <http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml>

Tanner MA. i Wong WH. (1987): The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82: 528-549.

Thomas RW. (1990): *Spatial epidemiology*. Londres: Pion.

Wilkinson P., Thakrar B., Shaddick G., Stevenson S., Patternden S., Landon M., Grundy C. i Elliot P. (1997): Cancer incidence and mortality around the Pan Britannica Industries pesticide factory, Waltham Abbey. *Occupational and Environmental Medicine*, 54, 101-107.

6. Resultats de l'estudi.

6.1. Descriptius de la base de dades utilitzada.

Amb la intenció de realitzar una aplicació pràctica dels mètodes presentats s'ha treballat amb un arxiu de dades referent a individus residents a la Regió Sanitària de Girona (província de Girona excepte la comarca de la Cerdanya), que van estar ingressats com a mínim una vegada durant l'any 2003, amb diagnòstic de malaltia pulmonar obstructiva crònica (MPOC).

Aquesta malaltia és una de les malalties pulmonars obstructives més freqüent ja que a Catalunya la pateixen un 9% de les persones entre 40 i 70 anys, malgrat que només un 22 % són diagnosticades a temps per rebre algun tipus de tractament o consells preventius que millorin la seva evolució. En un 90 % dels casos és produïda per l'addicció al tabac i la resta com a conseqüència d'infeccions greus en la infantesa, defectes genètics, treballar en determinats ambients contaminats i altres causes encara no conegudes. Apareix de manera lenta i progressiva a partir dels 50-60 anys, generalment en persones fumadores. Al principi només provoca tos matutina, expectoració i bronquitis agudes prolongades que el malalt no consulta al metge ja que considera que és "la tos normal del fumador". Al cap d'uns anys apareix l'ofec progressiu en l'esforç amb xiulets i aguditzacions més greus i freqüents que invaliden el malalt i en redueixen la qualitat de vida. En els últims anys la gran obstrucció dels bronquis, la manca d'oxigen i l'excés d'anhídrid carbònic a la sang provoquen una dificultat per fer simplement petits esforços: pujar uns graons, dutxar-se, vestir-se i se succeeixen els ingressos hospitalaris, cal oxigen al domicili i utilitzar cada vegada més medicaments.

Actualment és la tercera causa de mort, que en la meitat dels casos es produeix durant els deu anys següents al diagnòstic; a més a més, la MPOC provoca un important cost sanitari directe i indirecte, amb una proporció molt elevada de consultes a l'atenció primària i especialitzada, molts ingressos hospitalaris i un alt índex d'absentisme laboral.

Segons la Organització Mundial de la Salut (OMS), aquesta malaltia suposa el 3% de la despesa de la sanitat pública (uns 3.000 milions d'euros anuals); el 84% de la qual és deguda a hospitalitzacions. El cost anual mig per pacient en atenció primària és de 1876 euros, sense tenir en compte els costos indirectes per baixa laboral. El d'un

pacient greu pot arribar fins a 3538 euros i en alguns casos multiplicar-se per cinc. (Sánchez, 2008).

Aquesta aplicació pràctica té un doble objectiu. Per una banda il·lustrar aquesta tesi doctoral amb un exemple amb dades recurrents, i per altra banda, el seguiment d'aquesta malaltia posa de manifest la qualitat de l'atenció primària ja que si es diagnostica a temps es poden evitar els reingressos hospitalaris.

L'arxiu de treball s'ha construït a partir de dues fonts d'informació diferents. Per una banda, es disposava de les dades provinents del Conjunt Mínim Bàsic d'Altes Hospitalàries, CMBDAH, de l'any 2003. En aquesta base de dades s'hi troben tots els ingressos hospitalaris que es van produir als hospitals de la RSG (Annex III) durant l'any 2003 tant de residents a la RSG com fora d'aquesta. Per cada ingrés es disposa de diferent informació personal del pacient (edat, sexe, municipi de residència, data de naixement, etc) i altra informació referent a l'ingrés (diagnòstic principal i secundaris, procediments aplicats, data d'ingrés, data d'alta, etc).

L'altra font d'informació provinent de l'Institut Català de la Salut (ICS) va proporcionar informació referent a tots els usuaris atesos l'any 2003 a qualsevol dels seus centres d'atenció primària (Annex III), com ara el nombre i tipus de visites realitzades per cada usuari o els seus problemes de salut.

Finalment, per construir la base de dades definitiva es van creuar aquestes dues fonts d'informació tenint en compte els següents criteris de selecció; primer només s'han estudiat pacients residents a la Regió Sanitària de Girona amb diagnòstic principal i/o secundaris MPOC; alta hospitalària catalogada com a episodi mèdic; categoria major diagnòstica cardíaca o respiratòria (agrupació en grans categories, mútuament excloents, a partir del diagnòstic principal de l'episodi); i s'han eliminat els ingressos on el tipus d'activitat fos hospital de dia ja que aquests no es consideren com un ingrés. Després de tot això, s'ha obtingut un arxiu amb 2303 individus diferents, els quals han produït 3276 altes.

A continuació es presenten les variables més importants que formen part de la base de dades definitiva.

Sexe: variable qualitativa que fa referència al sexe del pacient i té dues categories, home i dona.

Edat: variable numèrica que conté l'edat del pacient.

Unitat productiva assignada: variable qualitativa que ens informa de l'Àrea Bàsica de Salut assignada al malalt segons la seva residència i conté 39 categories diferents.

Municipi: variable qualitativa que ens informa del municipi de residència del malalt.

Problemes de salut: conjunt de variables qualitatives que informen sobre si el pacient presenta algun dels següents problemes de salut: alcoholisme, MPOC, insuficiència cardíaca, hipertensió arterial, dislipèmia, diabetis mellitus tipus 2, tabaquisme.

Circumstància d'admissió: variable qualitativa que informa sobre si l'ingrés ha estat urgent o programat.

Circumstància d'alta: variable qualitativa que informa del tipus d'alta que ha generat cada ingrés. Les categories d'aquesta variable són: destinació al propi domicili, trasllat a un hospital general d'aguts o a un hospital monogràfic psiquiàtric, trasllat a un centre o servei sociosanitari de mitjana o llarga durada, trasllat a una residència social, alta voluntària, defunció, evasió o alta administrativa i hospitalització domiciliària.

Procedència: variable qualitativa que indica si un pacient és derivat des d'un altre recurs sanitari o no. Les seves categories són: domicili o residència social, trasllat des d'un hospital d'aguts o un hospital monogràfic psiquiàtric, trasllat des d'un recurs de la xarxa sociosanitària, trasllat des d'un altre centre de la xarxa ambulatoria de salut mental, trasllat des d'un centre d'atenció primària, trasllat des d'hospitalització domiciliària o trasllat des de la consulta externa del mateix hospital.

GRD: variable qualitativa que informa del Grup Relacionat amb el Diagnòstic de cada ingrés hospitalari. El Sistema de Classificació en Grups Relacionats amb el Diagnòstic (GRD) és el sistema per mesurar el producte hospitalari més desenvolupat i internacionalment validat actualment. Aquest sistema classifica cada estada hospitalària en un dels 511 grups possibles (en versió 9.0 HCFA2) d'acord amb el consum de recursos que ha generat i la patologia causant de l'estada.

² Health Care Financing Administration

Pes del GRD: variable numèrica que conté el pes relatiu del GRD, és a dir, la relació entre el cost mig de l'estada del GRD en concret, sobre el cost promig de l'estada, en el conjunt global de tots els GRD. Un pes relatiu més gran que la unitat implica que el GRD té un cost més gran, i segurament també més complexitat, que el promig.

Dies d'estada: variable numèrica que informa de la durada de l'estada.

6.1.1. Anàlisi global de les dades

En aquest apartat es mostra un anàlisi descriptiu de les variables més importants de l'arxiu d'estudi. Es recorda que aquesta base de dades consta de 2303 individus diferents, els quals han produït 3276 altes hospitalàries.

A la taula 7 s'observa que del total d'individus, el 71.5% són homes i el 28.5% dones.

Taula 7. Distribució dels individus segons el sexe.

	Nº d'individus	% d'individus
Homes	1646	71.5
Dones	657	28.5
Total	2303	100.0

Tal com es pot veure a la taula 8, la majoria dels pacients, un 72.9%, han ingressat només una vegada i la resta de dues a sis vegades. El nombre d'ingressos presenta diferències estadísticament significatives entre ambdós sexes (p-valor: 0.005); observant-se que el grup d'homes mostra un nombre de reingressos més elevat que el de dones.

Taula 8. Distribució per sexes del nombre d'ingressos.

Nº d'ingressos	Sexe		Total n(%)
	Homes n(%)	Dones n(%)	
1	1160(70.5)	519(79.0)	1679(72.9)
2	324(19.7)	92(14.0)	416(18.1)
3	99(6.0)	26(4.0)	125(5.4)
4	35(2.1)	13(2.0)	48(2.1)
5	17(1.0)	3(0.5)	20(0.9)
6	7(0.4)	2(0.3)	9(0.4)
7	1(0.06)	2(0.3)	3(0.1)
8	2(0.12)	0(0.0)	2(0.1)
9	1(0.06)	0(0.0)	1(0.0)

Els individus d'aquesta base de dades tenen una mitjana d'edat de 68.59 anys (taula 9). La meitat tenen una edat inferior a 74 anys i l'altra meitat superior. Això indica que els ingressos han estat realitzats per individus amb edat avançada i que per això la distribució és asimètrica a l'esquerra. Aquesta idea es reforça quan s'observen els quartils; el primer quartil ja pren un valor molt alt (65 anys), la qual cosa indica que el 75% dels individus tenen una edat superior a 65 anys, i si s'observa el tercer quartil es veu com encara el 25% té una edat superior a 80 anys. L'edat més repetida és de 78 anys amb un valor màxim de 102 i un mínim de 1 any.

Taula 9. Descriptius de l'edat del pacient (anys).

Mitjana	68.59	Moda	78.00
Mediana	74.00	Mínim	1.00
Mitjana retallada el 5%	70.75	Màxim	102.00
Desviació típica	19.11	Quartils	
		Q1	65.00
		Q3	80.00

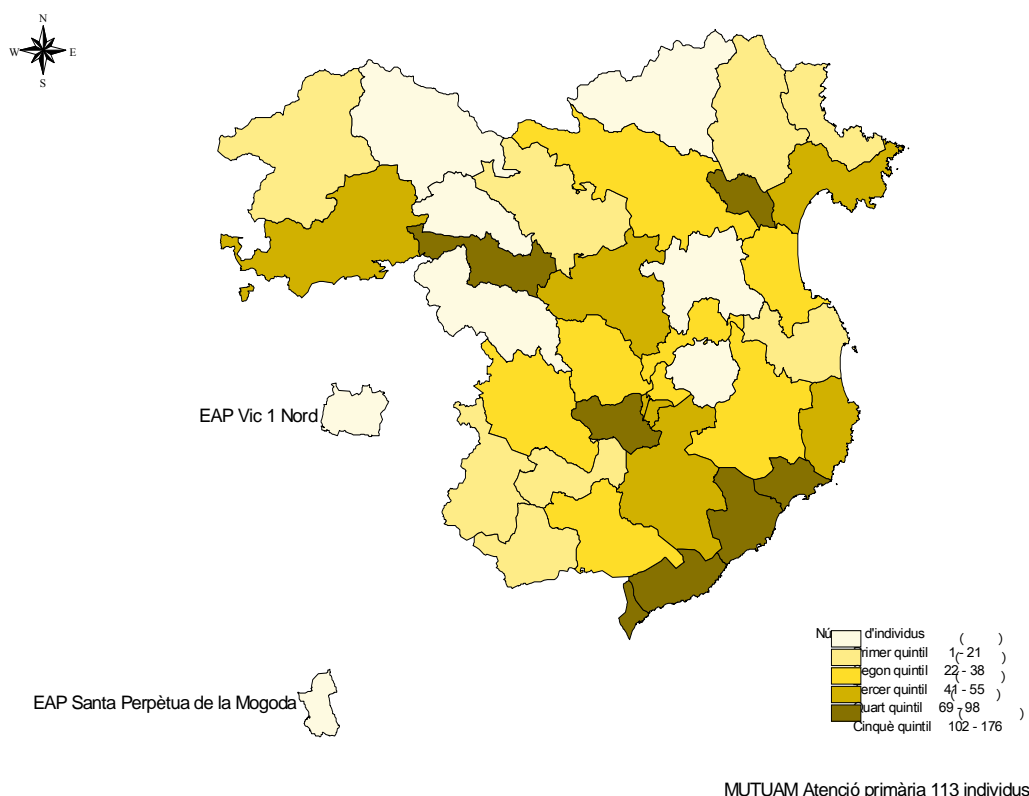
Per sexes s'observa com les distribucions continuen mostrant individus amb edat avançada tot i que les dones tenen una mitjana i una mediana inferior a la dels homes (taula 10). Per altra banda, la distribució de l'edat en les dones és més dispersa que en els homes, i el 50% de les dades centrals es mouen en un interval (58.00-81.00) més gran que en el cas dels homes (65.00-80.00). Aquesta diferència entre sexes respecte l'edat mitjana ha resultat ser estadísticament significativa (p -valor<0.0001).

Taula 10. Descriptius de l'edat del pacient segons el sexe (anys).

	SEXE	
	Home	Dona
Mitjana	69.82	65.53
Mediana	74.00	73.00
Mitjana retallada el 5%	71.89	67.50
Desviació típica	17.34	22.68
Moda	78.00	77.00
Mínim	1.00	1.00
Màxim	102.00	96.00
Quartils		
Q1	65.00	58.00
Q3	80.00	81.00

A continuació es mostra un mapa (mapa 1) on s'ha representat la distribució dels individus ingressats segons l'àrea bàsica de salut (ABS) assignada a cada un d'ells. S'ha pintat el mapa de manera que com més fosc és el color corresponent a l'ABS més individus té assignats. S'observa com a les zones litoral i prelitoral és on es concentren les ABS amb més individus.

Mapa 1. Distribució dels individus segons l'àrea bàsica de salut assignada



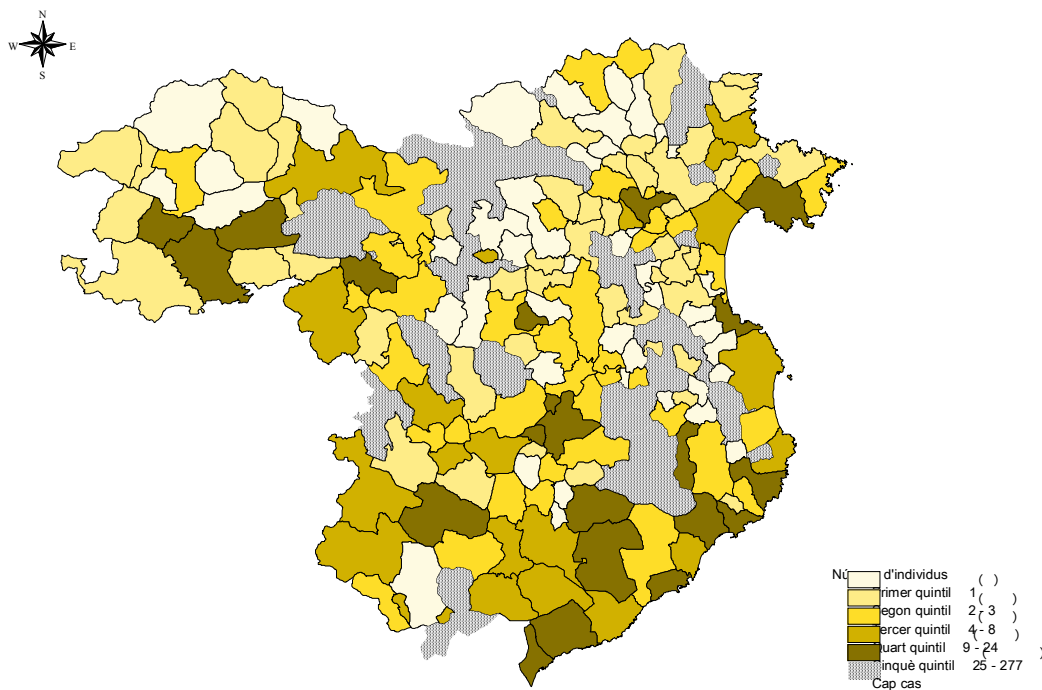
La següent variable analitzada és el municipi de residència. A la taula 11 es representen els 10 municipis de residència més freqüents.

El mapa següent (mapa 2) mostra la distribució dels individus segons el municipi de residència. S'observa com es manté la tendència vista en el mapa 1 on els municipis d'on provenen més pacients són els situats més cap a la costa.

Taula 11. Nombre d'individus segons municipi de residència més freqüent.

Codi	Descripció municipi	Nº d'individus (%)
17079	Girona	277(12.03)
17066	Figueres	163 (7.08)
17023	Blanes	162 (7.03)
17114	Olot	111 (4.82)
17155	Salt	101 (4.39)
17160	Sant Feliu de Guíxols	91 (3.95)
17095	Lloret de Mar	85 (3.69)
17118	Palamós	78 (3.39)
17147	Ripoll	75 (3.26)
17117	Palafrugell	70 (3.04)
-	Altres	1090 (47.32)

Mapa 2. Distribució dels individus segons el municipi de residència.



A continuació es mostra el nombre d'individus que presenten cada un dels problemes de salut que poden influir en la malaltia segons el sexe (taula 12). S'observa que el

problema més comú en general és MPOC (14.6%) seguit de la hipertensió arterial (11.3%).

Per sexes, els homes segueixen la tendència global ja que el problema més comú és MPOC (16.0%); en canvi en les dones el més comú és la hipertensió arterial (12.9%).

Es mostren diferències significatives entre ambdós sexes pels problemes alcohol (p-valor <0.0001); MPOC (p-valor:0.003) i tabac (p-valor <0.0001). Destacar també que el problema del tabac és molt més comú en homes (8.3%) que en dones (4.0%).

Taula 12. Distribució dels individus segons sexe i problemes de salut que pateixen.

	No presenten el problema			Presenten el problema		
	Homes	Dones	Total	Homes	Dones	Total
Consum d'alcohol	1588 (96.5%)	656 (99.8%)	2244 (97.4%)	58 (3.5%)	1 (0.2%)	59 (2.6%)
MPOC	1382 (84.0%)	584 (88.9%)	1966 (85.4%)	264 (16.0%)	73 (11.1%)	337 (14.6%)
Insuficiència cardíaca	1614 (98.1%)	645 (98.2%)	2259 (98.1%)	32 (1.9%)	12 (1.8%)	44 (1.9%)
Hipertensió arterial	1470 (89.3%)	572 (87.1%)	2042 (88.7%)	176 (10.7%)	85 (12.9%)	261 (11.3%)
Dislipèmia	1528 (92.8%)	606 (92.2%)	2134 (92.7%)	118 (7.2%)	51 (7.8%)	169 (7.3%)
Diabetis mellitus 2	1540 (93.6%)	620 (94.4%)	2160 (93.8%)	106 (6.4%)	37 (5.6%)	143 (6.2%)
Fumador	1509 (91.7%)	631 (96.0%)	2140 (92.9%)	137 (8.3%)	26 (4.0%)	163 (7.1%)
Ex-fumador	1615 (98.1%)	651 (99.1%)	2266 (98.4%)	31 (1.9%)	6 (0.9%)	37 (1.6%)

Per tal d'estudiar la distribució dels individus segons el nombre de problemes que pateixen s'ha creat una nova variable que conté la suma de problemes per individu. A la taula 13 es mostren els resultats obtinguts. Destacar que existeixen 7 individus que pateixen 6 dels 8 problemes estudiats (tots homes); i 16 que en pateixen 5 (15 homes i una dona).

Taula 13. Distribució dels individus segons sexe i nº de problemes que pateixen.

Nº de problemes	SEXE		Total
	Home	Dona	
0	1210 (73.5%)	496 (75.2%)	1704 (74.0%)
1	171 (10.4%)	71 (10.8%)	242 (10.5%)
2	121 (7.4%)	61 (9.3%)	182 (7.9%)
3	86 (5.2%)	22 (3.3%)	108 (4.7%)
4	36 (2.2%)	8 (1.2%)	44 (1.9%)
5	15 (0.9%)	1 (0.2%)	16 (0.7%)
6	7 (0.4%)	0 (0.0%)	7 (0.3%)
Total	1646 (71.5%)	657 (28.5%)	2303 (100.0%)

A partir d'ara es passa a analitzar els diferents ingressos. Tal com s'ha dit a l'inici d'aquest apartat, es disposa de 3276 ingressos.

En gairebé el 80% del total d'ingressos (concretament 2603), la circumstància d'admissió ha estat una urgència. La resta (668 ingressos) corresponen a admissions programades. Per sexe els percentatges rondan els mateixos valors que els globals (taula 14).

Taula 14. Distribució dels individus segons sexe i circumstància d'ingrés.

	SEXE		Total
	Home	Dona	
Admissió urgent	1922 (80.0%)	681 (78.3%)	2603 (79.6%)
Admissió programada	479 (20.0%)	189 (21.7%)	668 (20.4%)
Total	2401 (73.4%)	870 (26.6%)	3271 (100.0%)

*El nombre total d'individus no coincideix en totes les taules degut als valors mancants.

La circumstància d'alta més freqüent, en un 88.8% dels casos, és la destinació al propi domicili (taula 15). La defunció del pacient és la segona causa d'alta en un 5.3% del total dels casos. Per sexes la destinació al propi domicili és la circumstància d'alta més freqüent seguida, també, de la defunció; les dues circumstàncies amb un percentatge entorn dels globals esmentats.

Taula 15. Distribució dels individus segons sexe i circumstància d'alta.

	SEXE		Total
	Home	Dona	
Destinació al propi domicili	2115 (88.1%)	790 (90.8%)	2905 (88.8%)
Trasllat a un hospital general d'aguts	79 (3.3%)	24 (2.8%)	103 (3.1%)
Trasllat a un centre de mitjana o llarga durada	38 (1.6%)	11 (1.3%)	49 (1.5%)
Trasllat a una residència social	29 (1.2%)	8 (0.9%)	37 (1.1%)
Alta voluntària	1 (0.0%)	0 (0.0%)	1 (0.0%)
Defunció	137 (5.7%)	37 (4.3%)	174 (5.3%)
Evasió o alta administrativa	1 (0.0%)	0 (0.0%)	1 (0.0%)
Hospitalització domiciliària	2 (0.1%)	0 (0.0%)	2 (0.1%)
Total	2402 (73.4%)	870 (26.6%)	3272 (100.0%)

La següent variable analitzada és la procedència del pacient en el moment d'ingressar. A la taula 16 s'observa com la majoria dels ingressats (81.4%) provenien del propi domicili seguit, amb força diferència, del trasllat des de la consulta externa del mateix hospital (14.7%). Per sexes aquesta tendència es manté amb uns valors molt semblants a aquests globals.

Taula 16. Distribució dels individus segons sexe i procedència del ingrés.

	SEXE		Total
	Home	Dona	
Domicili o residència social	1975 (82.4%)	683 (78.6%)	2658 (81.4%)
Trasllat des d'un hospital general d'aguts	25 (1.0%)	14 (1.6%)	39 (1.2%)
Trasllat des d'un recurs de la xarxa sociosanitària	11 (0.5%)	11 (1.3%)	22 (0.7%)
Trasllat des d'un centre d'atenció primària	44 (1.8%)	23 (2.6%)	67 (2.1%)
Trasllat des de consulta externa del mateix hospital	341 (14.2%)	138 (15.9%)	479 (14.7%)
Total	2396 (73.4%)	869 (26.6%)	3265 (100.0%)

A continuació s'ha analitzat el Grup Relacionat amb el Diagnòstic (GRD) de cada ingrés i el pes relatiu de cada un d'aquests GRD.

Començant per el pes relatiu, a la taula 17 s'observa com globalment la mitjana del pes relatiu és pràcticament la unitat (0.9862), la qual cosa indica que el cost mitjà global és lleugerament inferior a la mitjana del conjunt global de tots els GRD.

Aquesta característica es manté per sexes, només destacar que en el cas de les dones es distancia més de la unitat. El valor mínim és 0 i el màxim és 3.6936 tan globalment com per sexes. La moda es troba en el punt 0.9317, tant globalment com per sexes, per tant el valor més freqüent també es troba per sota la unitat.

Taula 17. Descriptius del pes relatiu del GRD globals i segons sexe.

	SEXE		Descriptius Globals
	Home	Dona	
Mitjana	0.9993	0.9499	0.9862
Mediana	0.9317	0.9317	0.9317
Mitjana retallada el 5%	0.9811	0.9155	0.9638
Desviació típica	0.2723	0.3506	0.2959
Moda	0.9317	0.9317	0.9317
Mínim	0.0000	0.0000	0.0000
Màxim	3.6936	3.6936	3.6936
Quartils			
Q1	0.9317	0.7871	0.9317
Q3	1.0647	1.0130	1.0647

A la taula 18 es presenten els 20 GRD més freqüents. Tal com és d'esperar, el més comú és la malaltia que s'està estudiant, MPOC, en un 44.4% del total dels casos. La resta estan molt relacionats amb malalties respiratòries i pulmonars.

Per sexes continua essent la MPOC el GRD més comú, però el percentatge d'homes és força més elevat que el de dones, un 48% i un 34.5% respectivament.

Taula 18. Distribució dels individus segons sexe i grup relacionat amb el diagnòstic.

	SEXE		Total
	Home	Dona	
088 MALALTIA PULMONAR OBSTRUCTIVA CRONICA	1153 (48.0%)	301 (34.5%)	1454 (44.4%)
089 PNEUMONIA SIMPLE/PLEURITIS >17a a/cc	304 (12.6%)	66 (7.6%)	370 (11.3%)
127 INSUFICIENCIA CARDIACA i XOC	167 (6.9%)	81 (9.3%)	248 (7.6%)
096 BRONQUITIS/ASMA >17a a/cc	53 (2.2%)	125 (14.3%)	178 (5.4%)
082 NEOPLASIES RESPIRATORIES	114 (4.7%)	7 (0.8%)	121 (3.7%)
098 BRONQUITIS/ASMA 0-17a	64 (2.7%)	56 (6.4%)	120 (3.7%)
097 BRONQUITIS/ASMA >17a s/cc	22 (0.9%)	72 (8.3%)	94 (2.9%)
140 ANGOR PECTORIS	62 (2.6%)	6 (0.7%)	68 (2.1%)
087 EDEMA PULMONAR/INSUFICIENCIA RESPIRATORIA	36 (1.5%)	31 (3.6%)	67 (2.0%)
138 ARITMIES CARDIAQUES/TRASTORNS CONDUCCIO a/cc	47 (2.0%)	12 (1.4%)	59 (1.8%)
079 INFECCIONS/INFLAMACIONS RESPIRAT >17a a/cc	50 (2.1%)	4 (0.5)	54 (1.6%)
122 TRAST CIRCULAT a/IAM s/COMPL CVASC, ALTA VIU	41 (1.7%)	7 (0.8%)	48 (1.5%)
144 ALTRES DIAGNOSTICS CIRCULATORIS a/cc	31 (1.3%)	11 (1.3%)	42 (1.3%)
130 TRASTORNS VASCULARS PERIFERICIS a/cc	30 (1.2%)	7 (0.8%)	37 (1.1%)
121 TRAST CIRCULAT a/IAM/COMPL CVASC, ALTA VIU	26 (1.1%)	10 (1.1%)	36 (1.1%)
099 SIGNES/SIMPTOMES RESPIRATORIS a/cc	25 (1.0%)	9 (1.0%)	34 (1.0%)
143 DOLOR TORACIC	19 (0.8%)	6 (0.7%)	25 (0.8%)
094 PNEUMOTORAX a/cc	20 (0.8%)	4 (0.5%)	24 (0.7%)
101 ALTRES DIAGNOSTICS RESPIRATORIS a/cc	18 (0.7%)	6 (0.7%)	24 (0.7%)
475 DIAG PPAL RESPIRATORI a/VENTILACIO MECANICA	11 (0.5%)	9 (1.0%)	20 (0.6%)

L'estada mitjana és de 7 dies i la mediana és de 6 per tant es pot dir que la distribució de la variable estada és una mica asimètrica a la dreta amb algun possible valor atípic (taula 19). L'estada màxima és de 64 dies i la mínima de 0; això pot ser degut a individus que han estat ingressats i traslladats abans de produir un ingrés, o també individus que s'hagin mort. Per sexes s'observa que l'estada mitjana de les dones és una mica superior a la dels homes tot i que la resta d'estadístics es mantenen força semblants, excepte l'estada màxima que en els homes és de 64 dies i per les dones de 41.

Taula 19. Descriptius de l'estada globals i per sexe.

	SEXE		Descriptius Globals
	Home	Dona	
Mitjana	6.94	7.12	6.99
Mediana	6.00	6.00	6.00
Mitjana retallada el 5%	6.41	6.61	6.46
Desviació típica	5.30	5.16	5.27
Moda	4.00	4.00	4.00
Mínim	0.00	0.00	0.00
Màxim	64.00	41.00	64.00
Quartils			
Q1	3.00	4.00	3.00
Q3	9.00	9.00	9.00

Per tal d'analitzar el nombre d'ingressos que s'han generat durant l'any 2003 degut a la malaltia pulmonar obstructiva crònica s'ha construït una nova variable on apareix la suma d'ingressos per individus. A la taula 20 es mostren els resultats d'aquest anàlisi.

Taula 20. Distribució del nombre d'individus segon el nombre d'ingressos i el sexe.

	Nº d'individus		
	Home	Dona	Total
1 ingrés	1160 (70.5%)	519 (79.0%)	1679 (72.9%)
2 ingressos	324 (19.7%)	92 (14.0%)	416 (18.1%)
3 ingressos	99 (6.0%)	26 (4.0%)	125 (5.4%)
4 ingressos	35 (2.1%)	13 (2.0%)	48 (2.1%)
5 ingressos	17 (1.0%)	4 (0.6%)	21 (0.9%)
6 ingressos	7 (0.4%)	2 (0.3%)	9 (0.4%)
7 ingressos	1 (0.1%)	1 (0.2%)	2 (0.1%)
8 ingressos	2 (0.1%)	0 (0.0%)	2 (0.1%)
9 ingressos	1 (0.1%)	0 (0.0%)	1 (0.0%)
Total	1646 (71.5%)	657 (28.5%)	2303 (100.0%)

6.1.2. Resum de l'anàlisi dels individus segons el nombre d'ingressos.

En aquest apartat es presenta mitjançant una única taula (taula 21), un resum de l'anàlisi realitzat dels individus segons el nombre d'ingressos que han tingut durant l'any d'estudi, o sigui durant l'any 2003. L'anàlisi complet es pot consultar a l'Annex VI.

En aquesta taula (taula 21) es pot observar com el percentatge d'homes és sempre superior al de les dones i que aquest percentatge augmenta amb el nombre d'ingressos, és a dir, que els homes han realitzat més ingressos que les dones.

Respecte a la variable edat es pot destacar que l'edat mitjana ronda sempre els 70 anys i que la mediana és lleugerament superior a aquesta edat mitjana en tots els casos excepte en el quart ingrés.

Per altra banda, el problema de salut més freqüent és, en tots els casos, la malaltia pulmonar obstructiva crònica seguit, en alguns casos, de la hipertensió arterial i en alguns altres de la dislipèmia.

La circumstància d'ingrés més comú és sempre l'admissió urgent i la circumstància d'alta, la destinació al propi domicili. La procedència de l'ingrés més freqüent és el domicili o residència social també en tots els casos.

Quan al grup relacionat amb el diagnòstic més freqüent és la malaltia pulmonar obstructiva crònica i el pes relatiu es troba sempre entorn de 1. En els casos que és inferior a 1 indica que el cost mitjà global és inferior a la mitjana del cost global de tots els GRD, i quan és superior, indica que aquest cost mitjà és superior a la mitjana del cost global.

Finalment es pot veure com l'estada mitjana augmenta amb el nombre d'ingressos igual que la moda i que l'estada mínima i màxima. Quan a la desviació típica d'aquesta variable es pot veure com sempre es manté més o menys entorn dels 5 dies.

Taula 21. Resum de l'anàlisi dels individus segons els nombre d'ingressos.

	Un ingrés	Dos ingressos	Tres ingressos	Quatre ingressos	Cinc ingressos
Sexe (homes)	69.1%	77.9%	79.2%	72.9%	81.0%
Edat (anys)					
<i>Mitjana</i>	67.58	71.38	70.9	73.04	69.81
<i>Mediana</i>	73	75	73	72.50	76
<i>Desviació</i>	20.32	15.73	13.75	9.16	17.56
<i>Moda</i>	78	78	65	71	62
Problemes de salut més comuns					
<i>MPOC</i>	14.8%	15.6%	11.2%	10.4%	15.0%
<i>Hipertensió arterial</i>	11.7%	11.8%	9.6%	10.4%	10.0%
<i>Dislipèmia</i>	7.6%	6.0%	7.2%	6.3%	5.0%
Circumstància d'ingrés					
<i>Admissió urgent</i>					
<i>Admissió programada</i>	79.1%	80.2%	78.3%	80.2%	84.8%
Circumstància d'alta més comú					
<i>Destinació al propi domicili</i>	89.7%	87.4%	87.2%	90.6%	89.5%
Procedència de l'ingrés més comú.					
<i>Domicili o residència social</i>	80.4%	80.7%	83.4%	82.8%	81.0%
GRD més freqüent					
<i>MPOC</i>	37.6%	45.7%	54.9%	56.8%	60.0%
Pes relatiu dels GRD					
<i>Mitjana</i>	0.9749	1.0119	1.0007	0.9729	0.9648
<i>Mediana</i>	0.9317	0.9317	0.9317	0.9317	0.9317
<i>Desviació</i>	0.2796	0.3329	0.3163	0.1564	0.1324
Estada (dies)					
<i>Mitjana</i>	6.70	7.09	7.32	7.79	8.00
<i>Mediana</i>	6.00	6.00	7.00	6.50	7.00
<i>Desviació</i>	5.23	5.47	4.84	5.36	5.64
<i>Moda</i>	3.00	5.00	6.00	4.00	6.00
<i>Mínim</i>	1.00	0.00	0.00	0.00	0.00
<i>Màxim</i>	64.00	45.00	29.00	27.00	30.00

6.1.3. Anàlisi del nombre de visites entre diferents ingressos

A continuació es mostra l'anàlisi del nombre de visites fetes a primària entre els diferents reingressos.

Es comença amb el nombre de visites que han realitzat els individus entre el primer i el segon ingrés. S'observa com la majoria dels individus, un 93.6%, no han realitzat cap visita a primària, només un 1.8% han estat visitats una vegada i un 1.2% dues vegades. Destacar els dos individus que han realitzat 12 visites i un individu que ha estat visitat fins a 11 vegades entre el primer i el segon ingrés.

Taula 22. Nombre de visites a primària entre el primer i el segon ingrés.

	Nº d'individus	%
0	2155	93.6
1	42	1.8
2	28	1.2
3	17	0.7
4	15	0.7
5	13	0.6
6	12	0.5
7	5	0.2
8	5	0.2
9	6	0.3
10	2	0.1
11	1	0.0
12	2	0.1
Total	2303	100.0

A les taules que segueixen (taules 23-28), es mostren el nombre de visites realitzades entre els diferents ingressos segons el nombre d'ingressos realitzats.

Taula 23. Nombre de visites a primària entre el segon i el tercer ingrés dels dividus amb tres reingressos.

	Nº d'individus	%
0	2272	98.7
1	8	0.3
2	9	0.4
3	7	0.3
4	2	0.1
5	2	0.1
7	1	0.0
8	1	0.0
11	1	0.0
Total	2303	100.0

Taula 24 Nombre de visites a primària entre el segon i el tercer ingrés dels individus amb quatre reingressos.

	Nº d'individus	%
0	2292	99.5
1	1	0.0
2	1	0.0
3	1	0.0
4	1	0.0
6	4	0.2
8	2	0.1
10	1	0.0
Total	2303	100.0

Taula 25. Nombre de visites a primària entre el tercer i el quart ingrés dels individus amb quatre reingressos.

	Nº d'individus	%
0	2291	99.5
1	12	0.5
Total	2303	100.0

Taula 26. Nombre de visites a primària entre el tercer i el quart ingrés dels individus amb cinc reingressos.

	Nº d'individus	%
0	2302	100.0
1	1	0.0
Total	2303	100.0

Taula 27. Nombre de visites a primària entre el quart i el cinquè ingrés dels individus amb cinc reingressos.

	Nº d'individus	%
0	2302	100.0
1	1	0.0
Total	2303	100.0

Taula 28. Nombre de visites a primària entre el cinquè i el sisè ingrés dels individus amb sis reingressos.

	Nº d'individus	%
0	2302	100.0
1	1	0.0
Total	2303	100.0

6.2. Resultats obtinguts en l'aplicació dels mètodes d'Andersen i de Goldstein sense efectes espacials.

Una de les innovacions que s'introdueixen tan a l'aplicació d'Andersen com a la de Goldstein, es troba en l'estimació del model. En aquest cas s'ha optat per utilitzar l'estadística bayesiana (capítol 5) perquè es treballa amb matrius de dimensió molt gran i amb l'aproximació freqüentista aquesta és una tasca inabordable amb mitjans computacionals.

Quan es modelitza la variabilitat espacial utilitzant l'estadística bayesiana, es solen escollir distribucions a priori per les desviacions típiques dels efectes aleatoris "vagament" o "dèbilment" informatives. El problema és que l'elecció d'una distribució a priori estàndard, és a dir, no informativa i "propia" (en el sentit que la seva integral sigui la unitat) implicaria que la inferència posterior depengués dels paràmetres de la distribució a priori. Aquesta (excessiva) sensibilitat és particularment problemàtica si

no es té suficient evidència de l'existència d'heterogeneïtat entre unitats. Per aquest motiu s'ha utilitzat una distribució a priori que no presenti aquest problema, una distribució uniforme en un rang finit (0 a 5) de la desviació típica tant per els efectes aleatoris espacials (dependència espacial) com per els efectes aleatoris desestructurats (heterogeneïtat no espacial) (Gelman, 2005; Barceló *et al.*, 2008).

Com a conseqüència de la seva forma iterativa, la convergència de les simulacions resulta crucial per validar els resultats obtinguts a partir dels mètodes bayesians complets (Barceló *et al.*, 2008). Per aquest motiu s'utilitzen tres cadenes per cada paràmetre a estimar i s'executen un gran nombre de simulacions (50.000 com a mínim), de les quals no es té en compte el 10% de les inicial (procediment anomenat *burn-in*) i per cada cert nombre d'iteracions es guarden els paràmetres estimats, descartant la resta d'iteracions amb la finalitat d'evitar l'autocorrelació que es dona a les cadenes (procediment anomenat *thinning*).

Després d'obtenir els resultats del model, es comprova la convergència mitjançant l'estadístic de Brooks-Gelman-Rubin (Brooks i Gelman, 1998) (anomenat *R-hat*) i la mida mostral de les cadenes (*n.eff*) (Carlin *et al.*, 1998 i Gelman *et al.*, 2004). Els criteris de convergència acordats van ser que el *R-hat* es trobés entre 0.95 i 1,05 i el número efectiu (*n.eff*) major que 100 en les estimacions de les desviacions típiques dels efectes aleatoris (Barceló *et al.*, 2008), ja que són els paràmetres més problemàtics quan a la convergència. Si no es complien aquests criteris s'augmentava el número de simulacions fins que es produís la convergència.

També es van utilitzar, a més a més, els entorns de software lliure R (versió 2.4.1) (R Development Core Team, 2007) i WinBUGS (Spiegelhalter *et al.*, 2003) i la llibreria de R, R2WinBUGS (Gelman *et al.*, 2007), que permet connectar els dos entorns.

6.2.1. Resultats pel mètode d'Andersen sense efectes espacials.

A la taula 29 es mostren els resultats obtinguts en el model final sense efectes espacials utilitzat per aplicar el mètode d'Andersen a l'arxiu de dades. Només han resultat ser significatives les variables número de recurrències i número de visites i les dues influeixen de forma negativa a la variable resposta. En aquest sentit, si s'ha produït més d'un ingrés, la probabilitat de reingressar és més petita que si només se n'ha produït un; i com més visites a primària menys probabilitat de reingressar.

Taula 29. Model final d'Andersen sense efectes espacials (model 4.10)

	OR (sd)	p-valor
Sexe (home)		
Dona	0.982 (0.019)	0.3588
Número de recurrències (una recurrència)		
Dues recurrències	0.164 (0.034)	0.0000
Tres recurrències	0.093 (0.072)	0.0000
Quatre recurrències	0.070 (0.126)	0.0000
Cinc recurrències	0.158 (0.129)	0.0000
Sis recurrències	0.087 (0.201)	0.0000
Edat (< 64 anys)		
64-72 anys	0.973 (0.027)	0.3083
72-77 anys	0.971 (0.027)	0.2663
77-82 anys	1.014 (0.028)	0.6257
>82 anys	1.017 (0.027)	0.5194
Alcohol (no)		
si	0.992 (0.063)	0.9061
MPOC (no)		
si	1.002 (0.028)	0.9288
Insuficiència cardíaca (no)		
si	0.985 (0.065)	0.8124
Hipertensió arterial (no)		
si	1.022 (0.032)	0.5017
Dislipèmia (no)		
si	0.960 (0.038)	0.2833
Diabetis Mellitus 2 (no)		
si	0.997 (0.040)	0.9347
Tabaquisme (no)		
si	0.957 (0.040)	0.2750
Exfumador (no)		
si	1.049 (0.069)	0.4857
Número de visites	0.984 (0.006)	0.0061

Variables significatives en negreta. Valors de referència entre parèntesis.

Tot i no ser estadísticament significativa s'ha trobat una relació positiva en el fet de patir algun dels següents problemes: malaltia pulmonar obstructiva crònica (MPOC), hipertensió arterial o ser exfumador; és a dir, si l'individu pateix algun d'aquests problemes té més probabilitat de tornar a ingressar a l'hospital que si no els pateix.

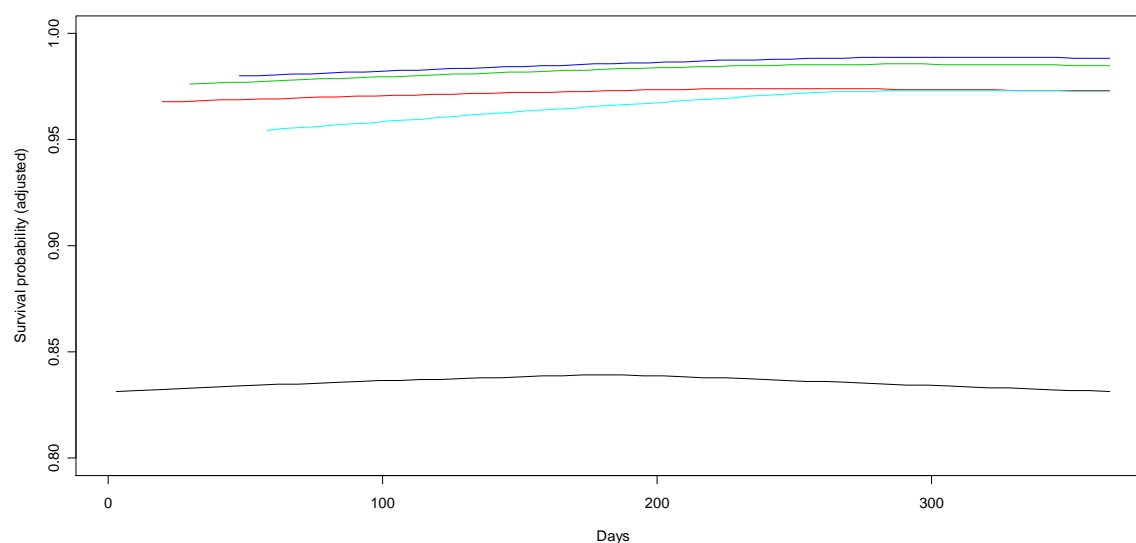
També s'ha trobat una relació positiva en el fet de pertànyer ens els dos grups d'edat més grans, així, els individus que tenen més de 77 anys tenen més probabilitat de

reingressar a l'hospital que els que tenen menys de 64. Per la resta de variables la relació que s'ha trobat amb la probabilitat de reingressar és negativa, és a dir, les dones tenen menys probabilitat de reingressar que els homes, si l'individu pateix alcoholisme, insuficiència cardíaca, dislipèmia, diabetis mellitus 2 o tabaquisme té menys probabilitat de reingressar a l'hospital que si no els pateix; i finalment, els individus que tenen menys de 77 anys tenen menys probabilitat d'ingressar que els que tenen menys de 64.

A continuació es mostren els gràfics més interessants del conjunt de gràfics realitzats a partir dels resultats obtinguts (veure Annex VII).

En el gràfic 4 es mostra la probabilitat de supervivència segons els nombre d'ingressos realitzats durant el període d'estudi. Cal observar que en el primer ingrés, la supervivència és inferior a la resta.

Gràfic 4. Probabilitat de supervivència per cada ingrés

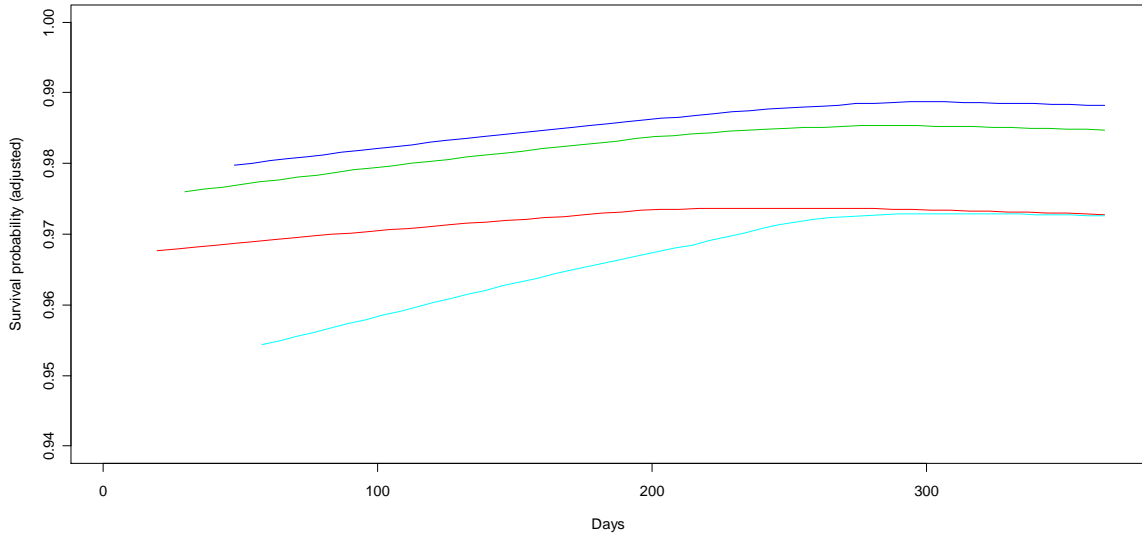


Negre:	primer	ingrés
Vermell:	segon	ingrés
Verd:	tercer	ingrés
Blau:	quart	ingrés
Turquesa:	cinquè	ingrés

Per tal d'estudiar millor la supervivència en els casos en que s'ha realitzat més d'un ingrés es mostra en el gràfic 5 les corbes de supervivència anteriors excloent la corba corresponent al primer ingrés. Dins aquest grup es veu com la supervivència en el

quart ingrés és sempre superior a la resta i el cinquè ingrés correspon a la supervivència més baixa.

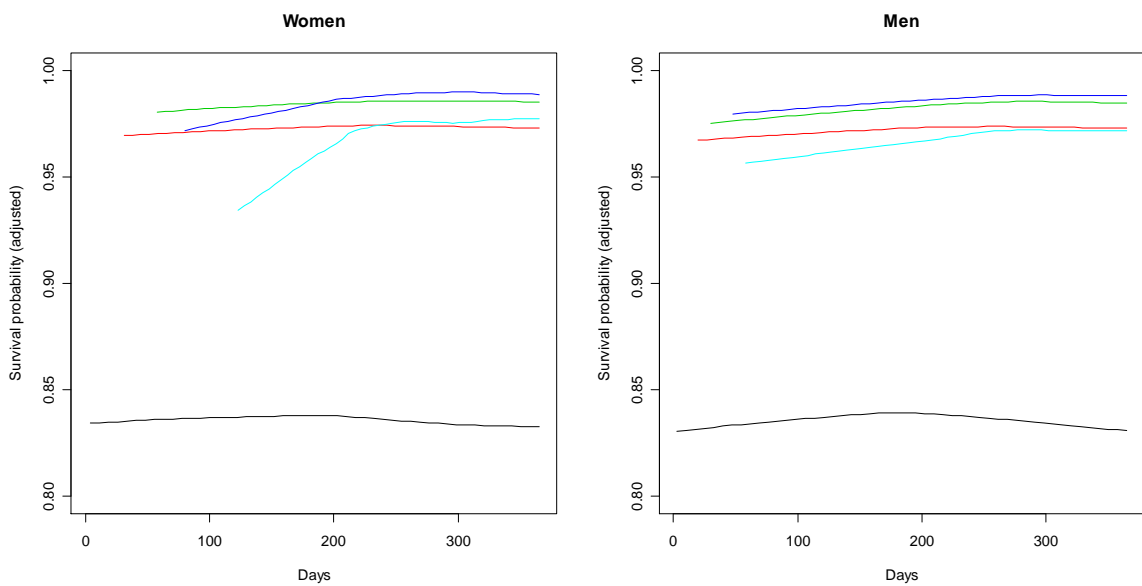
Gràfic 5. Probabilitat de supervivència pels casos amb més d'un ingrés



Vermell: segon ingrés Verd: tercer ingrés Blau: quart ingrés Turquesa: cinquè ingrés

En el gràfic 6 es mostra la supervivència segons els nombre d'ingressos separant els homes i les dones. No s'observen diferències significatives entre sexes i a més, en els dos casos es segueix la tendència general.

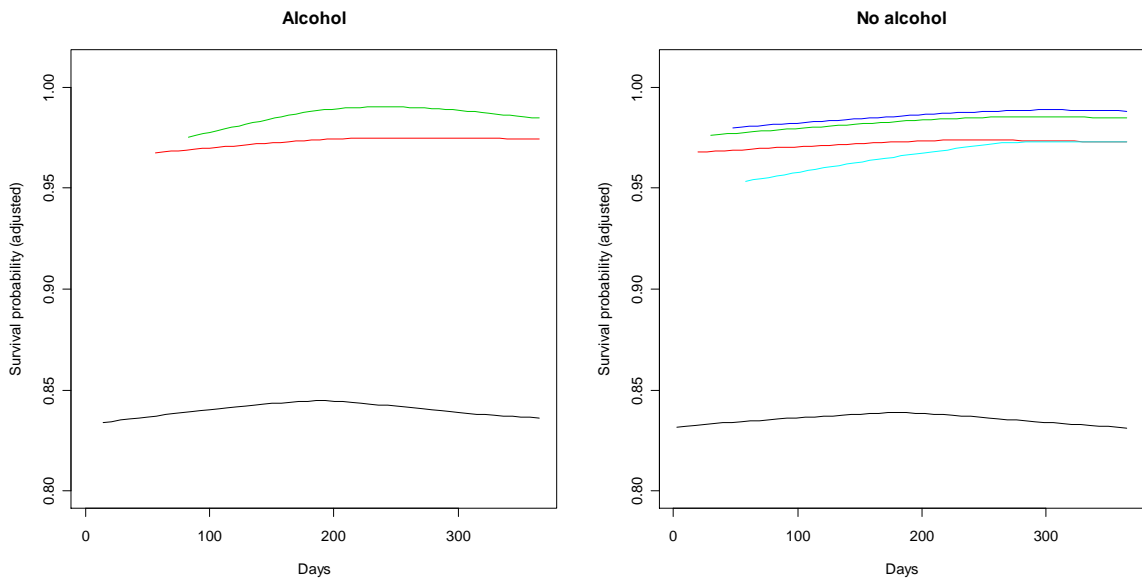
Gràfic 6. Probabilitat de supervivència segons el nombre d'ingressos i el sexe.



Negre: primer ingrés Vermell: segon ingrés Verd: tercer ingrés Blau: quart ingrés Turquesa: cinquè ingrés

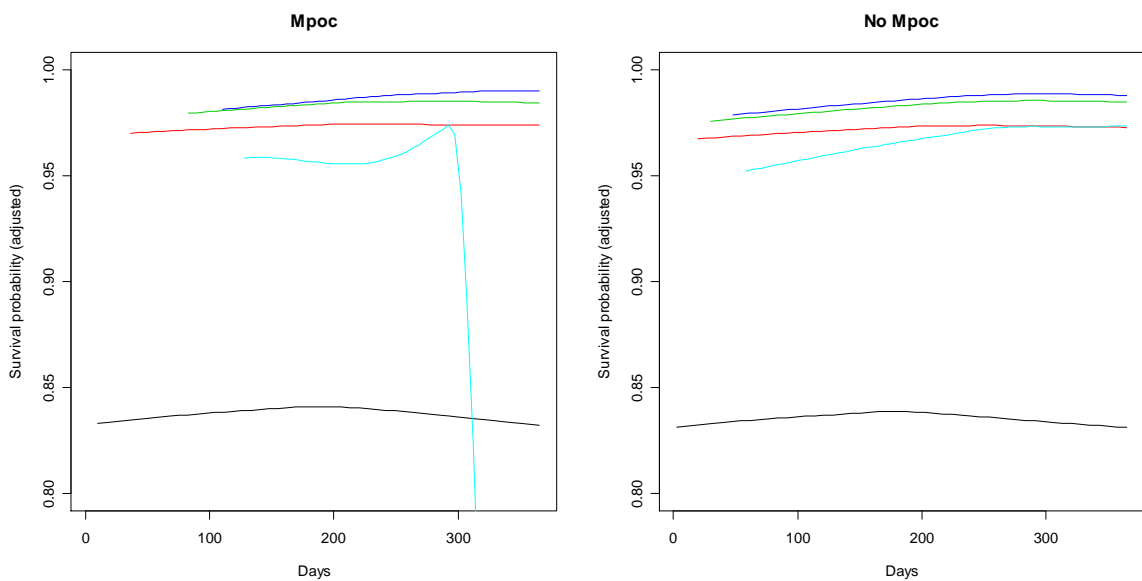
En la resta de gràfics (7-14) es mostra la probabilitat de supervivència segons els diferents problemes crònics que pot presentar el malalt.

Gràfic 7. Probabilitat de supervivència segons el nombre d'ingressos i alcohol.



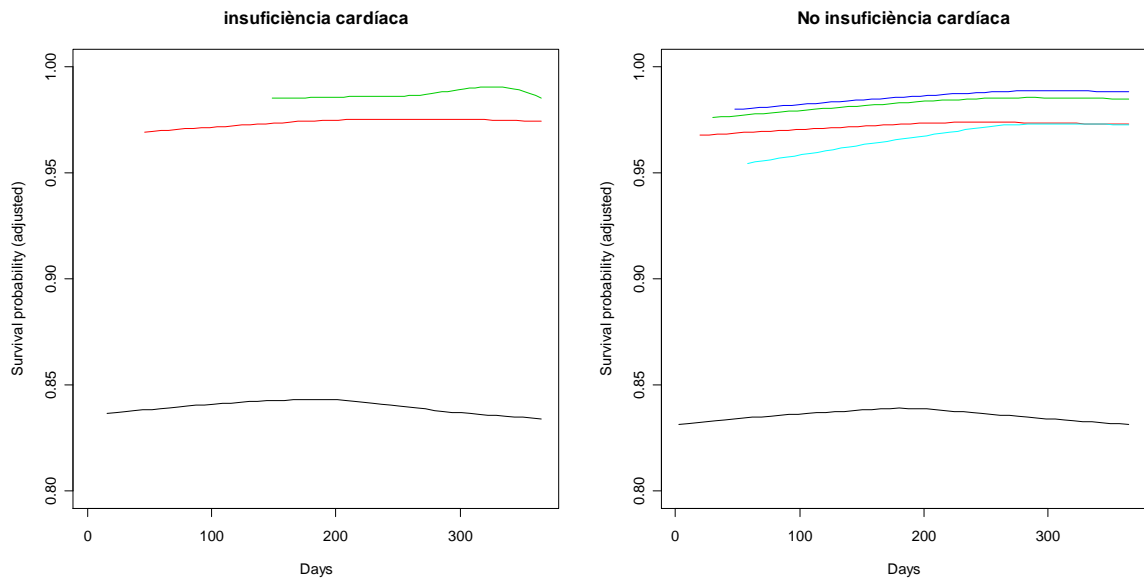
Negre: primer ingrés Vermell: segon ingrés Verd: tercer ingrés Blau: quart ingrés Turquesa: cinquè ingrés

Gràfic 8. Probabilitat de supervivència segons el nombre d'ingressos i mpoc.



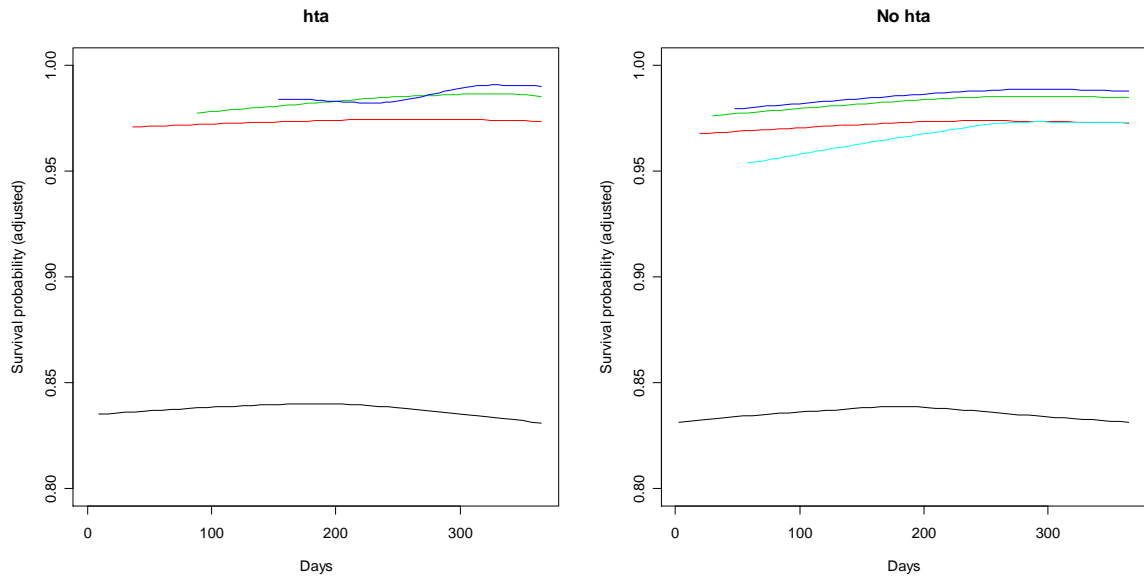
Negre: primer ingrés Vermell: segon ingrés Verd: tercer ingrés Blau: quart ingrés Turquesa: cinquè ingrés

Gràfic 9. Probabilitat de supervivència segons el nombre d'ingressos i insuficiència cardíaca.



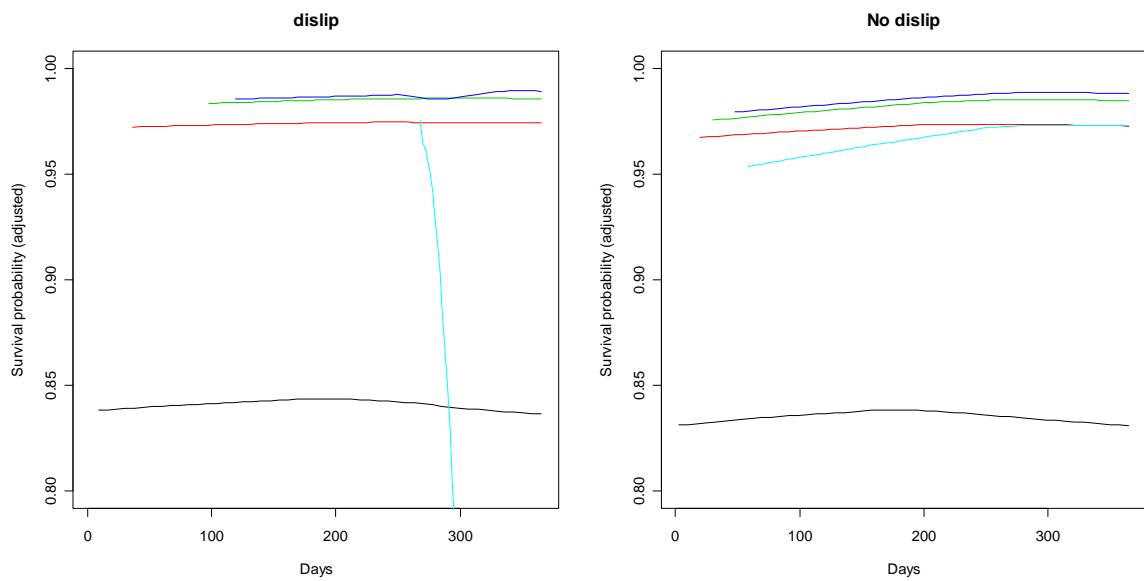
Negre: primer ingrés Vermell: segon ingrés Verd: tercer ingrés Blau: quart ingrés Turquesa: cinquè ingrés

Gràfic 10. Probabilitat de supervivència segons el nombre d'ingressos i hipertensió arterial.



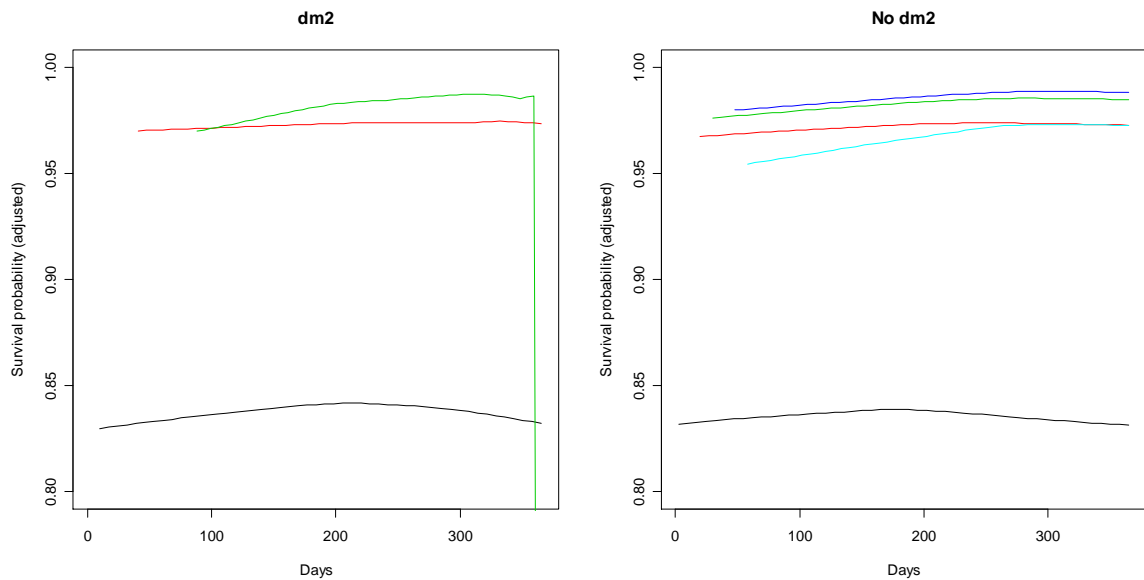
Negre: primer ingrés Vermell: segon ingrés Verd: tercer ingrés Blau: quart ingrés Turquesa: cinquè ingrés

Gràfic 11. Probabilitat de supervivència segons el nombre d'ingressos i dislipèmia.



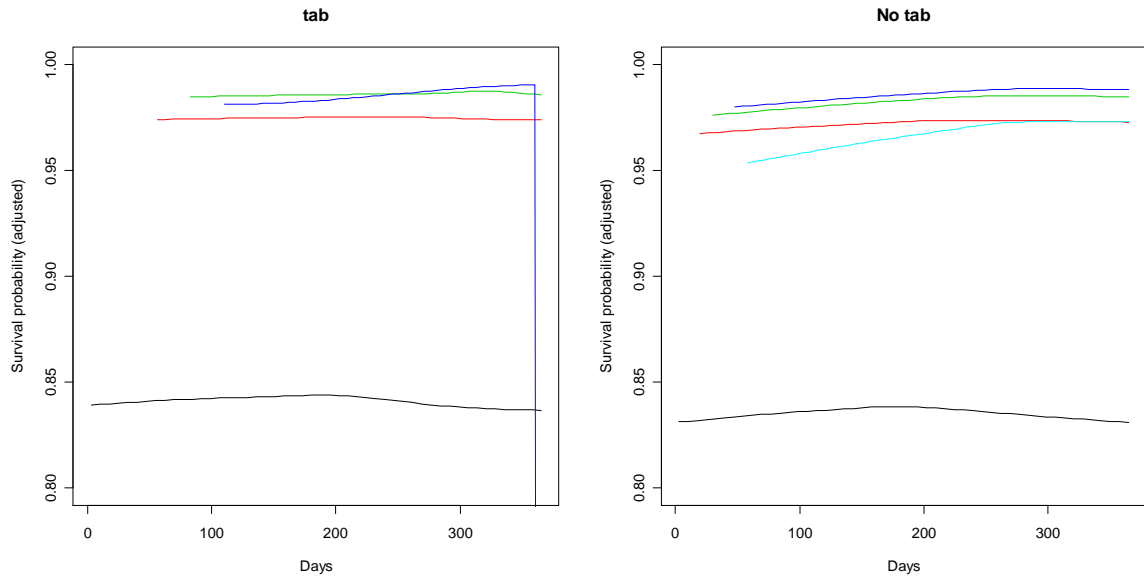
Negre: primer ingrés Vermell: segon ingrés Verd: tercer ingrés Blau: quart ingrés Turquesa: cinquè ingrés

Gràfic 12. Probabilitat de supervivència segons el nombre d'ingressos i diabetis mellitus 2.



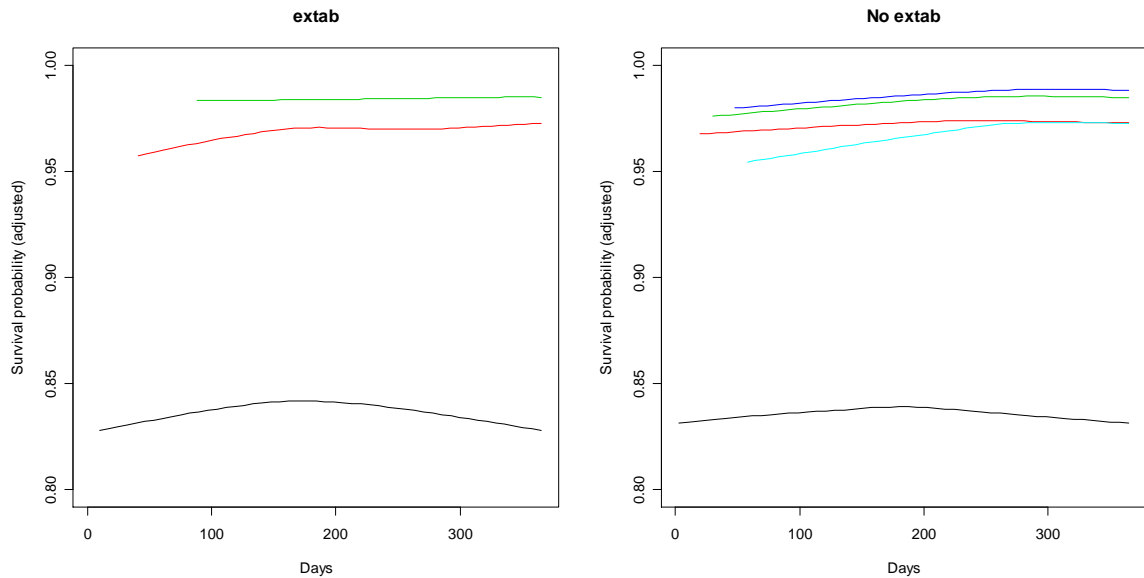
Negre: primer ingrés Vermell: segon ingrés Verd: tercer ingrés Blau: quart ingrés Turquesa: cinquè ingrés

Gràfic 13. Probabilitat de supervivència segons el nombre d'ingressos i tabac.



Negre: primer ingrés Vermell: segon ingrés Verd: tercer ingrés Blau: quart ingrés Turquesa: cinquè ingrés

Gràfic 14. Probabilitat de supervivència segons el nombre d'ingressos i extabac.



Negre: primer ingrés Vermell: segon ingrés Verd: tercer ingrés Blau: quart ingrés Turquesa: cinquè ingrés

6.2.2. Resultats pel mètode de Goldstein sense efectes espacials.

Per l'aplicació del mètode de Goldstein només es van tenir en compte els ingressos realitzats durant el primer trimestre de l'any 2003 ja que la construcció de la base de dades que requereix aquesta aplicació esdevenia intractable per la seva grandària.

Així, durant el primer trimestre del 2003 s'han comptabilitzat 753 individus que han ingressat com a mínim una vegada per MPOC, dels quals el 72.6% són homes i la resta dones.

Tal com es pot veure a la taula 30, la majoria dels pacients, un 83.7%, han ingressat només una vegada i la resta de dues a cinc vegades. El nombre d'ingressos presenta diferències estadísticament significatives entre ambdós sexes (p-valor: 0.008); observant-se que el grup d'homes mostra un nombre de reingressos més elevat que el de dones.

Taula 30. Distribució per sexes de les variables.

	Homes n(%)	Dones n(%)
N^a reingressos		
1	447(81.7)	183(88.8)
2	78(14.3)	20(9.7)
3	16(2.9)	2(1.0)
4	4(0.7)	1(0.5)
5	2(0.4)	0(0.0)
Consum alcohol		
Si	19(3.5)	0(0.0)
MPOC		
Si	85(15.5)	22(10.7)
Insuficiència cardíaca		
Si	12(2.2)	2(1.0)
Hipertensió arterial		
Si	56(10.2)	26(12.6)
Dislipèmia		
Si	36(6.6)	14(6.8)
Diabetis mellitus 2		
Si	38(6.9)	10(4.9)
Altres diabetis		
Si	2(0.4)	0(0.0)
Fumador		
Si	43(7.9)	2(1.0)
Ex-fumador		
Si	12(2.2)	2(1.0)

Respecte als problemes que pateixen els individus, a la mateixa taula 30 es pot observar que el més comú és MPOC en un 14.2% dels casos seguit de la hipertensió arterial en un 10.9%. Per sexes, els homes segueixen la tendència global ja que el problema més comú és MPOC (15.5%); en canvi en les dones el més comú és la hipertensió arterial (12.6%). Destacar també que el problema del tabac és molt més comú en homes (7.9%) que en dones (1.0%).

De mitjana els homes han ingressat 1.24 vegades (interval de confiança al 95%, IC 1.19-1.29) i les dones 1.13 (IC 1.08-1.19) (taula 31). L'edat mitjana global és de 71.04 anys (IC 69.90-72.19), i el nombre mig de visites realitzades durant aquest període és de 0.4 (IC 0.28-0.51).

Per sexes, l'edat mitjana de les dones és de 70.9 anys (IC 68.56-73.25) i pels homes 71.1 (IC 69.79-72.41); i respecte al nombre de visites, la mitjana per les dones és de 0.48 (IC 0.24-0.72) i per els homes de 0.37 (IC 0.24-0.49) (taula 31). Aquestes diferències entre sexes també són estadísticament significatives observant-se que l'edat mitjana en els homes és superior que en les dones (p-valor: 0.000) i que les dones realitzen més visites als serveis d'atenció primària que els homes (p-valor: 0.001).

Taula 31. Descriptius de les variables numèriques.

	Sexe		Total Mitjana (sd);mediana; [95% IC]
	Home Mitjana (sd);mediana; [95% IC]	Dona Mitjana (sd);mediana; [95% IC]	
Edat	71.10(15.61);75.00; [69.79-72.41]	70.90(17.08);75.50; [68.56-73.25]	71.04(16.01);75.00; [69.90-72.19]
Nº de visites	0.37(1.52);0.00; [0.24-0.49]	0.48(1.75);0.00; [0.24-0.72]	0.40(1.58);0.00; [0.28-0.51]
Nº ingressos	1.24(0.53);1.00; [1.19-1.29]	1.13(0.40);1.00; [1.08-1.19]	1.21(0.53);1.00; [1.17-1.25]

A la taula 32 es mostren els resultats obtinguts en el model final on només han resultat ser significatives les variables associades al temps de permanència dins de cada interval. Aquesta variable s'ha trencat en quatre parts obtenint una relació no lineal. En aquest sentit, si el temps de permanència es troba en la primera, en la tercera o en la quarta part, aquesta variable influeix de forma positiva a la variable resposta, és a dir, més probabilitat de reingressar; en canvi si el temps de permanència es troba en la segona part aquesta variable influeix de forma negativa.

Taula 32. Model final sense efectes espacials (model 4.18)

	OR (sd)	[95% IC]*
Sexe (home)		
Dona	0.897 (0.061)	[0.806- 1.015]
Temps de permanència en cada interval		
Primer break	2.002 (2.043)	[0.094- 40.987]
Segon break	0.161 (0.193)	[0.109- 0.229]
Tercer break	3.850(0.325)	[1.914- 6.568]
Quart break	5.562 (0.167)	[3.908- 7.706]
Edat (< 64 anys)		
64-72 anys	0.951 (0.096)	[0.780- 1.119]
72-77 anys	0.908 (0.098)	[0.738- 1.083]
77-82 anys	0.886 (0.095)	[0.730- 1.048]
>82 anys	0.916 (0.097)	[0.749- 1.083]
Alcohol (no)		
si	0.975 (0.186)	[0.669- 1.364]
MPOC (no)		
si	0.999 (0.815)	[0.838- 1.166]
Insuficiència cardíaca (no)		
si	1.068 (0.199)	[0.717- 1.586]
Hipertensió arterial (no)		
si	0.989 (0.102)	[0.821- 1.217]
Dislipèmia (no)		
si	1.011 (0.121)	[0.794- 1.257]
Diabetis Mellitus 2 (no)		
si	0.968 (0.115)	[0.772- 1.206]
Tabaquisme (no)		
si	0.838 (0.124)	[0.658- 1.073]
Exfumador (no)		
si	1.033 (0.195)	[0.724- 1.482]
Número de visites	0.492 (2.041)	[0.025- 10.581]
DIC		16125.300
AIC		16143.136
BIC		16193.446

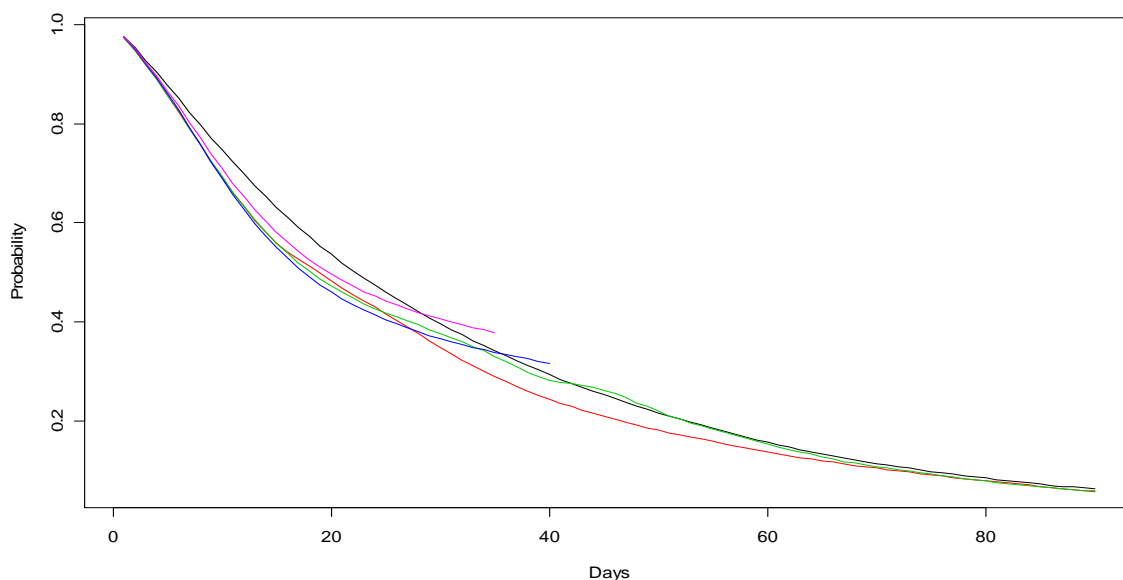
*Interval de confiança al 95%. Valors de referència entre parèntesis. Variables significatives en negreta

Tot i no ser estadísticament significativa s'ha trobat una relació positiva en el fet de patir algun dels següents problemes: insuficiència cardíaca, dislipèmia o ser exfumador; és a dir, si l'individu pateix algun d'aquests problemes té més probabilitat

de tornar a ingressar a l'hospital que si no els pateix. Per altra banda, per la resta de variables s'ha trobat una relació negativa amb la probabilitat d'ingressar, és a dir, les dones tenen menys probabilitat de reingressar que els homes, com més edat tingui l'individu menys probabilitat té de tornar a ingressar, si l'individu pateix algun dels següents problemes: alcoholisme, malaltia pulmonar obstructiva crònica (MPOC), hipertensió arterial, diabetis mellitus 2 o tabaquisme té menys probabilitat de reingressar a l'hospital que si no els pateix; i finalment, com més visites s'han realitzat a primària, menys probabilitat té l'individu de tornar a ingressar.

A continuació es mostren els gràfics més destacables d'entre tots els realitzats amb els resultats obtinguts (veure Annex VIII). En el gràfic 15 es mostra la probabilitat que es produeixi un canvi d'estat, és a dir, d'ingressar a no ingressar o a l'inrevés, segons el número d'ingrés. Primer cal observar que per els tres primers ingressos aquesta probabilitat existeix per a tot el període d'estudi (90 dies) mentre que per el quart i el cinquè ingrés només es tenen dades fins a la meitat del període. S'observa, també, que durant els primers trenta dies la probabilitat de canvi d'estat del primer ingrés és superior a la resta, i que el quart ingrés és el que té la probabilitat més baixa. Durant la resta de període, la probabilitat de canvi d'estat del primer ingrés i del tercer són molt properes i la del segon ingrés és la més baixa.

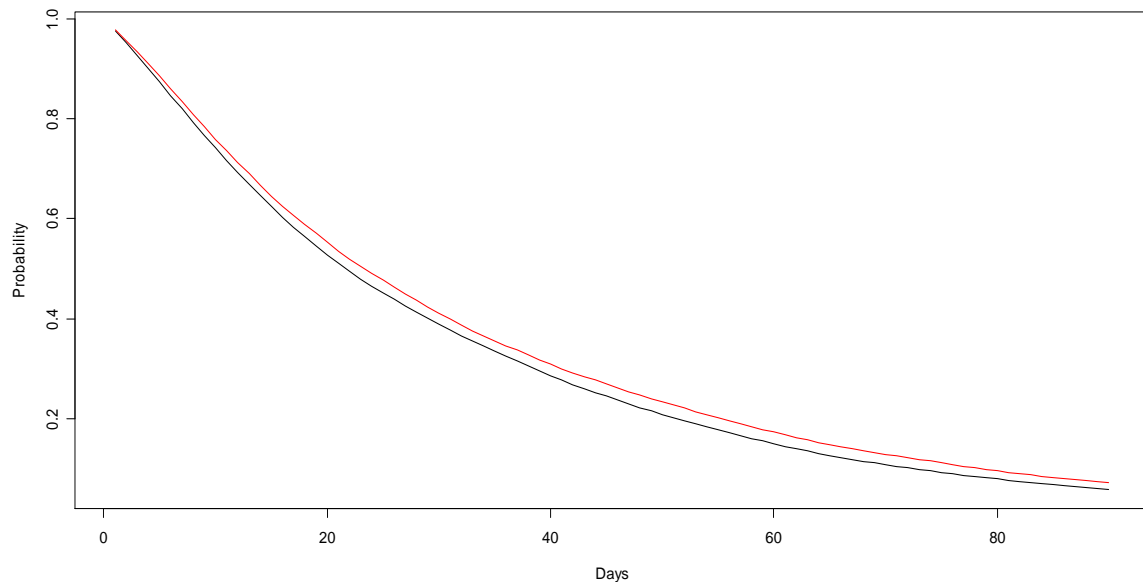
Gràfic 15. Probabilitat per cada ingrés.



Negre: primer ingrés Vermell: segon ingrés Verd: tercer ingrés Blau: quart ingrés Rosa: cinquè ingrés

En el gràfic 16 es mostra la probabilitat que es produeixi un canvi d'estat segons el sexe pel primer ingrés i es pot observar com les dones tenen una probabilitat superior als homes durant tot el període d'estudi.

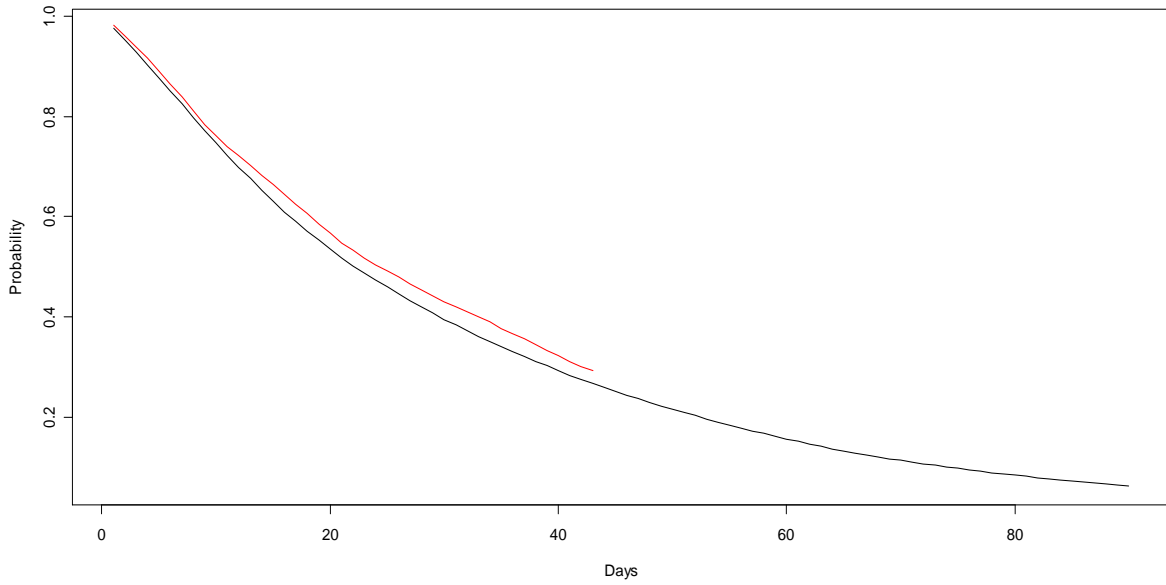
Gràfic 16. Probabilitat del primer ingrés per sexe.



Negre: homes Vermell: dones

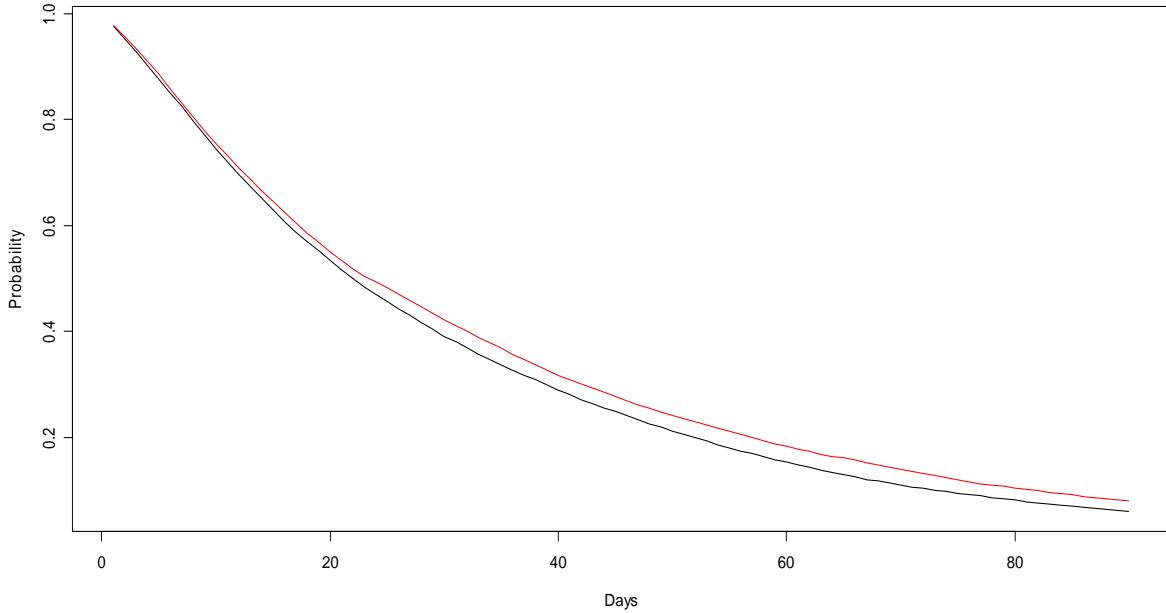
A la resta de gràfics (gràfics 17-21) es pot veure com la probabilitat que es produeix un canvi d'estat pel primer ingrés segons alguns dels problemes que pot patir l'individu (alcohol, mpoc, hipertensió arterial, insuficiència cardíaca i tabac) segueix en tots la mateixa tendència, si l'individu pateix el problema té més probabilitat que es produeixi el canvi d'estat que si no el pateix (encara que cap relació sortís estadísticament significativa).

Gràfic 17. Probabilitat del primer ingrés segons si es pateix el problema alcohol.



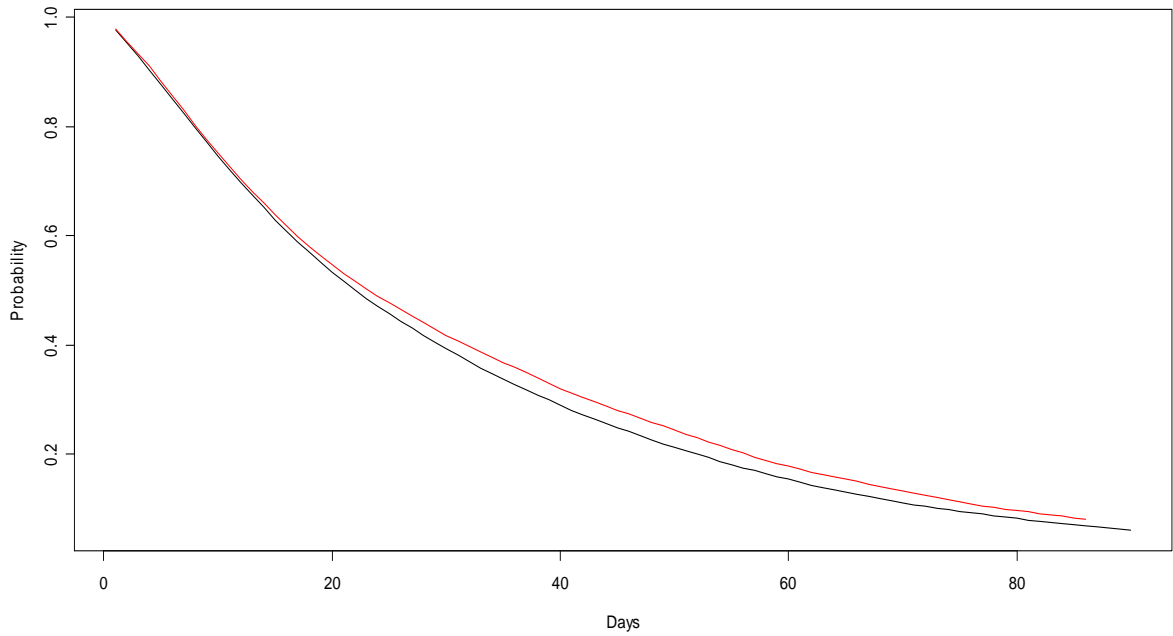
Negre: no Vermell: si

Gràfic 18. Probabilitat del primer ingrés segons si es pateix el problema mpoc.



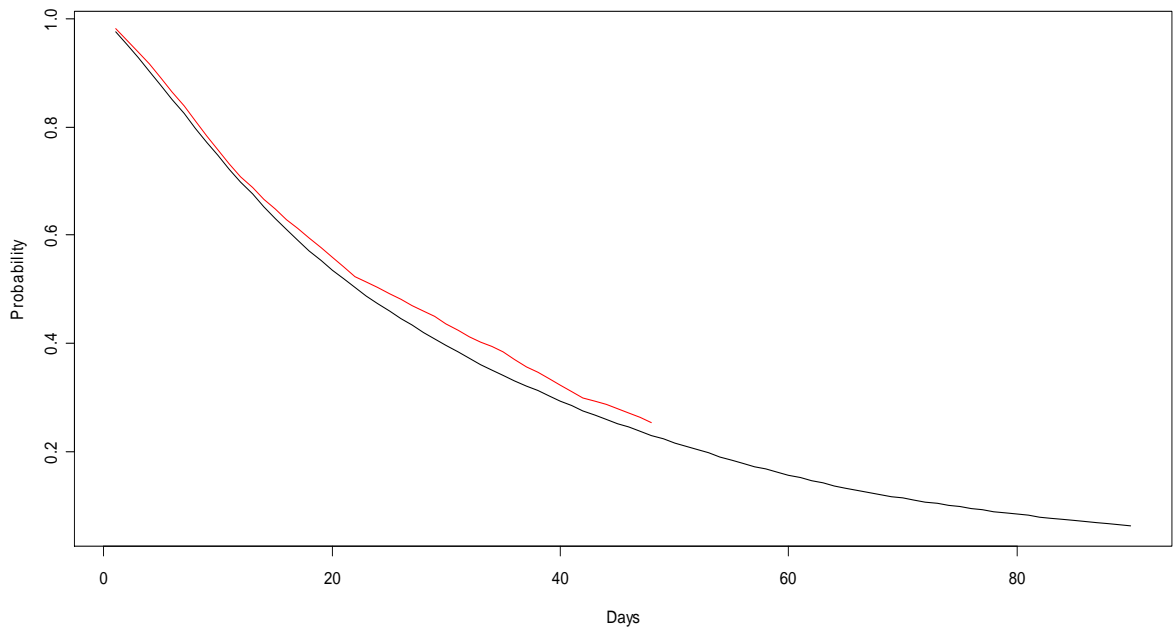
Negre: no Vermell: si

Gràfic 19. Probabilitat del primer ingrés segons si es pateix el problema hipertensió arterial.



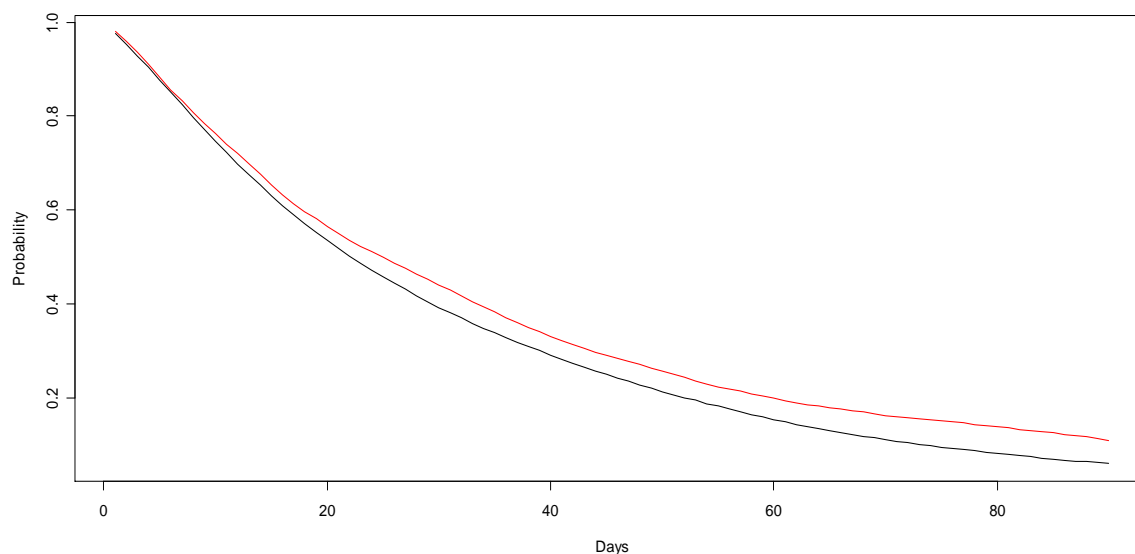
Negre: no
Vermell: si

Gràfic 20. Probabilitat del primer ingrés segons si es pateix el problema insuficiència cardíaca.



Negre: no
Vermell: si

Gràfic 21. Probabilitat del primer ingrés segons si es pateix el problema tabac.



Negre: no
Vermell: si

6.3. Resultats obtinguts en l'aplicació dels mètodes d'Andersen i de Goldstein amb efectes espacials.

Una altra de les innovacions que presenta aquest treball és la introducció de la variació espacial en els models utilitzats. Després d'introduir aquests efectes i tornar a estimar els models, els resultats que s'obtenen no varien pràcticament gens dels vistos a l'apartat anterior.

En l'aplicació de Goldstein les variables significatives continuen essent les associades al temps de permanència dins de cada interval (taula 32). En aquest cas si el temps de permanència es troba en la primera o en la segona part, la influència d'aquesta variable sobre la variable resposta és negativa; és a dir, menys probabilitat de reingressar; en canvi, si es troba en la tercera o la quarta part, la influència és positiva. Això es pot interpretar com que més temps ingressat implica més gravetat i per tant més probabilitat de tornar a ingressar.

Per la resta de variables, tot i no ser estadísticament significatives, es pot dir que influeix de forma positiva sobre la variables resposta el fet de patir insuficiència cardíaca o ser exfumador, així com el fet d'haver realitzat més visites a primària. La

variable edat també influeix de forma positiva en tots els seus trams excepte en el tercer, o sigui, es pot dir que en general com més edat més probabilitat de reingressar excepte en l'interval de 77 a 82 anys que la influència és negativa. En canvi, per la resta de variables s'ha trobat una relació negativa amb la probabilitat de reingressar, és a dir, les dones tenen menys probabilitat de reingressar que els homes i el fet de patir alcoholisme, MPOC, hipertensió arterial, dislipèmia, diabetis mellitus 2 o tabaquisme també disminueix la probabilitat de reingrès

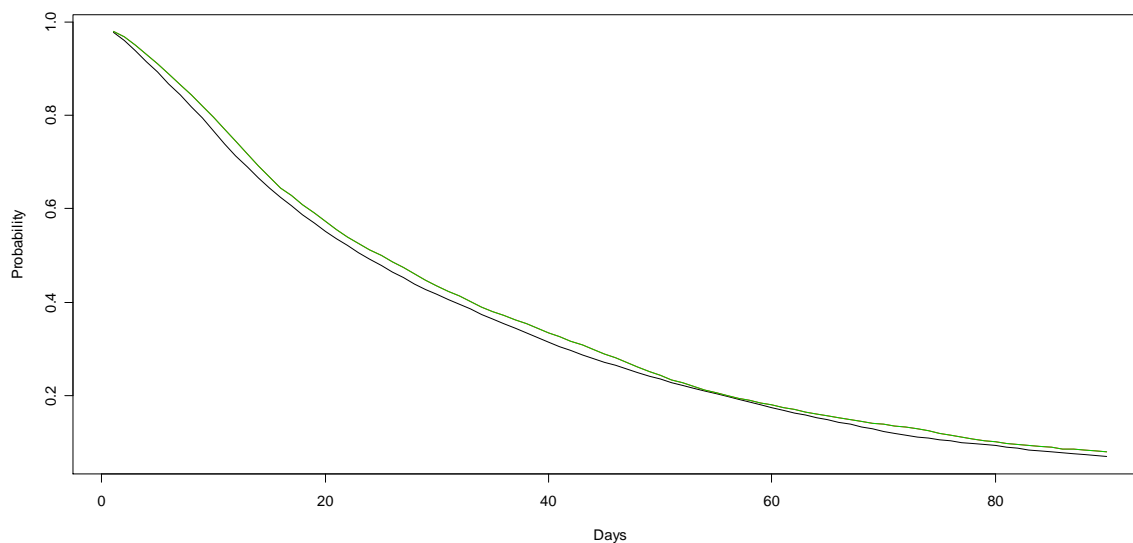
Taula 32. Model final amb efectes espacials (model 5.4)

	OR (sd)	[95% IC]*
Sexe (home)		
Dona	0.931 (0.062)	[0.823- 1.048]
Temps de permanència en cada interval		
Primer break	0.477 (0.900)	[0.111- 2.936]
Segon break	0.193 (0.151)	[0.143- 0.261]
Tercer break	5.409(0.208)	[3.532- 8.103]
Quart break	5.268 (0.160)	[3.830- 7.330]
Edat (< 64 anys)		
64-72 anys	1.047 (0.085)	[0.883- 1.244]
72-77 anys	1.000 (0.089)	[0.837- 1.194]
77-82 anys	0.986 (0.087)	[0.833- 1.162]
>82 anys	1.011 (0.084)	[0.860- 1.183]
Alcohol (no)		
si	0.981 (0.204)	[0.653- 1.464]
MPOC (no)		
si	0.901 (0.088)	[0.763- 1.072]
Insuficiència cardíaca (no)		
si	1.346 (0.211)	[0.903- 2.040]
Hipertensió arterial (no)		
si	0.937 (0.113)	[0.744- 1.174]
Dislipèmia (no)		
si	0.967 (0.135)	[0.738- 1.274]
Diabetis Mellitus 2 (no)		
si	0.951 (0.128)	[0.746- 1.236]
Tabaquisme (no)		
si	0.806 (0.143)	[0.604- 1.064]
Exfumador (no)		
si	1.288 (0.220)	[0.842- 1.945]
Número de visites	2.038 (0.900)	[0.335- 8.750]

*Interval de confiança al 95%. Valors de referència entre parèntesis. Variables significatives en negreta

Respecte als gràfics més destacables (veure Annex VIII per la resta), es comença mostrant un gràfic de la probabilitat que es produeixi un canvi d'estat segons el sexe pel primer ingrés (gràfic 22). Es pot observar com les dones tenen una probabilitat lleugerament superior als homes durant tot el període d'estudi.

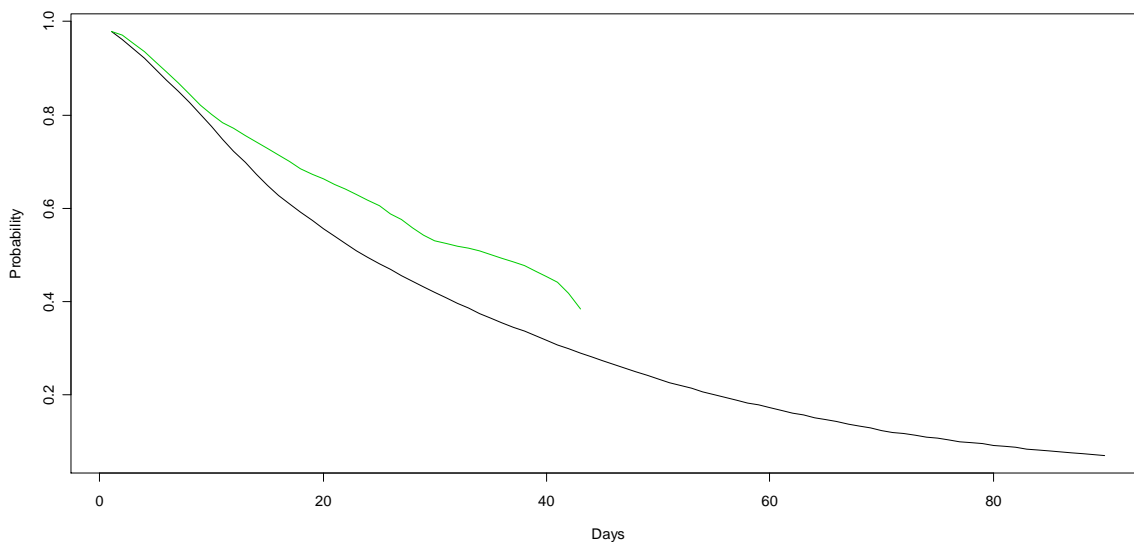
Gràfic 22. Probabilitat del primer ingrés per sexe.



Negre: homes Verd: dones

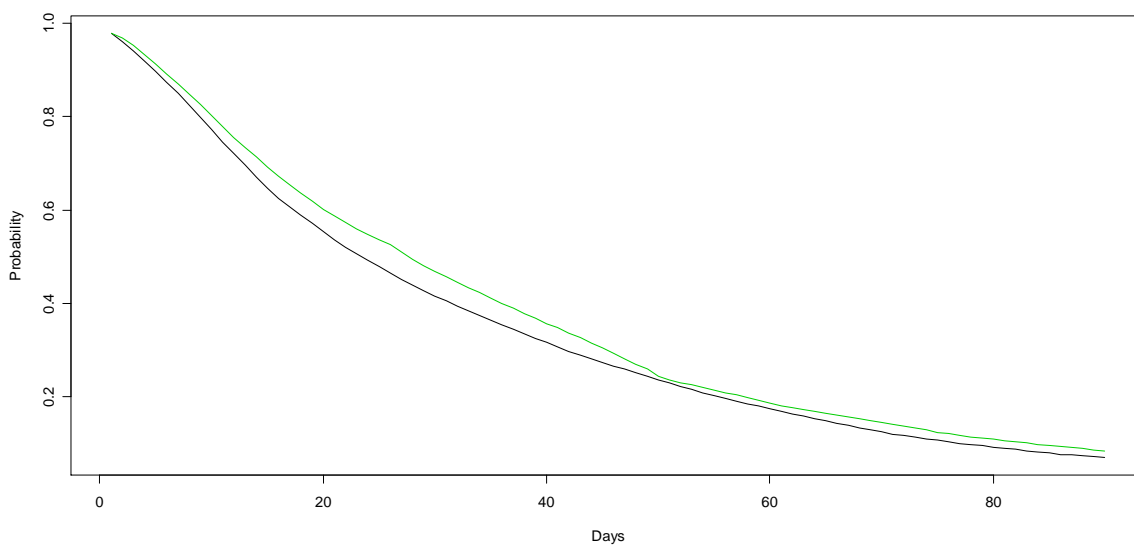
A la resta de gràfics (gràfics 23-27) es pot veure com la probabilitat que es produeix un canvi d'estat pel primer ingrés segons alguns dels problemes que pot patir l'individu (alcohol, mpoc, hipertensió arterial, insuficiència cardíaca i tabac) segueix en tots la mateixa tendència, si l'individu pateix el problema té més probabilitat que es produeixi el canvi d'estat que si no el pateix (encara que cap relació sortís estadísticament significativa).

Gràfic 23. Probabilitat del primer ingrés segons si es pateix el problema alcohol.



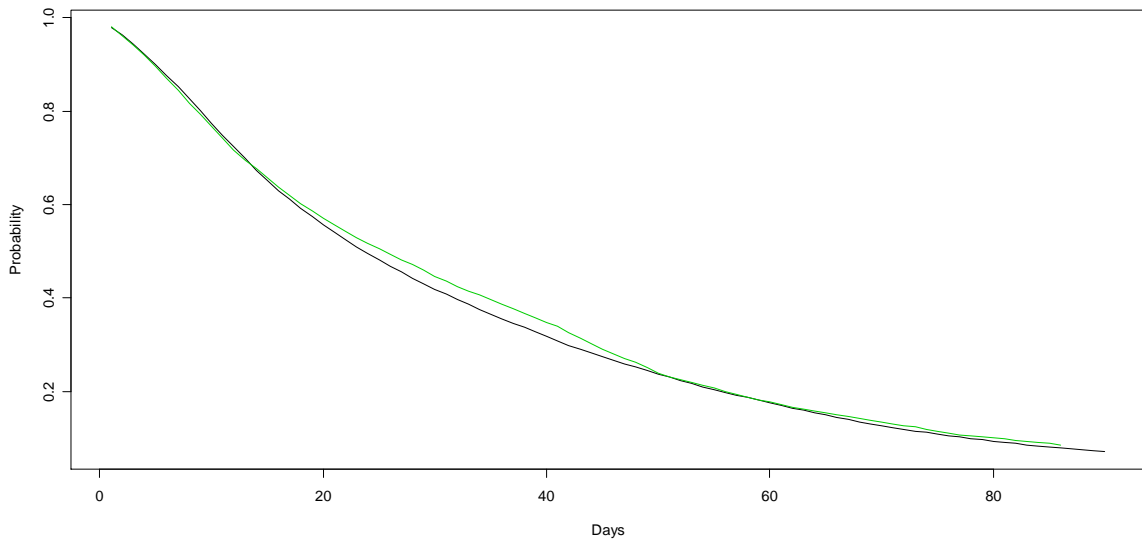
Negre: no pateix el problema Verd: pateix el problema

Gràfic 24. Probabilitat del primer ingrés segons si es pateix el problema mpoc



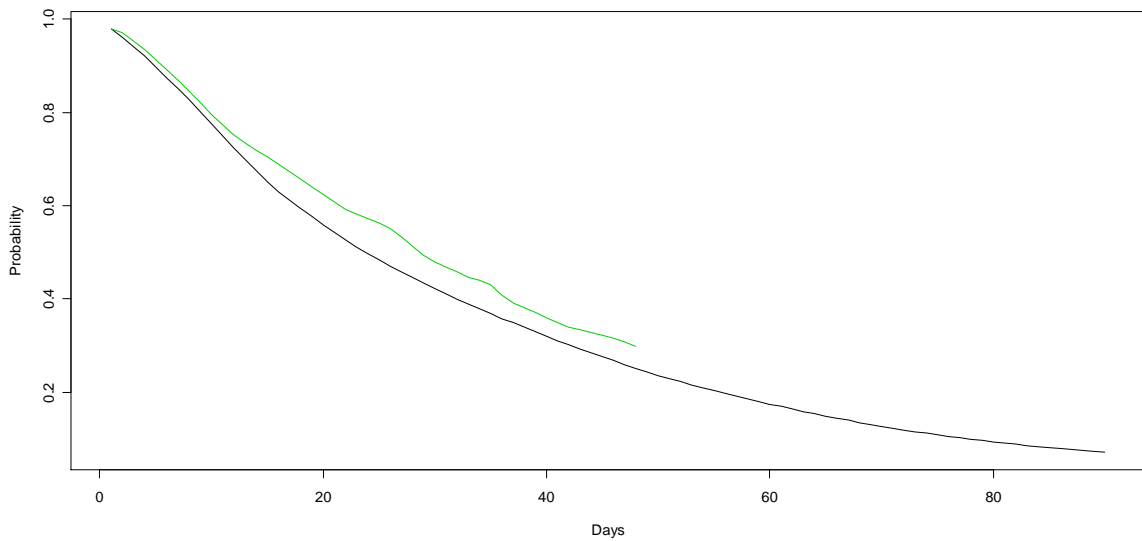
Negre: no pateix el problema Verd: pateix el problema

Gràfic 25. Probabilitat del primer ingrés segons si es pateix el problema hipertensió arterial.



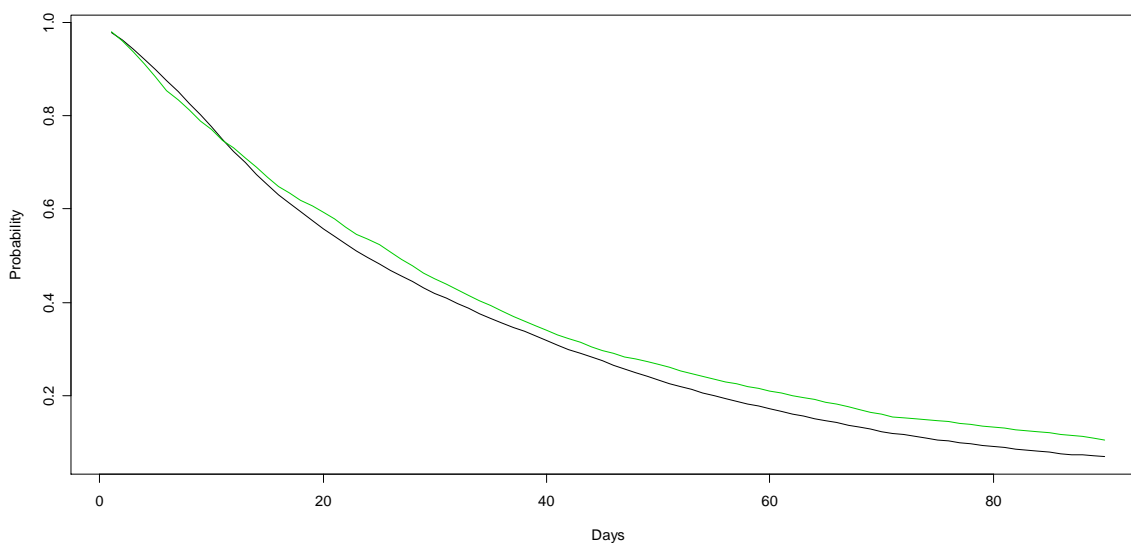
Negre: no pateix el problema Verd: pateix el problema

Gràfic 26. Probabilitat del primer ingrés segons si es pateix el problema insuficiència cardíaca.



Negre: no pateix el problema Verd: pateix el problema

Gràfic 27. Probabilitat del primer ingrés segons si es pateix el problema tabac.



Negre: no pateix el problema Verd: pateix el problema

De la mateixa manera que es va fer amb el mètode Goldstein es van introduir els efectes espacials a l'aproximació d'Andersen. La següent taula (taula 33) mostra un resum dels resultats obtinguts en les dues aplicacions:

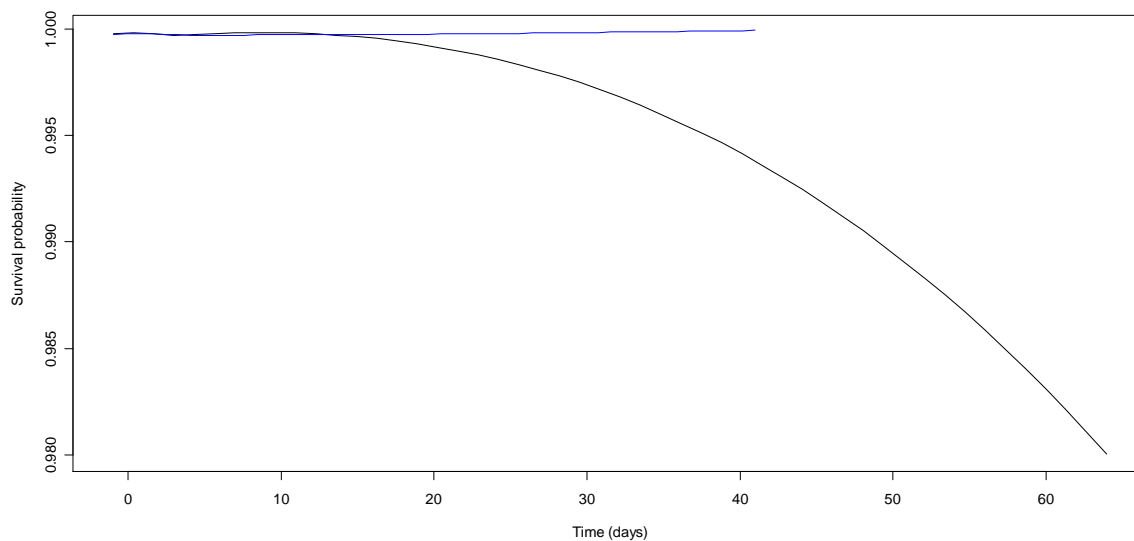
Taula 33. Comparació models amb efectes espacials (models 5.4 i 5.5).

	Mètode Golstein	Mètode Andersen
	Mitjana (desviació típica)	Mitjana (desviació típica)
sds	0.228 (0.107)	4.952 (0.051)

Per l'aproximació d'Andersen amb efectes espacials els gràfics més destacables de la probabilitat de supervivència són els que es mostren a continuació (a l'Annex VII s'adjunten tots els gràfics realitzats per aquests resultats).

En el gràfic 28 es mostra com la supervivència en els homes disminueix amb el temps mentre que per les dones es manté aproximadament entorn del 100%.

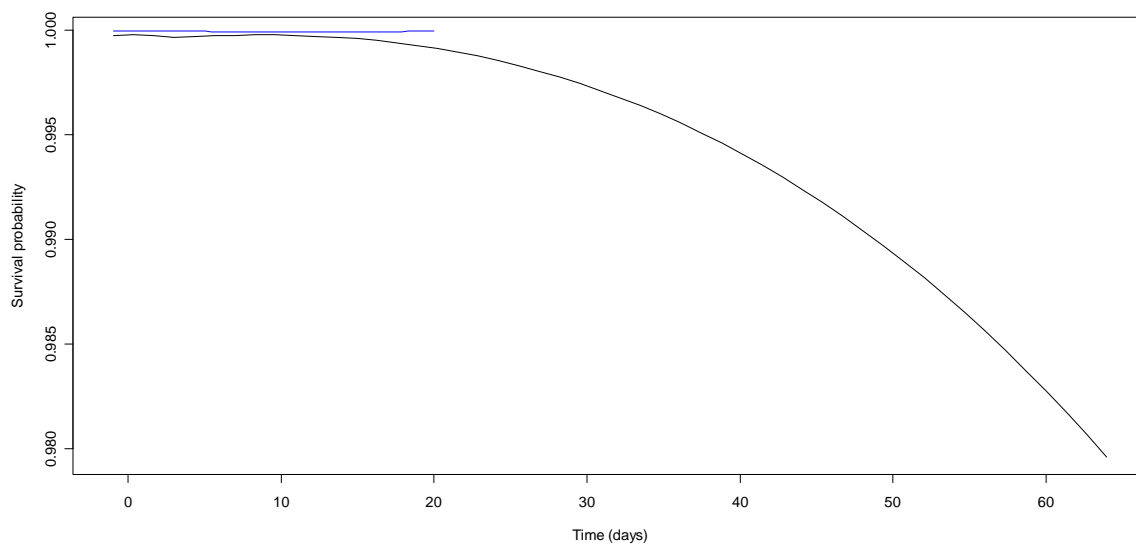
Gràfic 28. Supervivència segons el sexe.



Negre: homes
Blau: dones

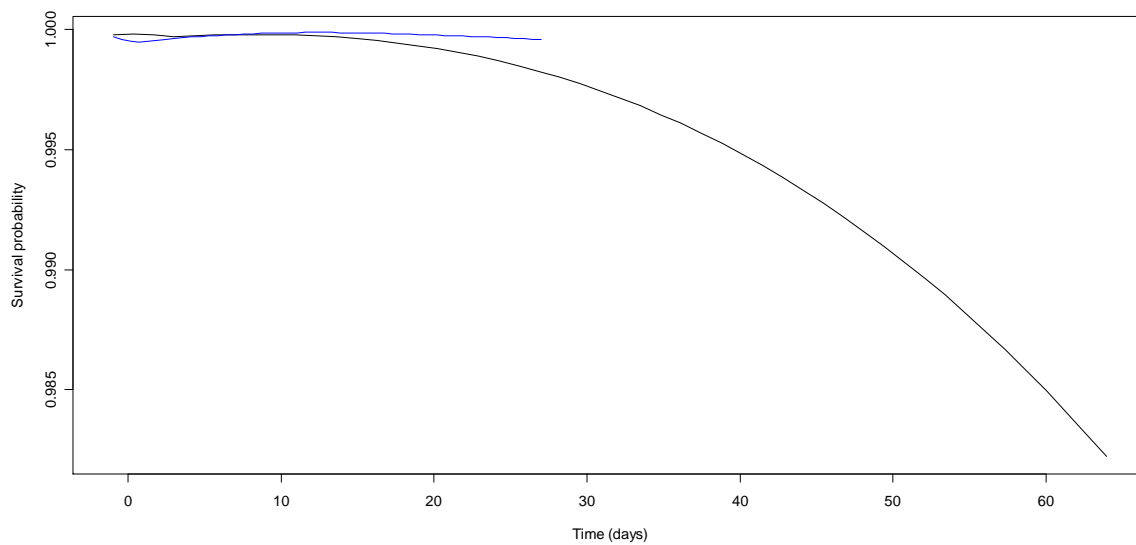
En els altres gràfics (gràfics 29-32) es mostra la supervivència segons si es pateix o no cadascun dels problemes crònics de l'arxiu de dades. En tots els casos excepte per el problema d'hipertensió arterial la supervivència en el cas de no patir el problema corresponent disminueix al llarg del temps aproximadament de la mateixa manera. En el cas de la hipertensió però, aquesta corba es mou en sentit contrari. Per altra banda, la corba corresponent a la supervivència en el cas de partir la malaltia es comporta de forma diferent per cada cas.

Gràfic 29. Supervivència segons el problema alcohol.



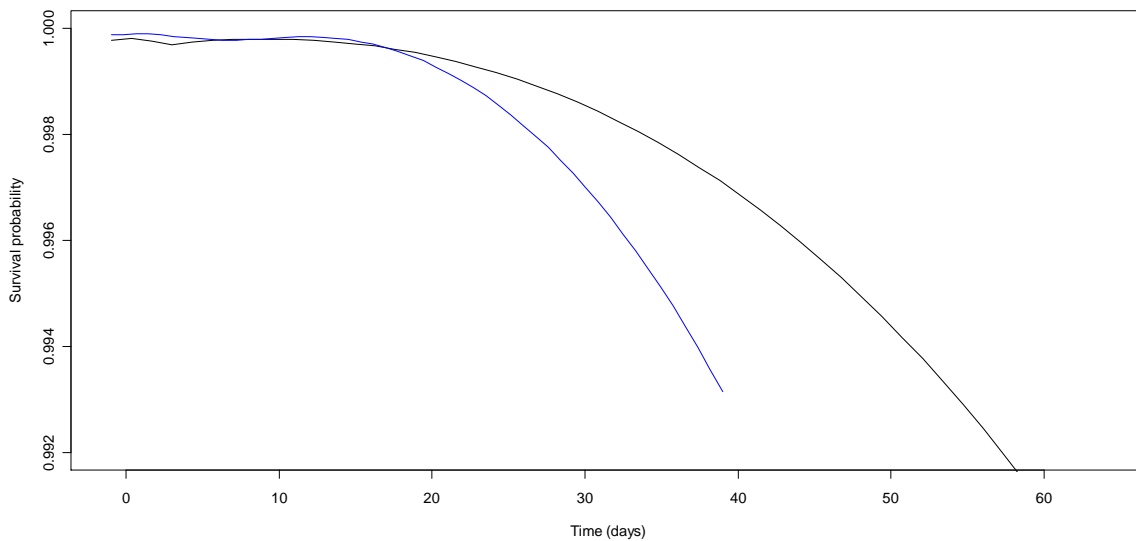
Negre: no patir el problema
Blau: patir el problema

Gràfic 30. Supervivència segons el problema mpoc.



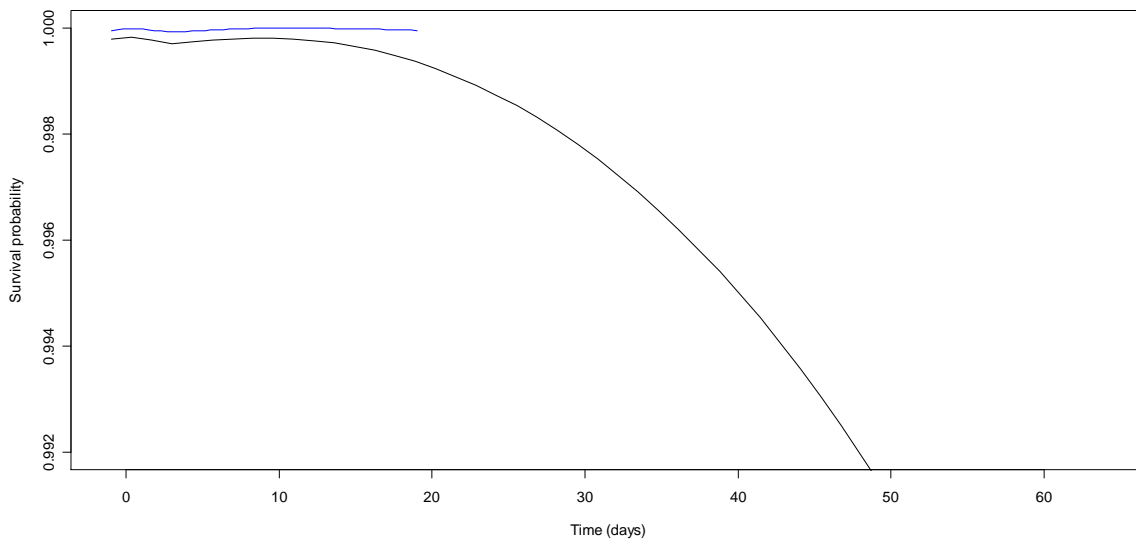
Negre: no patir el problema
Blau: patir el problema

Gràfic 31. Supervivència segons el problema tabac.



Negre: no patir el problema
Blau: patir el problema

Gràfic 32. Supervivència segons el problema extabac.



Negre: no patir el problema
Blau: patir el problema

6.4. Comparació dels resultats obtinguts pels diferents mètodes aplicats.

Globalment parlant, els resultats obtinguts en les dues aplicacions, Andersen i Goldstein, no presenten massa diferències, cosa que es pot interpretar com que les dues aproximacions són vàlides per tal de treballar amb dades de supervivència multivariant a partir de GLMM que era l'objectiu principal d'aquest treball de recerca, tot i que en els dos casos les variables significatives han estat poques degut, segurament, a que la probabilitat de reingressar a l'hospital està influenciada per altres coses diferents a les que s'han tingut en compte en aquest estudi.

En l'aplicació del mètode d'Andersen han resultat ser significatives les variables nombre de recurrències que havia experimentat l'individu durant el període d'estudi (l'any 2003) i el nombre de visites que havia generat en aquest mateix període als serveis d'atenció primària, i les dues afecten de forma negativa en la probabilitat de reingressar, és a dir, com més ingressos i com visites realitzades menys probabilitat de tornar a ingressar a l'hospital.

Quan a l'aplicació del mètode de Goldstein, només ha resultat ser significativa la variable relacionada amb el temps de permanència en cada estat de manera que com més gran és el temps de permanència més gran és la probabilitat de reingressar a l'hospital. Respecte a la variable nombre de visites realitzades a primària, en l'aplicació de Goldstein no ha resultat ser significativa però la seva influència en la probabilitat de reingressar és negativa igual com passava en l'aplicació d'Andersen, és a dir, com més visites realitzades als serveis d'atenció primària menys probabilitat de reingressar té l'individu. Aquest fet corrobora una de les hipòtesis que es plantejava en aquest treball de recerca, la qualitat dels serveis d'atenció primària, ja que com més visites realitzades més probable és detectar a temps la malaltia i per tant evitar possibles posteriors ingressos.

A més, en les dues aplicacions s'ha obtingut que les dones sempre tenen una probabilitat inferior als homes de tornar ingressar, tot i que aquesta influència no ha resultat ser significativa en cap de les aplicacions. En el mateix sentit, els resultats obtinguts en les dues aplicacions també coincideixen en el fet que com més joves sigui l'individu més petita és la probabilitat de reingrés a l'hospital, tot i que aquesta influència tampoc ha resultat ser significativa.

Per altra banda, la introducció d'efectes espacials als models no aporta una gran millora en els resultats, ja que aquests resultats continuen essent semblants als obtinguts sense tenir en compte aquests efectes.

Comparant els resultats obtinguts en les aplicacions quan s'introdueixen els efectes espacials, cal destacar la diferència entre els dos models en la magnitud de la desviació típica de l'efecte aleatori que recull la dependència espacial. De fet, aquesta desviació típica és molt més gran en el mètode d'Andersen. És molt probable que aquest fet sigui degut a que el mètode d'Andersen es realitza en dues etapes. En la primera, es construeixen els pseudovalors i posteriorment, en la segona, s'estima el model GLMM final. L'aplicació de Goldstein, en canvi, utilitza una única etapa, estimant directament el model GLMM final. Aquesta aproximació, com es sap, sempre és més eficient que les seves alternatives polietàpiques.

Referències

Barceló MA., Saez M., Cano-Serral G., Martínez-Beneito MA., Martínez JM., Borrell C., Ocaña-Riola R., Montoya I., Calvo M., López-Abente G., Rodríguez-Sanz M., Toro S., Alcalá JT., Saurina C., Sánchez-Villegas P. i Figueiras A. (2008): Métodos para la suavización de indicadores de mortalidad: aplicación al análisis de desigualdades en mortalidad en ciudades del estado Español (proyecto MEDEA). *Gaceta Sanitaria*, 22(6), 596-608.

Brooks SP. i Gelman A. (1998): Alternatives methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7, 434-455.

Carlin BP, Gelman A, Neal RM. (1998); Markov Chain Monte Carlo in Practice: A Roundtable Discussion. *The American Statistician*, 52,93-100.

Gelman A, Carlin JB, Stern HS, Rubin DB. (2004): *Bayesian data analysis*. Londres:Chapman & Hall/CRC.

Gelman A. (2005): Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1(2):1-19.

Gelman A, Sturtz S, Ligges U. The R2WinBUGS Package. [citado el 20 de Julio de 2007]. Disponible en: <http://www.stat.columbia.edu/~gelman/bugsR/>.

R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2007, ISBN 3-900051-07-0, [citado el 20 de Julio de 2007]. Disponible en: <http://www.R-project.org>.

Sánchez M. (2008): El 80% de los pacientes con EPOC desconoce sufrir la enfermedad. [http://www.elpais.com/articulo/sociedad/pacientes/EPOC/desconoce/sufrir/enfermedad/elpepusoc/20081119elpepusoc_9/Tes]

Spiegelhalter D, Thomas A, Best N, Lunn D. WinBUGS User Manual. Version 1.4. Enero 2003. <http://www.mrc-bsu.cam.ac.uk/bugs>.

ANNEXOS

ANNEX I: ANÀLISI DE LA SUPERVIVÈNCIA UNIVARIANT

A.1.1. Introducció.

Seguint Barceló (2001), l'anàlisi de la supervivència es defineix com el conjunt de tècniques que s'apliquen quan la variable d'interès és el temps que passa des de l'inici d'un estudi fins que es produeix l'esdeveniment d'interès. En aquest context la variable resposta acostuma a mesurar el temps transcorregut entre l'inici del tractament i l'execució d'un cert esdeveniment tal com la data de remissió d'un càncer, la recuperació d'un transplantament de cor, o la mort deguda al sida en epidemiologia; el temps que passa des de que se li ofereix una feina a un individu fins que l'accepta (en economia, anàlisi de duració); el temps que passa fins que un individu es casa (en demografia, *event analysis*), etc. En general, qualsevol temps d'interès s'anomenarà *temps de supervivència* encara que no representi el temps fins a la mort.

La pregunta fonamental, en aquest moment, és: què fa diferent aquest tipus de dades de totes les altres? O en altres paraules, perquè es necessita una teoria estadística diferent?

En termes generals, el que fa especials les dades de supervivència és que la resposta és el temps i el temps no es mesura de la mateixa manera que les altres variables. Qualsevol variable es pot, en general, mesurar de forma instantània, no obstant, el temps de supervivència no es pot mesurar fins que l'esdeveniment d'interès s'ha produït.

El fet que el temps es mesuri de forma seqüencial comporta diverses conseqüències:

1.- Les dades de supervivència no estan distribuïdes simètricament. En la majoria dels casos, l'histograma corresponent als temps de supervivència d'un grup d'individus similars tendeix a tenir una asimetria positiva, és a dir, les dades són asimètriques a la dreta (coeficient d'asimetria positiu).

2.- Els temps de supervivència no poden ser negatius.

3.- Els temps de supervivència poden estar censurats, és a dir, que no s'ha observat l'esdeveniment d'interès per aquest individu. Quan això és degut a que l'estudi s'acaba abans de presentar-se l'esdeveniment d'interès, o que l'individu falla per causes

diferents a les analitzades, o que s'ha perdut l'individu abans de finalitzar l'estudi, llavors es parla de censura a la dreta. Però existeix un altre tipus de censura, la censura a l'esquerra que no és tan habitual com l'anterior i es produeix quan el temps de supervivència real és menor que el temps de supervivència observat. Si es suposa, per exemple, que interessa estudiar el temps que passa fins que es reproduïx un determinat càncer després d'una primera intervenció quirúrgica, pot passar que alguns dels pacients morin abans d'aquesta intervenció. Aquests pacients estaran censurats a l'esquerra.

Totes aquestes característiques fan que les dades de supervivència no es distribueixin normalment i per això, és més apropiat adoptar altres models per aquest tipus de dades.

A.1.2. Funcions importants dins l'anàlisi de la supervivència

En l'anàlisi de la supervivència hi ha dues funcions d'especial interès la funció de supervivència i la funció de risc, que es passen a detallar a continuació.

El temps de supervivència d'un individu t es pot considerar un valor d'una variable aleatòria T , la qual no pren cap valor negatiu i que s'anomena variable aleatòria associada al temps de supervivència.

Els diferents valors que aquesta variable pot prendre segueixen una distribució de probabilitat. Si es suposa que $f(t)$ és la funció de densitat d'aquesta variable T , llavors la seva funció de distribució ve donada per (Collet, 1994):

$$F(t) = P(T < t) = \int_0^t f(u) du \quad (\text{A.1.2.1})$$

i representa la probabilitat que el temps de supervivència sigui més petit que un determinat valor t , és a dir, que l'esdeveniment d'interès s'hagi produït abans del moment t .

La *funció de supervivència*, $S(t)$, es defineix com la probabilitat que el temps de supervivència sigui més gran o igual que un determinat valor t .

$$S(t) = P(T \geq t) = 1 - F(t) \quad (\text{A.I.2.2})$$

Aquesta funció representa la probabilitat que un individu sobrevisqui, com a mínim fins al temps t .

$S(t)$ compleix les següents característiques:

$$S(t) = 1 \text{ si } t < 0$$

$$S(\infty) = \lim_{t \rightarrow \infty} S(t) = 0$$

$S(t)$ és una funció monòtona decreixent.

Per altra banda, la *funció de risc* representa la probabilitat que un individu falli en temps t condicionada a que hagi sobreviscut fins a t . Per tant, la funció de risc representa la taxa de mortalitat per un individu que ha sobreviscut fins al moment t .

D'una manera més formal, es pot definir la funció de risc com el següent límit:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq t + \Delta t / T \geq t)}{\Delta t} \quad (\text{A.I.2.3})$$

A partir d'aquesta definició i utilitzant alguns resultats estàndard de la teoria de probabilitat, es poden obtenir algunes relacions entre aquestes dues funcions:

$$\lambda(t) = \frac{f(t)}{S(t)} \qquad \lambda(t) = -\frac{d}{dt} \{\log S(t)\}$$

$$S(t) = \exp\{-\Lambda(t)\} \text{ on } \Lambda(t) = \int_0^t \lambda(u) du \text{ és la funció de risc acumulada.} \quad (\text{A.I.2.4})$$

Així, $\Lambda(t) = -\log S(t)$.

A.I.3. Estimació de la funció de supervivència i de la funció de risc.

Un pas inicial en l'anàlisi d'un conjunt de dades de supervivència és presentar resums numèrics o gràfics dels temps de supervivència pels individus que s'estan estudiant.

Això es pot fer donant una estimació de la funció de supervivència i de la funció de risc definides a l'apartat anterior.

Els mètodes que es mostren a continuació s'anomenen no paramètrics degut a que no necessiten especificar la forma de la funció de densitat de probabilitat dels temps de supervivència (T).

A.I.3.1. Estimació de la funció de supervivència.

La forma més senzilla d'estimar la funció de supervivència S(t) definida en l'equació (A.I.2.2) si s'han observat totes les dades (no hi ha dades censurades, ni dos temps de supervivència iguals) és a partir dels temps de supervivència observats mitjançant la funció de supervivència empírica (Collet,1994):

$$\tilde{S}(t) = \frac{\text{nombre d'individus amb temps de supervivència major o igual a } t}{\text{nombre total d'individus}} \quad (\text{A.I.3.1.1})$$

Equivalentment, $\tilde{S}(t) = 1 - \tilde{F}(t)$ on $\tilde{F}(t)$ és la funció de distribució empírica i dóna la probabilitat que un individu mori en temps t.

Exemple 1:

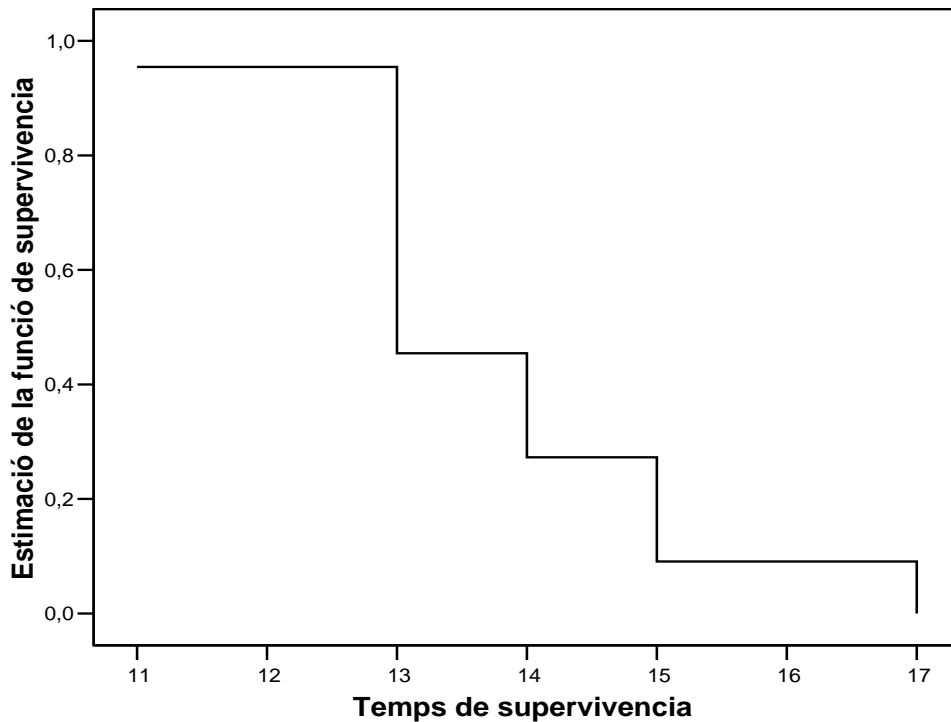
En un estudi relacionat amb el càncer de pulmó, Burdette i Gehan (1970) van seguir el temps de supervivència d'onze homes afectats per aquesta malaltia (Collet,1994). Els temps en mesos i l'estimació de la funció de supervivència empírica corresponent apareixen a la taula següent:

Taula A.I.1. Estimació de la funció de supervivència.

Temps de supervivència (mesos)	Individus a risc (temps de supervivència major o igual a t)	Individus morts	$\tilde{S}(t)$
11	11	1	11/11=1
13	10	5	10/11=0.909
14	5	2	5/11=0.455
15	3	2	3/11=0.273
17	1	1	1/11=0.091

El valor de l'estimació de la funció de supervivència és 1 fins al temps d'origen 11 mesos i 0 a partir dels 17 mesos. A continuació es mostra un gràfic d'aquesta estimació:

Gràfic A.I.1. Estimació de la funció de supervivència.

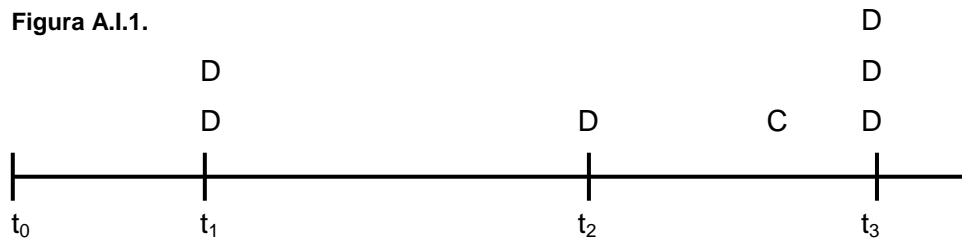


El problema d'aquest mètode d'estimació són els individus amb temps de supervivència censurat. Kaplan i Meier (1958) van suggerir una extensió d'aquest mètode que permet l'existència d'aquest tipus d'observacions i es basa en la descomposició de la supervivència d'un individu després de t anys en un producte de probabilitats condicionades, i llavors es troba cada un dels factors d'aquest producte.

Per determinar l'estimador de KM de la funció de supervivència d'una mostra de dades de supervivència censurades es construeixen una sèrie d'interval de manera que en cada interval apareix un únic cop l'esdeveniment d'interès i aquest es produeix a l'inici de l'interval.

Per exemple, si es suposa que t_1 , t_2 i t_3 representen tres temps de supervivència ordenats en ordre ascendent, és a dir, $t_1 < t_2 < t_3$ i C és un temps de supervivència censurat produït entre els moments t_2 i t_3 , els intervals que es construeixen comencen en t_1 , t_2 i t_3 , i en cada un d'ells s'ha produït l'esdeveniment d'interès una única vegada tot i que pot ser que més d'un individu l'hagi experimentat en aquest temps. La situació

proposada es mostra a la figura 1 on D representa l'esdeveniment d'estudi (en aquest cas es suposa que és la mort):



S'observa que en temps t_1 existeixen dos individus que han experimentat l'esdeveniment d'interès, en t_2 només un i en t_3 tres individus han mort; i que no existeix cap interval que comenci en la dada censurada C.

El temps inicial es denota per t_0 , i per tant, s'obté un interval inicial que comença en t_0 i acaba just abans de t_1 , temps en què s'ha produït per primera vegada l'esdeveniment d'interès. En aquest primer interval, per tant, no s'ha produït cap mort. El següent interval comença en t_1 i acaba just abans de t_2 , temps en què s'ha tornat a produir l'esdeveniment, i aquest interval, per tant, inclou les morts que s'han produït en el moment t_1 . El segon interval comença en t_2 i acaba just abans de t_3 i inclou la mort produïda en el moment t_2 i el temps de censura C. Finalment, l'últim interval comença en t_3 i conté el temps de supervivència més llarg, el que s'ha produït en el moment t_3 .

En general, es suposa que es tenen n individus amb temps de supervivència t_1, t_2, \dots, t_n . Algunes d'aquestes observacions poden estar censurades per la dreta i també pot donar-se el cas que més d'un individu tingui el mateix temps de supervivència observat.

Es suposa també que existeixen r d'aquests n temps observats on s'ha detectat l'esdeveniment que s'està estudiant, essent $r \leq n$. Siguin $t_1 < t_2 < \dots < t_r$ els temps anteriors ordenats de forma ascendent, es denota per n_j el nombre d'individus a risc en temps t_j , és a dir, els individus vius en t_j inclosos aquells que van morir en aquest moment; i d_j el nombre de morts en t_j .

Llavors, la probabilitat que un individu mori en l'interval de $t_{j-\delta}$ fins a t_j es pot estimar utilitzant la funció de supervivència empírica (A.I.3.1.1) com $\frac{n_j - d_j}{n_j}$.

A vegades pot passar que un temps de supervivència censurat es produeixi en el mateix moment que l'esdeveniment d'estudi. En aquest cas per calcular els individus a risc en aquest temps determinat (n_j) es suposa que el temps censurat s'ha produït immediatament després de l'esdeveniment.

I finalment notar que si es suposa que els individus fallen independentment uns dels altres, la probabilitat de sobreviure com a mínim j és una probabilitat condicionada a que sobrevisqui fins a j ; i per tant, aquesta probabilitat no serà res més que multiplicar la probabilitat de sobreviure en j per la probabilitat de sobreviure en tots els temps de supervivència anteriors. Així s'obté l'estimador de Kaplan-Meier per la funció de supervivència:

$$\hat{S}_{KM}(t) = \prod_{j=1}^k \left(\frac{n_j - d_j}{n_j} \right) \quad (\text{A.I.3.1.2})$$

on $k = 1, \dots, r$ i fixant $\hat{S}(t) = 1$ quan $t < t_1$ i $\hat{S}(t) = 0$ quan $t > t_r$.

Estrictament parlant, si el temps de supervivència més gran és un temps de supervivència censurat, C , llavors $\hat{S}(t)$ no està definida per $t > C$. En cas contrari, és a dir, t_r no està censurat ($n_r = d_r$), llavors $\hat{S}(t)$ és zero per tot $t > t_r$. El dibuix de l'estimador de KM de la funció de supervivència és una funció constant a trossos; les probabilitats de supervivència estimades són constants entre dos temps de fallada consecutius i decreix en cada temps de fallada.

Cal esmentar que si no existís censura, $n_j - d_j = n_{j+1}$ per $j = 1, 2, \dots, k$ i per tant:

$$\hat{S}_{KM}(t) = \frac{n_2}{n_1} \times \frac{n_3}{n_2} \times \dots \times \frac{n_{k+1}}{n_k} \quad (\text{A.I.3.1.3})$$

que es redueix a $\frac{n_{k+1}}{n_1}$, per $k = 1, 2, \dots, k$; i $\hat{S}(t) = 1$ quan $t < t_1$ i $\hat{S}(t) = 0$ quan $t \geq t_r$.

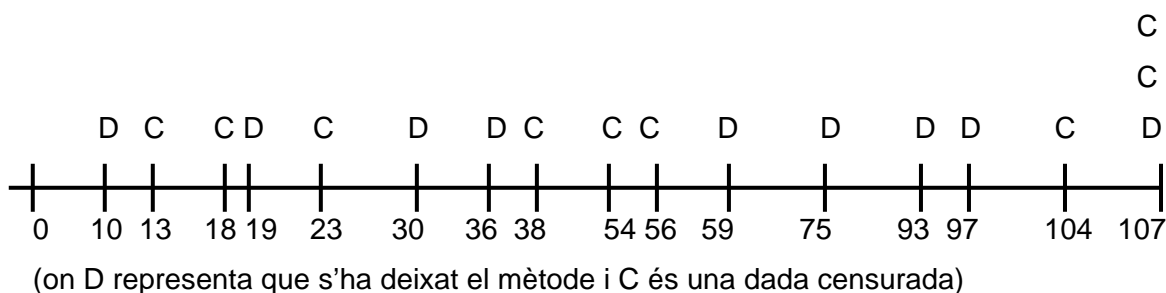
Així, doncs, es pot concloure que quan no existeixen temps de supervivència censurats, l'estimador de Kaplan-Meier és idèntic a la funció de supervivència empírica vista a l'inici d'aquest apartat.

Exemple 2: Temps fins que s'ha deixat d'utilitzar el DIU

Les dades que es donen a continuació representen el nombre de setmanes que 18 dones amb edats compreses entre 18 i 35 anys les quals han tingut dos embarassos prèviament, han estat utilitzant el DIU com a mètode anticonceptiu (Collet,1994). Les dades amb asterisc són dades censurades:

10	13*	18*	19	23*	30	36	38*	54*
56*	59	75	93	97	104*	107	107*	107*

Figura A.I.2.



Per aquestes dades, la funció de supervivència, $S(t)$, representa la probabilitat que una dona continuï utilitzant aquest mètode anticonceptiu després d'un temps t qualsevol. Utilitzant l'equació donada en (A.I.3.1.2) s'obté l'estimador de Kaplan-Meier de la funció de supervivència que es mostra en la següent taula:

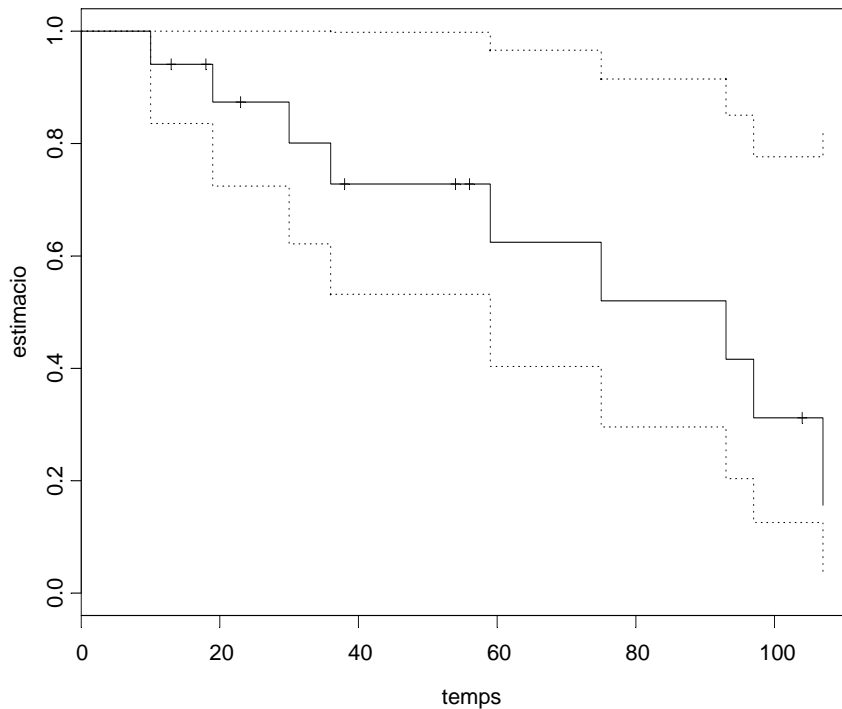
Taula A.I.2. Estimació de la funció de supervivència per KM.

Temps	n_j	d_j	$(n_j - d_j) / n_j$	$\hat{S}(t_j)$
0	18	0	1.0000	1.0000
10	18	1	0.9444	0.9444
19	15	1	0.9333	0.8815
30	13	1	0.9231	0.8137
36	12	1	0.9167	0.7459
59	8	1	0.8750	0.6526
75	7	1	0.8571	0.5594
93	6	1	0.8333	0.4662
97	5	1	0.8000	0.3729
107	3	1	0.6667	0.2486

A partir dels resultats obtinguts es pot afirmar, per exemple, que la probabilitat que una dona continuï utilitzant el DIU com a mètode anticonceptiu després de 30 setmanes és del 81.37%.

L'estimació de la funció de supervivència també es pot veure en el gràfic que segueix :

Gràfic A.I.2. Estimació de la funció de supervivència.



A.I.3.2. Estimació de la funció de risc.

Una forma natural d'estimar la funció de risc és calculant la relació entre el nombre de morts en un temps donat (d_j) i el nombre d'individus a risc en aquest temps (n_j). Si es suposa que la funció de risc és constant entre morts successives, el risc per unitat de temps es pot calcular dividint la proporció anterior per l'allargada de l'interval.

Així doncs, si en temps t_j s'han produït d_j morts, $j=1, 2, \dots, r$, i n_j representa el nombre d'individus a risc en aquest moment, la funció de risc a l'interval de t_j fins a t_{j+1} es pot estimar per:

$$\hat{\lambda}(t) = \frac{d_j}{n_j \tau_j} \tag{A.I.3.2.1}$$

per $t_j \leq t \leq t_{j+1}$, on $\tau_j = t_{j+1} - t_j$, és a dir, la longitud de l'interval.

Cal observar que no és possible estimar la funció de risc a partir de l'estimador vist en (A.1.3.2.1) quan es tracta de l'últim interval, és a dir, l'interval que comença en l'últim temps de supervivència, ja que aquest no té final.

També de forma natural s'obté un estimador de la funció de risc acumulada a partir de la relació vista en (A.1.1.2.4) prenent logaritmes i substituint la funció de supervivència per l'estimador de KM:

$$\hat{\Lambda}(t) = -\log S_{KM}(t)$$

No obstant és més utilitzat l'anomenat *estimador de Nelson-Aalen* (Nelson, 1972; Aalen, 1978), definit com:

$$\hat{\Lambda}_{NA}(t) = \sum_{j=1}^k \frac{d_j}{n_j} \quad (\text{A.1.3.2.2})$$

Aquests dos estimadors són en realitat asimptòticament equivalents, essent l'estimador de Nelson-Aalen una aproximació de primer ordre del primer estimador. A la pràctica, amb mostres de mida petita l'estimador de Nelson-Aalen sembla comportar-se millor. En qualsevol cas, les diferències entre les dues estimacions solen ser perceptibles per valors de t grans.

Exemple 3: Temps fins que s'ha deixat d'utilitzar el DIU

Es consideren les dades utilitzades en l'exemple 2 referents al nombre de setmanes que 18 dones han estat utilitzant el DIU com a mètode anticonceptiu (Collet, 1994). En l'exemple citat s'ha calculat l'estimador de Kaplan-Meier per la funció de supervivència i en la taula que segueix es mostra l'estimador de Kaplan-Meier per la funció de risc utilitzant l'equació (3.2.1):

Taula A.I.3. Estimació de la funció de risc per KM.

Temps	τ_j	n_j	d_j	$\hat{\lambda}(t)$
0	10	18	0	0.0000
10	9	18	1	0.0062
19	11	15	1	0.0061
30	6	13	1	0.0128
36	23	12	1	0.0036
59	16	8	1	0.0078
75	18	7	1	0.079
93	4	6	1	0.0417
97	10	5	1	0.0200

Els resultats d'aquesta estimació es poden interpretar com: el risc que té una dona de deixar d'utilitzar el DIU com a mètode anticonceptiu després de 30 setmanes d'utilitzar-lo és del 1.28%.

Tot i la seva senzillesa, l'estimador de Kaplan-Meier presenta l'inconvenient que no permet considerar la influència de covariables en la supervivència dels individus i en la majoria d'estudis existeixen variables explicatives o covariables.

A.I.4. El model de regressió de Cox o model de riscos proporcionals.

Davant un conjunt de dades de supervivència, més que trobar un estimador de la funció de supervivència i de la funció de risc és interessant estudiar l'efecte que poden produir les variables explicatives o covariables en les variacions de la funció de risc en el temps. És per aquest motiu que són importants els models de regressió dins l'anàlisi de supervivència. Però, a diferència dels models de regressió estàndard, aquí no es modela l'esperança dels temps de supervivència condicionada a diferents variables explicatives, sinó que es modela directament la funció de risc. No obstant això, molts dels procediments i principis utilitzats en els models de regressió estàndard es poden utilitzar per modelar dades de supervivència.

El model de regressió més utilitzat en el context dels estudis de supervivència és el model de Cox o model de riscos proporcionals (Cox, 1972; Cox i Oakes, 1984; Collet, 1994; Hougaard, 1999; Hougaard, 2000; Therneau i Grambsch, 2000). El seu nom és degut a que suposa que les funcions de risc són proporcionals, és a dir, la relació entre els riscos de dos individus es manté constant al llarg del temps. Però aquesta hipòtesis no sempre es verifica i per això és necessari utilitzar altres models que no la requereixin, tals com el model d'Andersen i Gill, els models condicionals o els models marginals que s'expliquen en el primer capítol d'aquesta tesi.

En el següent apartat doncs, es passa a explicar el model de Cox clàssic i després s'adjunta també la formulació d'aquest model des del punt de vista de processos comptadors (veure Barceló, 2001).

A.1.4.1. El model de Cox o model de riscos proporcionals

Es denota per $X_i = \{X_{ij}\}$ ($X_{1i}, X_{2i}, \dots, X_{pi}$) el vector de p covariables associat a l'individu i -èssim per tot $i=1, \dots, n$ que es suposa que tenen un efecte sobre la supervivència i que es recullen a l'inici de l'estudi; a més, com ja s'ha dit, T_i representa el temps de supervivència d'aquest mateix individu. D'aquesta manera, es pot afirmar que el que s'observa ara són els triplets $(\tilde{T}_1, D_1, X_1), \dots, (\tilde{T}_n, D_n, X_n)$ on:

$$\tilde{T}_i = \min(T_i, C_i)$$

$$D_i = I(T_i \leq C_i) = \begin{cases} 1 & \text{si } T_i \leq C_i \\ 0 & \text{si } T_i > C_i \end{cases}$$

Quan s'analitza la influència de les diferents variables en la supervivència és usual considerar un grup de referència o basal respecte al qual es realitzaran les inferències. S'estableix per conveni que el valor d'una covariable d'un pacient en el grup basal és 0.

Ara ja s'han introduït les condicions necessàries per tal de definir el model de regressió de Cox per l'individu i -èssim. Aquest model suposa que les covariables es relacionen amb el temps de supervivència mitjançant una relació multiplicativa sobre la funció de risc i en concret, estableix una relació de proporcionalitat entre la funció de risc d'un

individu amb vector de covariables X_i i la funció de risc del grup basal tal com es mostra a continuació:

$$\lambda_i(t|X_i)=\lambda_0(t)\exp(\beta'X_i)=\lambda_0(t)\exp(\beta_1X_{i1}+\beta_2X_{i2}+\dots+\beta_pX_{ip}) \quad (\text{A.I.4.1.1})$$

on λ_0 és una funció no negativa que representa el risc basal; i $\beta'=(\beta_1,\dots,\beta_p)'$ és un vector columna de paràmetres desconeguts.

Tal com s'ha dit anteriorment, el supòsit més important del model de Cox és el que s'anomena hipòtesi de riscos proporcionals. Aquesta hipòtesi suposa que el ràtio de dos individus amb vectors de covariables X_i i X_j és constant al llarg del temps:

$$\frac{\lambda_i(t)}{\lambda_j(t)} = \frac{\lambda_0(t)e^{\beta'X_i(t)}}{\lambda_0(t)e^{\beta'X_j(t)}} = e^{\beta'(X_i(t)-X_j(t))} \quad (\text{A.I.4.1.2})$$

Exemple 4:

La situació, de nou, és l'estudi de Burdette i Gehan (1970) relacionat amb el càncer de pulmó. Ara es suposa que es tenen les següents covariables edat, sexe i status definides com:

$$\text{Edat} = \begin{cases} 0 & \text{si edat} < 25 \\ 1 & \text{si } 25 \leq \text{edat} < 40 \\ 2 & \text{si } 40 \leq \text{edat} < 60 \\ 3 & \text{si edat} \geq 60 \end{cases}$$

$$\text{Sexe} = \begin{cases} 0 & \text{si l'individu és un home} \\ 1 & \text{si l'individu és una dona} \end{cases}$$

$$\text{Status} = \begin{cases} 0 & \text{si no ha experimentat cap recaiguda} \\ 1 & \text{si ha experimentat una recaiguda} \end{cases}$$

S'observa que el grup de referència correspon a un home de menys de 25 anys que no ha experimentat cap recaiguda.

Es suposa que en el moment t s'observa una dona d'entre 25 i 40 anys que ha experimentat una recaiguda, $X_1=(1,1,1)$, i també s'observa una dona del mateix grup d'edat que no ha experimentat cap recaiguda, $X_2=(1,1,0)$. Llavors, les seves funcions de risc segons el model de Cox són:

$$\lambda_1(t|X_1)=\lambda_0(t)\exp(\beta_1+\beta_2+\beta_3)$$

$$\lambda_2(t|X_2)=\lambda_0(t)\exp(\beta_1+\beta_2)$$

Si es calcula ara el ràtio d'aquests dos individus,

$$\frac{\lambda_1(t|X_1)}{\lambda_2(t|X_2)} = e^{\beta_3}$$

es pot afirmar que el risc d'una dona d'entre 25 i 40 anys que ha experimentat una recaiguda és e^{β_3} vegades el risc d'una dona del mateix grup d'edat que no ha experimentat cap recaiguda; i la hipòtesi del model de Cox suposa que aquesta relació es manté al llarg del temps.

A més d'aquesta hipòtesi, el model de Cox també suposa que:

- Les variables explicatives actuen de manera multiplicativa en la funció de risc.
- Els temps de fallada dels individus i i j , condicionats a X_i i X_j , són independents.
- Els coeficients dels regressors, en la funció de risc, són constants al llarg del temps.
- Tots els individus experimenten la fallada només una vegada durant la seva vida.
- Tots els individus estan a risc d'experimentar la fallada.

Per estimar aquest model, es té que tant el risc basal com els paràmetres són desconeguts. Cox (1972) suggereix un mètode d'estimació que elimina l'efecte del risc basal. Aquest mètode consisteix en maximitzar l'anomenada funció de versemblança parcial de Cox (1975), que depèn únicament de les dades i dels paràmetres β . Amb això, s'estimen els paràmetres de la regressió $\hat{\beta}$, i aquests s'utilitzen per construir un estimador de la funció de risc basal, $\hat{\lambda}_0(t)$.

La funció de versemblança parcial que s'utilitza es pot definir com un producte de n factors, un per a cada temps de supervivència, tal com es mostra a continuació:

$$L(\beta) = \prod_{i=1}^n P_i(\beta)^{D_i} \quad (\text{A.I.4.1.3})$$

on

$$P_i(\beta) = \frac{e^{\beta'X_i}}{\sum_{j \in R_i} e^{\beta'X_j}} = \frac{e^{\beta'X_i}}{\sum_{j=1}^n Y_j e^{\beta'X_j}}$$

essent,

$$R_i = \{j : \tilde{T}_j \geq \tilde{T}_i\} = \text{conjunt d'individus en risc en el temps } \tilde{T}_i$$

$$Y_j(t) = I(\tilde{T}_j \geq t)$$

$$D_i = \begin{cases} 1 & \text{si } T_i \leq C_i \\ 0 & \text{si } T_i > C_i \end{cases}$$

Els paràmetres de la regressió, $\hat{\beta}$, s'estimaran maximitzant la funció $L(\beta)$, és a dir, l'estimador parcial de màxima versemblança $\hat{\beta}$ és la solució a $\{U(\beta)=0\}$, on

$$U(\beta) = \frac{\partial \log L(\beta)}{\partial \beta} = \sum_{i=1}^n D_i \left\{ X_i - \frac{S^{(1)}(\beta, \tilde{T}_i)}{S^{(0)}(\beta, \tilde{T}_i)} \right\} \quad (\text{A.I.4.1.4})$$

essent

$$S^{(0)}(\beta, t) = \sum_{j=1}^n Y_j(t) e^{\beta'X_j}$$

$$S^{(1)}(\beta, t) = \sum_{j=1}^n Y_j(t) X_j e^{\beta'X_j}$$

Per a n gran, l'estadístic $U(\beta)$ és aproximadament normal amb mitjana 0 i matriu de variàncies i covariàncies $I(\hat{\beta})$ i $\hat{\beta}$ és aproximadament normal amb mitjana β i matriu de variàncies i covariàncies $I^{-1}(\hat{\beta})$ on $I(\hat{\beta})$ és la matriu d'informació de Fisher.

Exemple 5:

Es consideren els resultats preliminars corresponents a un estudi clínic per avaluar l'eficàcia de la quimioteràpia de manteniment en 23 malalts de leucèmia. Aquest estudi es va realitzar l'any 1977 per Embury *et al.* i consistia en assignar aquests malalts en dos grups: el primer grup era tractat amb quimioteràpia de manteniment i el segon grup (grup control) no rebia cap tipus de tractament. L'objectiu de l'estudi era veure si la quimioteràpia de manteniment allargava el temps fins a la recaiguda.

Es disposa de les següents variables: temps fins a la recaiguda en setmanes (time); status del malalt (0=censurat i 1=recaiguda) i el grup al qual pertany el malalt, quimioteràpia de manteniment (maintained) o sense quimioteràpia (nonmaintained).

Si apliquem un model de regressió de Cox amb la variable grup com a variable explicativa els resultats que s'obtenen són:

Taula A.I.4. Resultats de l'estimació del model de Cox.

	$\hat{\beta}$	$\text{Exp}(\hat{\beta})$	$\text{Se}(\hat{\beta})$	Lower.95	Upper.95	z	p
Grup	-0.916	0.4	0.512	0.147	1.09	-1.79	0.074

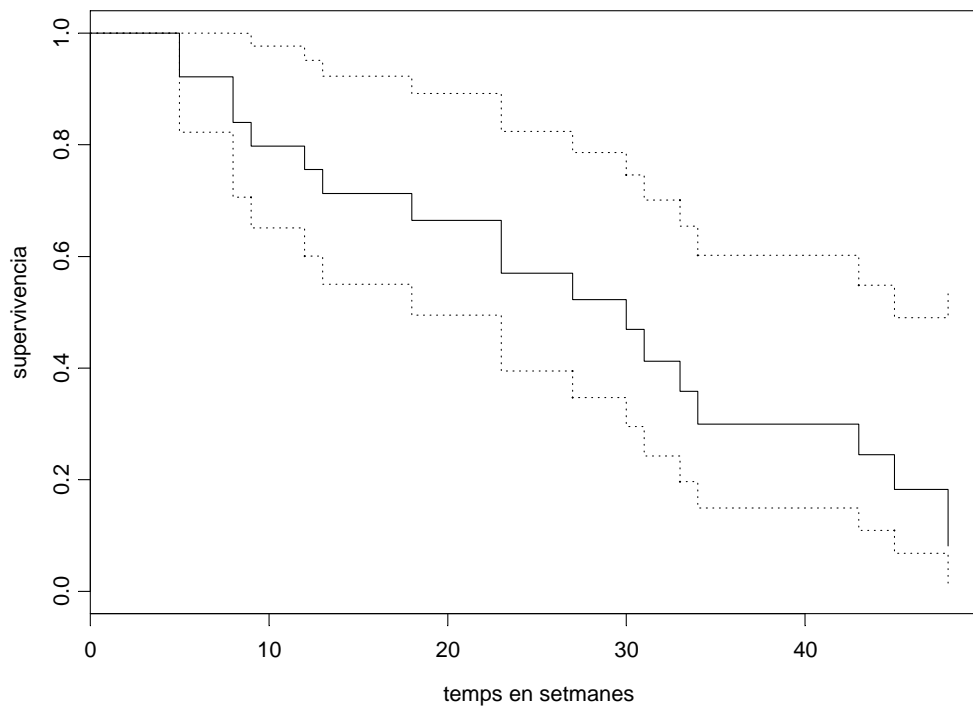
En aquesta taula es mostra l'estimació del coeficient β , $\hat{\beta}$, l'estimació del risc relatiu per una unitat de la variable $e^{\hat{\beta}}$, l'error estàndard de l'estimació, el z-test $\frac{\hat{\beta}}{\text{se}(\hat{\beta})}$, el p-valor i l'interval de confiança al 95%.

La interpretació del model de Cox no es fa directament a partir del coeficient estimat, sinó de la seva exponencial. Si la variable és dicotòmica, $\text{exp}(\hat{\beta})$ és un estimador de la raó dels riscos (hazard ratio) i s'interpreta com el risc que es té amb la presència de la covariable en relació amb la seva absència. Pel cas de variables contínues, $\text{exp}(\hat{\beta})$ representa la raó dels riscos (hazard ratio) en incrementar una unitat la covariable.

En l'exemple que s'estava estudiant, com que la variable grup és dicotòmica es pot afirmar que un pacient que pertany al grup amb quimioteràpia de manteniment té un 60% menys de risc de recaure que un pacient del grup sense quimioteràpia.

El gràfic corresponent es mostra a continuació:

Gràfic A.I.3. Estimació de la funció de risc a partir del model de Cox.



El principal problema que presenta aquest model és l'incompliment de la hipòtesi de riscos proporcionals. La causa més comú d'incompliment és la presència de covariables dependents del temps. En el cas que les covariables variïn amb el temps però es pugui predir la seva evolució futura (variables temps dependents externes com per exemple l'edat) llavors, el model de Cox es pot estendre fàcilment a partir de la següent expressió:

$$\lambda_i(t) = \lambda_0(t)e^{\sum_{j=1}^p \beta_j X_{ji}(t)} \quad (\text{A.I.4.1.5})$$

on el subíndex i fa referència a l'individu i j a la covariable. Aquesta extensió del model de Cox però, és poc utilitzada ja que en el seu lloc es sol utilitzar el model AG que s'explicarà en el segon capítol d'aquesta tesi i que inclou la mateixa.

Per altra banda, a vegades la hipòtesi de riscos proporcionals no és vàlida per alguna covariable, llavors una possibilitat és la d'estratificar segons aquesta covariable i utilitzar el model de riscos proporcionals per les altres. Aquesta extensió doncs, divideix els individus en diferents estrats on, en cada un d'ells es compleix la hipòtesis

de riscos proporcionals. A més, cada estrat té un risc basal diferent i mateixos valors pel vector de coeficients β .

Aquesta alternativa del model de Cox és l'anomenat *model de Cox estratificat* i es pot desenvolupar de la següent manera: si es suposa que la hipòtesi de riscos proporcionals no es compleix per la covariable X_1 i que X_1 és un factor amb k categories, llavors, el model de regressió de Cox estratificat per un individu i és:

$$\lambda_h(t) = \lambda_{0h}(t) \exp\{\beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_p X_{pi}\} \text{ per } h=1, \dots, k \quad (\text{A.I.4.1.6})$$

on λ_h (i λ_{0h}) fa referència a la funció de risc (i de risc basal) per l'individu i a l'estrat h .

L'estimació dels coeficients β_2, \dots, β_p es basa també en la maximització de la funció de versemblança parcial, construïda a partir de les versemblances parcials per cada estrat. El principal inconvenient d'aquesta alternativa és que no permet estimar la variable per la qual estratifiquem, i per tant, no podem estimar el seu efecte.

Altres motius que poden portar a la decisió de no utilitzar aquest mètode són per exemple: si la variable és continua la seva estratificació és una tasca complicada o si és politòmica les interaccions augmenten el número de paràmetres, la qual cosa implica una reducció dels graus de llibertat i, per tant, una reducció de l'eficàcia

A.I.4.2. Anàlisi residual del model de Cox

Després d'aplicar un model a un conjunt de dades observades, cal veure si aquest model ha estat l'adequat. Per això, és necessari l'ús de procediments de diagnosi que ens permetin detectar si s'ha produït algun error d'especificació (omissió de variables rellevants, inclusió de variables supèrflues o forma funcional incorrecta) o la presència d'individus amb alguna característica problemàtica.

Igual que en els models de regressió lineal, també en l'anàlisi de supervivència la majoria dels procediments de verificació del model es basen en unes quantitats anomenades residus. S'anomena residu a la diferència entre el valor observat i el valor estimat per l'equació de regressió, és a dir, la part que l'equació de regressió deixa sense explicar per cada individu. Alguns dels procediments per determinar si el model està ben construït es basen en representar gràficament aquests residus i avaluar si

presenten patrons anòmals comparant amb la forma que teòricament haurien de presentar.

En el model de Cox, existeixen quatre tipus de residus que resulten d'interès: (i) els residus de Cox-Snell; (ii) els residus basats en martingales, utilitzats per verificar la forma funcional de les covariables; (iii) els residus basats en la deviance utilitzats per detectar valors atípics i (iv) els residus de Schoenfeld o Score utilitzats sobretot per verificar la hipòtesis de riscos proporcionals (Collet, 1994; Therneau i Grambsch, 2000). A continuació es presenta cada un dels residus esmentats.

Es suposa que s'ha ajustat un model de Cox tal com:

$$\lambda_i(t|X_i) = \lambda_0(t) \exp(\beta' X_i) = \lambda_0(t) \exp(\beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi})$$

a) Residus de Cox-Snell.

Aquests residus són els més utilitzats en l'anàlisi de supervivència i s'anomenen així perquè són un exemple particular de la definició més general de residus donada per Cox i Snell (1981).

El residu de Cox-Snell per l'individu i -èssim, $i=1,2,\dots,n$ s'expressa com:

$$r_{ci} = \exp(\hat{\beta}' x_i) \hat{H}_0(t_i) \tag{A.I.4.2.1}$$

on $\hat{H}_0(t_i)$ és l'estimador de la funció acumulada del risc basal en el temps t_i .

Es pot demostrar, també, que aquests residus es poden expressar com $\hat{H}_i(t_i) = -\log \hat{S}_i(t_i)$, on $\hat{H}_i(t_i)$ és l'estimació de la funció de risc acumulada per l'individu i en el temps t_i ; i $\hat{S}_i(t_i)$ és l'estimació de la funció de supervivència per l'individu i en el temps t_i .

Aquests residus tenen les propietats següents: no poden prendre valors negatius, no es distribueixen simètricament entorn del 0 i a més, segueixen una distribució exponencial amb mitjana i variància 1.

b) Residus de Martingala

Aquests residus es poden definir com la diferència entre el nombre d'esdeveniments observats per a un individu i a l'interval $(0, t_i)$ i el nombre d'esdeveniments esperats segons el model.

$$r_{Mi} = \delta_i - r_{Ci} \quad (\text{A.I.4.2.8})$$

on δ_i indica les morts observades per l'individu i en temps t_i , és a dir, $\delta_i=1$ si l'individu i pot morir en t_i i $\delta_i=0$ en cas contrari. El terme r_{Ci} , són els residus de Cox-Snell i estimen la funció integrada de risc per l'individu i en t_i .

El seu nom és degut a que també es poden definir utilitzant els processos comptadors, en concret les martingales (Therneau i Grambsch, 2000).

Els residus de martingala prenen valors entre $-\infty$ i 1 per les observacions no censurades i prenen valors negatius, entre $-\infty$ i 0 per les observacions censurades.

Per n gran, es pot demostrar que els residus de martingala no estan correlacionats entre si i tenen un valor esperat igual a zero. No obstant això, no es distribueixen simètricament entorn del zero encara que el model sigui correcte, i això complica la interpretació dels gràfics.

Aquests residus s'utilitzen per estudiar la forma funcional de les covariables. Es representen en un gràfic de dispersió els residus de martingala del model sense la variable d'interès a l'eix d'ordenades i la variable d'interès al d'abscisses. Si el gràfic que s'obté és lineal, significa que la variable està ben introduïda.

Exemple 6:

S'utilitza una base de dades de càncer de pulmó fet pel North Central Cancer Treatment Group. La base de dades inclou el temps de supervivència usual (time) i una variable indicador de censura o mort (status) a més de les variables següents: un codi numèric de la institució a on va ser ingressat el pacient (inst); l'edat del pacient (age), el sexe (sex: 1=home, 2=dona), la puntuació mèdica de l'estat del ECOG(0-4); la puntuació de Karnofsky (ph.karno), la valoració del pacient del seu/va puntuació de

karnofsky ; les calories consumides en els menjars sense incloure begudes i snacks (meal.acl) i el pes perdut durant els últims 6 mesos (wt-loss).

Es comença amb l'exemple ajustant un model amb totes les variables estratificat pel sexe i els resultats són els que es mostren a continuació:

Taula A.I.5. Estimació del model amb totes les variables.

	$\hat{\beta}$	$\text{Exp}(\hat{\beta})$	$\text{se}(\hat{\beta})$	z	p
age	9.05e-003	1.009	0.011601	0.78	0.4400
ph.ecog	7.07e-001	2.029	0.222773	3.17	0.0015
ph.karno	2.07e-002	1.021	0.011282	1.84	0.0660
pat.karno	-1.33e-002	0.987	0.008050	-1.65	0.0980
meal.cal	-5.27e-006	1.000	0.000263	-0.02	0.9800
wt.loss	-1.52e-002	0.985	0.007890	-1.93	0.0540

Els resultats d'aquesta primera estimació indiquen que les variables edat i nombre de calories consumides en els menjars no són estadísticament significatives, és a dir, que no influeixen en el risc que té el pacient de morir.

Per tant, es torna a estimar el model sense aquestes dues variables.

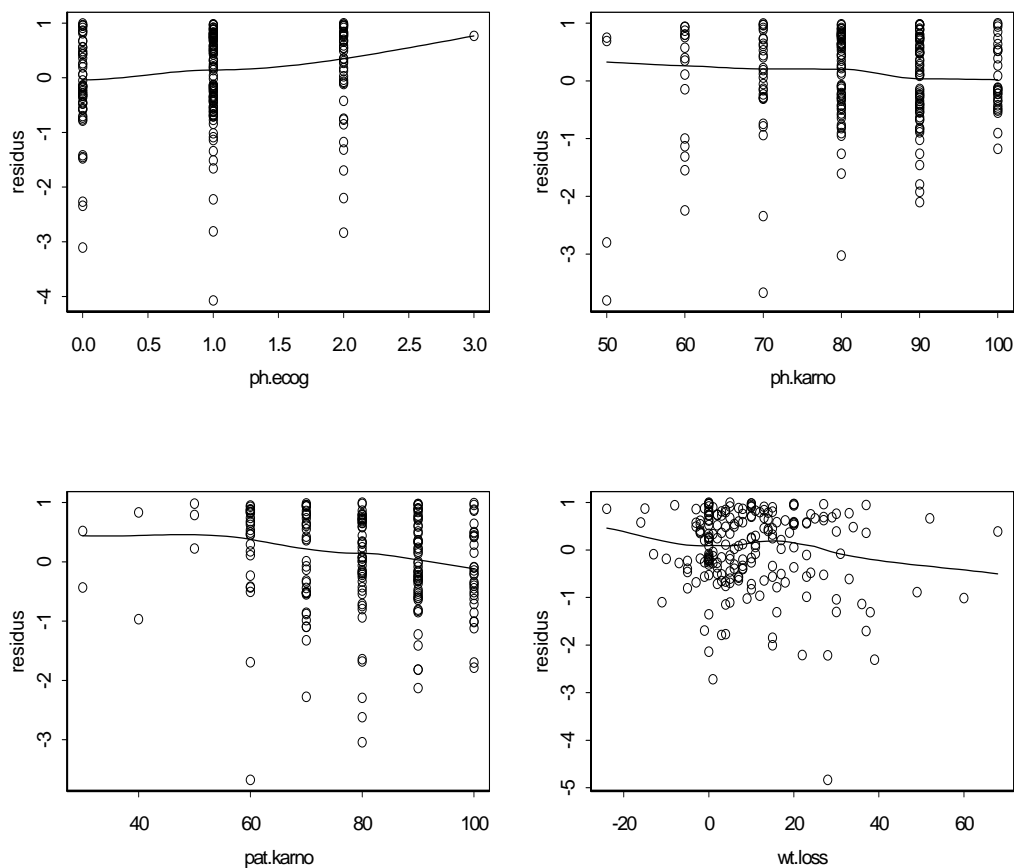
Taula A.I.6. Estimació del model sense les variables no significatives.

	$\hat{\beta}$	$\text{Exp}(\hat{\beta})$	$\text{se}(\hat{\beta})$	z	p
ph.ecog	0.6495	1.915	0.20070	3.24	0.0012
ph.karno	0.0173	1.017	0.01031	1.68	0.0930
pat.karno	-0.0167	0.983	0.00726	-2.30	0.0220
wt.loss	-0.0137	0.986	0.00691	-1.99	0.0470

D'aquesta manera totes les variables són significatives (nivell de significació 10%).

A continuació es passa a estudiar la forma funcional de la funció respecte cada una de les variables importants dels model. Això es farà a partir de gràfics dels residus de martingala del model amb la variable d'interès a l'eix d'abscisses. Però abans, per tal que la manipulació dels missings sigui una mica més simple, es crea una nova base de dades que conté només les variables del model i sense els missings. Els gràfics obtinguts són els que es mostren a continuació:

Gràfic A.I.4. Residus de martingala.



Com es pot veure, en cada gràfic s'ha representat una de les variables d'interès contra els residus del model obtingut sense aquesta variable. Per tal de considerar una bona forma funcional de la variable d'estudi cal que resulti un gràfic més o menys lineal.

c) Residus "Deviance"

Encara que els residus de martingala tenen propietats molt semblants als de la regressió lineal, no es distribueixen simètricament entorn del zero. Aquesta asimetria pot donar lloc a gràfics de residus difícils d'interpretar. En canvi, els residus deviance, que van ser introduïts per Therneau *et al.* (1990), a més de presentar propietats semblants als de la regressió lineal, es distribueixen simètricament entorn del zero.

Aquests residus es defineixen de la següent forma:

$$r_{Di} = \text{sgn}(r_{Mi}) \left[-2 \{ r_{Mi} + \delta_i \log(\delta_i - r_{Mi}) \} \right]^{1/2} \tag{A.I.4.2.9}$$

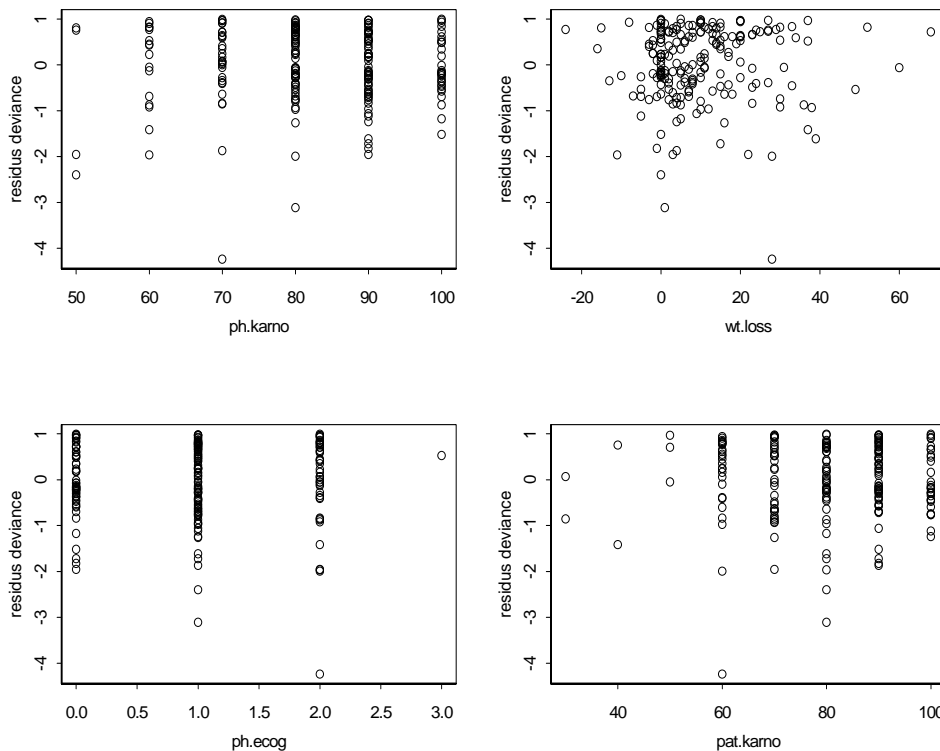
on r_{Mi} són els residus de martingala per l'individu i , i la funció $\text{sgn}()$ és la funció signe que pren el valor $+1$ si el seu argument és positiu i -1 si és negatiu. D'aquesta manera, els residus deviance tenen el mateix signe que els residus de martingala.

Aquests residus s'utilitzen per identificar valors atípics mitjançant gràfics de dispersió entre els residus i cada una de les covariables.

Exemple 7:

Es continua amb la mateixa base de dades de càncer de pulmó (exemple 6). Com que els individus amb residus deviance elevats no són ben predits pel model; aquests residus ens donaran valors atípics representant-los contra cada una de les variables d'interès. A continuació es mostra algun exemple:

Gràfic A.I.5. Residus Deviance.



d) Residus de Schoenfeld o Score

Per a l'individu i i per la variable X_j , aquests residus es defineixen com:

$$r_{S_{ji}} = \delta_i (x_{ji} - a_{ji}) \tag{A.I.4.2.10}$$

on:

$$a_{ji} = \frac{\sum_{k \in R(t_i)} x_{jk} \exp(\hat{\beta}' x_k)}{\sum_{k \in R(t_i)} \exp(\hat{\beta}' x_k)}$$

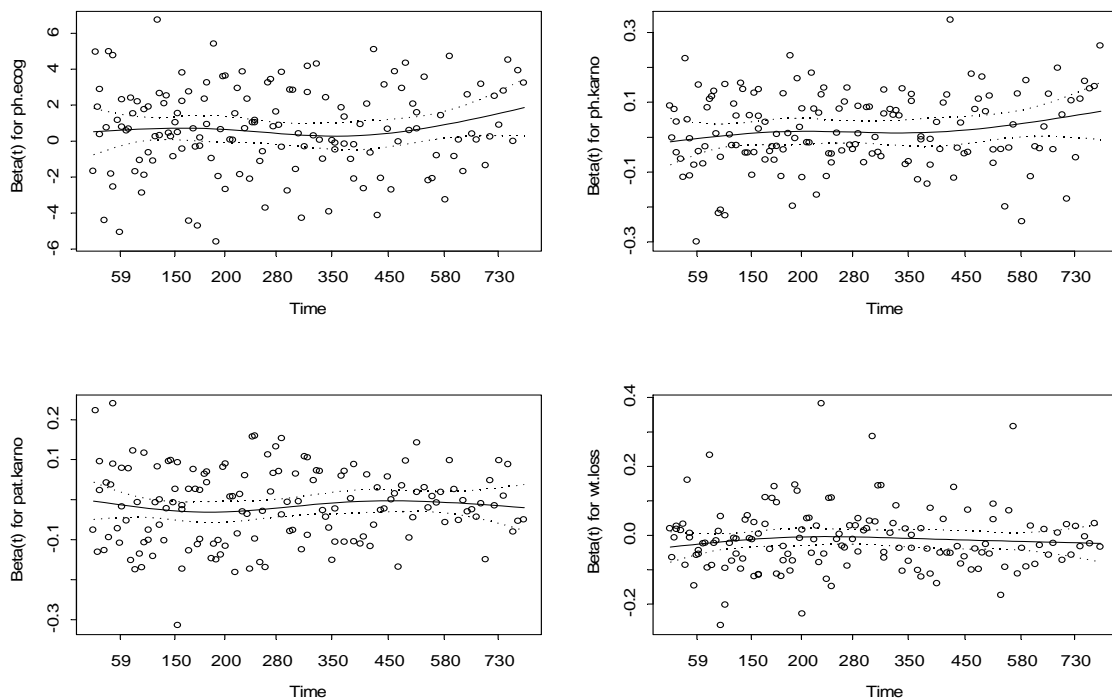
Aquests residus prenen p valors (un per cada covariable) per cada un dels individus. La seva suma i el seu valor esperat, s'aproximen a zero i, en un model ben especificat, no estan correlacionats.

S'utilitzen per contrastar la hipòtesis de riscos proporcionals representant, en un gràfic de dispersió, els residus a l'eix d'ordenades i els temps de supervivència al d'abscisses. Si es compleix la hipòtesis, el gràfic ha d'oscil·lar entorn del zero i no s'ha d'observar cap comportament sistemàtic.

Exemple 8:

Es continua amb la mateixa base de dades de càncer de pulmó (exemple 6). Finalment, es pot examinar la hipòtesis de riscos proporcionals gràficament tal com es mostra a continuació:

Gràfic A.I.6. Residus de Schoenfeld o Score.



En els quatre gràfics que es mostren s'observa com els valors oscil·len entorn del zero sense cap comportament sistemàtic i per tant, no s'observa una violació del supòsit.

A.1.4.3. Formulació del model de Cox a partir de processos comptadors.

El model de Cox de riscos proporcionals (Cox, 1972) i els mètodes estadístics associats basats en la versemblança parcial (Cox, 1975) es van situar en el context dels models d'intensitat multiplicatius per processos comptadors multivariants per tal de solucionar els problemes que aquest model presentava (Aalen, 1978; Andersen i Gill, 1982; Andersen i Borgan, 1985).

Aquesta nova formulació ha permès, a més d'ampliar el rang d'aplicació del model de Cox, una construcció molt més rigurosa i una justificació per la funció de versemblança parcial (Jacod, 1975; Arjas i Haara, 1984).

En aquest apartat es farà una descripció de les principals definicions de la teoria dels processos comptadors, i a continuació s'introduirà el model de Cox en aquest context. Algunes referències importants a la teoria dels processos comptadors i martingales són Brémaud i Jacod (1977), Jacod (1979), Gill (1980), Brémaud (1981) i Therneau *et al.*, (1990).

A.1.4.3.1. Introducció a la teoria dels processos comptadors.

Pràcticament és impossible realitzar un estudi de supervivència sense parlar de censura i del supòsit de censura independent que vol dir que la supervivència futura no es pot veure afectada per la supervivència o censura passada. Per formalitzar aquest concepte, cal parlar matemàticament sobre el passat i el futur i això serà possible a partir del concepte de *filtració*

Si es considera un espai de probabilitat (Ω, \mathcal{F}, P) , una *filtració* $\{\mathcal{F}_t; t \geq 0\}$, també anomenada "història"; és una família creixent de sub- σ -àlgebres de \mathcal{F} , que compleix les característiques següents:

$$1.- \mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F} \quad \forall s < t \quad (\text{creixent})$$

$$2.- \mathcal{F}_s = \bigcap_{t>s} \mathcal{F}_t \quad \forall s \quad (\text{continua per la dreta})$$

La σ -àlgebra \mathcal{F}_t conté tot el que passa en temps t i representa l'adquisició progressiva d'informació respecte el temps. Així, la σ -àlgebra \mathcal{F}_t tot el que passa estrictament abans del moment t ; és la σ -àlgebra més petita que conté totes \mathcal{F}_s amb $s < t$.

Llavors, el supòsit de censura independent es pot escriure com:

$$P(\tilde{T}_i \in [t, t + dt), D_i = 1 | \mathcal{F}_{t-}) = \begin{cases} \lambda(t) & \text{si } \tilde{T}_i \geq t \\ 0 & \text{si } \tilde{T}_i < t \end{cases}$$

Si es substitueix la probabilitat per l'esperança d'una variable aleatòria indicador i es suma per i s'obté la següent expressió:

$$\begin{aligned} E(\#\{i: \tilde{T}_i \in [t, t+dt), D_i=1\} | \mathcal{F}_{t-}) &= \\ = \#\{i: \tilde{T}_i \geq t\} \times \lambda(t) dt &= \\ = Y(t) \lambda(t) dt &= \\ = \alpha(t) dt & \end{aligned} \tag{A.I.4.3.1.1}$$

on es defineixen els processos Y i α com:

$$Y(t) = \#\{i: \tilde{T}_i \geq t\} \tag{A.I.4.3.1.2}$$

que compta el nombre d'individus a risc just abans del temps t , i

$$\alpha(t) = Y(t) \lambda(t). \tag{A.I.4.3.1.3}$$

Per altra banda, la fórmula (4.3.1.1) es pot interpretar com una martingala associada a un cert procés comptador; en aquest cas, al procés $N=(N(t)) \quad t \geq 0$ que compta el nombre de recurrències de l'esdeveniment que s'està estudiant en el moment t :

$$N(t) = \#\{i: \tilde{T}_i \leq t, D_i=1\}$$

i el seu procés d'intensitat associat serà α .

Per tal d'explicar el significat de martingala, primer cal definir el concepte de procés d'intensitat acumulat Λ :

$$\Lambda(t) = \int_0^t \alpha(s) ds, \quad t \geq 0$$

llavors, es pot definir el procés martingala M com:

$$M(t) = N(t) - \Lambda(t)$$

o, equivalentment,

$$dN(t) = d\Lambda(t) + dM(t) = \alpha(t)dt + dM(t) = Y(t)\lambda(t)dt + dM(t)$$

A partir d'aquí, si es considera l'esperança condicionada a \mathcal{F}_t del procés M en l'interval $[t, t+dt)$ es té:

$$\begin{aligned} E(dM(t) | \mathcal{F}_t) &= E(dN(t) - d\Lambda(t) | \mathcal{F}_t) \\ &= E(dN(t) - \alpha(t)dt | \mathcal{F}_t) \\ &= E(dN(t) | \mathcal{F}_t) - \alpha(t)dt \\ &= 0. \end{aligned}$$

Així, una martingala és un procés comptador que es caracteritza per la relació $E(dM(t) | \mathcal{F}_t) = 0$. O equivalentment, una martingala $M(t)$ és un procés estocàstic adaptat a una filtració $\{\mathcal{F}_t: t \geq 0\}$ tal que:

$$E\{M(t+s) | \mathcal{F}_t\} = M(t)$$

és a dir, si es coneix la història del procés fins al moment t , el valor de M després de s unitats és, en terme mig, el valor que tenia en t .

En general, qualsevol procés comptador N es defineix com un procés que pren els valors $0,1,2,\dots$ i registra amb un salt del valor $k-1$ al k el temps de la k -èssima recurrència d'un cert esdeveniment, i té un procés d'intensitat associat $\alpha(t)$ definit com $\alpha(t) = E(dN(t) | \mathcal{F}_{t-})$. Aquest procés d'intensitat es caracteritza pel fet que $M=N-\Lambda$, on Λ és el corresponent procés d'intensitat acumulada, és una martingala.

A.I.4.3.2. El model de Cox estàndard a partir de processos comptadors.

En el transcurs dels últims anys, s'ha anat creant una connexió entre el model de Cox, l'estudi dels processos comptadors i la teoria de martingales. A continuació es passa a introduir el model de Cox estàndard en el context dels processos comptadors.

Seguint la notació de l'apartat A.I.1.4. d'aquest annex; es denota per $X_i = (X_{1i}, X_{2i}, \dots, X_{pi})$ al vector de covariables associat a l'individu i per tot $i=1,\dots,n$; i es recorda que T_i representa el temps de supervivència d'aquest individu. D'aquesta manera, es pot afirmar que el que s'observa ara són els triplets $(\tilde{T}_1, D_1, X_1), \dots, (\tilde{T}_n, D_n, X_n)$ on:

$$\begin{aligned} \tilde{T}_i &= \min(T_i, C_i) \\ D_i &= I(T_i \leq C_i) = \begin{cases} 1 & \text{si } T_i \leq C_i \\ 0 & \text{si } T_i > C_i \end{cases} \end{aligned}$$

Llavors, el model de regressió de Cox defineix la funció de risc per l'individu i com:

$$\lambda_i(t) = \lambda_0(t) e^{\beta' X_i(t)}$$

on λ_0 és una funció no negativa anomenada risc basal i representa la funció de risc associada comú a tots els individus; i $\beta=(\beta_1,\dots,\beta_p)'$ és un vector columna de paràmetres desconeguts.

En el context dels processos comptadors, el parell de variables (\tilde{T}_i, D_i) es substitueixen per el parell de funcions $(N_i(t), Y_i(t))$, on

$N_i(t)$ = nombre d'esdeveniments o recurrències d'un mateix esdeveniment en l'interval $[0,t]$ per l'individu i -èssim.

$$Y_i(t) = \begin{cases} 1 & \text{si l'individu } i\text{-èssim està en risc en el moment } t. \\ 0 & \text{altrament.} \end{cases}$$

on la funció $N_i(t)$ és un procés comptador que compta 1 si l'individu i falla en el moment t . Així, N_i són funcions esglaonades, que prenen el valor zero en temps zero, i augmenten una unitat per cada fallada. Per altra banda, Y_i és un procés indicador que pren valors $\{0,1\}$; prenent el valor 1 quan l'individu i -èssim està sota observació (per tant, en particular, N_i només augmenta quan $Y_i=1$).

Tal com s'ha dit anteriorment, el procés comptador $N_i(t)$ està associat a un procés d'intensitat $\lambda=(\lambda_1, \dots, \lambda_n)$ de manera que el model de Cox es pot escriure ara com:

$$\lambda_i = Y_i(t)\lambda_0(t)\exp\{X_i(t)\beta\} \quad (\text{A.I.4.3.2.1})$$

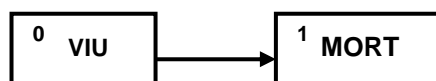
on β és un vector columna fix de p coeficients, λ_0 és la funció de risc basal, i $X_i=(X_{i1}, \dots, X_{ip})'$ és un vector columna de p covariables per l'individu i -èssim.

Sota aquest supòsit, es defineix el procés de martingala associat com:

$$M_i = N_i - \int_0^\tau \lambda_i(u)du, \quad i=1,\dots,n, \quad \tau \in [0, t]$$

Així es modelitzen els temps de supervivència mitjançant processos estocàstics amb espai de temps finit.

Resulta adequat il·lustrar aquest procés mitjançant un diagrama, on els estats es representen com a caixes, i on les fletxes entre les caixes indiquen les possibles transicions directes. El model més simple possible per dades de supervivència és el model de dos estats, 0 i 1 .



O sigui, l'individu i -èssim pot prendre dos possibles estats, estat 0 ("viu") o estat 1 ("mort"); $N_i(t)$ compta el nombre de transicions de l'estat 0 a l'estat 1 per l'individu i -èssim a l'interval de temps $[0,t]$, i $Y_i(t)=1$ si l'individu i s'observa en l'estat 0 just abans de l'instant t .

A partir d'aquí s'estimen els paràmetres del model solucionant la següent equació:

$$\frac{\partial l(\beta)}{\partial \beta} = 0$$

on la funció de versemblança parcial es defineix com:

$$l(\beta) = \sum_{i=1}^n \int_0^{\infty} \left[Y_i(t) \exp(X_i(t)\beta) - \log \left\{ \sum_j Y_j(t) \exp(X_j(t)\beta) \right\} \right] dN_i(t)$$

Finalment, es pot demostrar que els estimadors $\hat{\beta}$ són asimptòticament normals i consistents per β (Andersen et al., 1993).

Referències

Aalen O. (1978): Non-parametric inference for a family of counting processes. *The Annals of Statistics*, 6 (4), 701-726.

Arjas E. i Haara PA. (1984): A marked point process approach to censored failure data with complicated covariates. *Scandinavian journal of statistics*, 11, 193-209.

Andersen PK. i Gill RD. (1982): Cox's Regression Model for Counting Processes: A Large Sample Study. *Annals of Statistics*, 10(4), 1100-1120.

Andersen PK. i Borgan O. (1985): Counting process models for life history data (with discussion). *Scandinavian journal of statistics*, 12: 97-158.

Barceló MA. (2001): *Modelos de fragilidad en el análisis multivariante de supervivencia* [tesis doctoral]. Girona: Universitat de Girona.

Bremaud P. i Jacod J. (1977): Processus ponctuels et martingales: résultats récents sur la modélisation et filtrage. *Advances in Applied Probability*, 9(2), 362-416.

Bremaud P. (1981). *Point Processes and Queues, Martingale Dynamics*. Springer-Verlag, New York.

Burdette WJ., Gehan EA. (1970): *Planning and Analysis of Clinical Studies*. Charles C. Thomas, Springfield, Illinois.

Collet D. (1994): *Modelling Survival data in Medical Research*. Texts in Statistical Science. Chapman & Hall, London.

Cox DR. (1972): Regression models and life tables. *Journal of the Royal Statistical Society, Series B (Methodological)*, 34 (2), 187-220.

Cox DR. (1975): Partial likelihood. *Biometrika*, 62 (2), 269-276.

Cox DR., Oakes D. (1984): *Analysis of Survival Data*. Chapman & Hall, London.

Cox DR., Snell EJ. (1981) *Applied Statistics: Principles and examples*. Chapman and Hall, London.

Gill RD. (1980): *Censoring and stochastic integrals*. Math. Centre Tracts 124, Mathematical Centre, Amsterdam.

Hougaard P. (1999): Fundamentals of survival data. *Biometrics*, 55 (1), 13-22.

Hougaard P. (2000): *Analysis of Multivariate Survival Data*. Springer-Verlag, New York.

Jacod J. (1975): Multivariate point processes: predictable projection, Radon-Nikodym derivatives, representation of martingales. *Probability Theory and Related Fields*, 31:235-253.

Jacod J. (1979): *Calcul Stochastique et Problèmes de Martingales*. Springer-Verlag, Berlin.

Kaplan EL., Meier P. (1958): Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53, 457-481.

Therneau TM., Grambsch PM., Fleming TR. (1990): Martingale-based residuals for survival models. *Biometrika*, 77 (1), 147-160.

Therneau TM., Grambsch PM. (2000): *Modeling survival data. Extending the Cox model*. Springer-Verlag, New York.

ANNEX II: EL MÈTODE JACKKNIFE.

En aquest annex es farà una breu revisió del mètode Jackknife (veure més a Efron, 1982; Shao i Tu, 1995 i Wang *et al.*, 1997). Aquest mètode és una tècnica ideada per estimar el biaix i l'error estàndard d'un estimador (Quenouille, 1949; Tukey, 1958; Efron, 1992; Tibshirani, 1992).

Es suposa que es té una mostra $x = (x_1, x_2, \dots, x_n)$ i un estimador $\hat{\theta} = s(x)$; i que es vol estimar el biaix i l'error estàndard per $\hat{\theta}$. El mètode jackknife es centra en les mostres que s'obtenen en eliminar una a una les observacions de la mostra original:

$$X_{(i)} = (x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \quad (\text{A.II.1})$$

per $i=1, 2, \dots, n$, anomenades mostres jackknife. La mostra i -èsima jackknife consisteix en el conjunt de dades original sense l'observació i -èsima. Llavors es defineix com:

$$\hat{\theta}_{(i)} = s(x_{(i)}) \quad (\text{A.II.2})$$

com la i -èsima rèplica de l'estimador $\hat{\theta}$.

A partir d'aquí es defineix l'estimador jackknife pel biaix com:

$$\text{bias}_{\text{jack}} = (n-1)(\hat{\theta}_{(.)} - \hat{\theta}) \quad (\text{A.II.3})$$

on

$$\hat{\theta}_{(.)} = \sum_{i=1}^n \hat{\theta}_{(i)} / n \quad (\text{A.II.4})$$

L'estimador de jackknife per l'error estàndard es defineix com:

$$\text{se}_{\text{jack}} = \left[\frac{n-1}{n} \sum (\hat{\theta}_{(i)} - \hat{\theta}_{(.)})^2 \right]^{1/2} \quad (\text{A.II.5})$$

D'on provenen aquestes fórmules? Es comença amb se_{jack} . Igual que l'estimador bootstrap de l'error estàndard, la fórmula per aquest estimador és semblant a la mostra de desviacions estàndard dels n valors $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ obtinguts en eliminar les observacions de la mostra original una a una; excepte el factor inicial que és $(n-1)/n$ enlloc de $1/(n-1)$ o $1/n$. Naturalment, $(n-1)/n$ és més gran que $1/(n-1)$ o $1/n$. Aquest "factor d'augment" és necessari perquè les desviacions de jackknife

$$\left(\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)}\right)^2 \quad (\text{A.II.6})$$

tendeixen a ser més petites que les desviacions bootstrap

$$\left(\hat{\theta}^*(b) - \hat{\theta}^*(\cdot)\right)^2 \quad (\text{A.II.7})$$

i així, la mostra jackknife és més semblant a la mostra original x que la mostra bootstrap.

La forma exacta del factor $(n-1)/n$ prové de considerar el cas especial $\hat{\theta} = \bar{x}$. Llavors és fàcil veure que

$$se_{jack} = \left\{ \sum_1^n (x_i - \bar{x})^2 / \{(n-1)n\} \right\}^{1/2} \quad (\text{A.II.8})$$

És a dir, el factor $(n-1)/n$ és exactament el que es necessita per tal que l'estimador se_{jack} sigui igual a l'estimador sense biaix de l'error estàndard per la mitjana.

De forma similar, l'estimador de jackknife pel biaix (3) és un múltiple de la mitjana de les desviacions jackknife

$$\hat{\theta}_{(i)} - \hat{\theta}, \quad i=1,2,\dots,n. \quad (\text{A.II.9})$$

Aquestes quantitats es coneixen, a vegades, com els valors d'influència de jackknife. El factor multiplicador $(n-1)$ de (A.II.3) és un factor d'augment similar al que apareixia a la fórmula de l'estimador jackknife de l'error estàndard. Per veure d'on prové no

podem recórrer al cas especial $\hat{\theta} = \bar{x}$ tal com s'ha fet pel cas anterior, ja que \bar{x} no té biaix i $\hat{\theta}_{(i)} - \hat{\theta}$ és zero. Donat que aquest cas no explica la necessitat d'aquest factor, es considera la variància de la mostra

$$\hat{\theta} = \sum_1^n (x_i - \bar{x})^2 / n \quad (\text{A.II.10})$$

Aquesta té biaix igual a $-1/n$ vegades la variància de la població, i el factor $(n-1)$ multiplicant $\hat{\theta}_{(i)} - \hat{\theta}$ fa que l'estimador biasjack sigui igual a $-1/n$ vegades $\sum_1^n (x_i - \bar{x})^2 / (n-1)$, l'estimador sense biaix de la variància de la població.

Una altra forma de pensar en el mètode Jackknife és en termes dels pseudo-valors

$$\tilde{\theta}_i = n\hat{\theta} - (n-1)\hat{\theta}_{(i)} \quad (\text{A.II.11})$$

S'observa que en el cas especial $\hat{\theta} = \bar{x}$, tenim que $\tilde{\theta}_i = x_i$, l'i-èssim valor de les dades inicials. A més a més, per qualsevol $\hat{\theta}$, la fórmula per l'estimador se_{jack} es pot expressar com:

$$se_{jack} = \left\{ \sum_1^n (\tilde{\theta}_i - \tilde{\theta})^2 / \{(n-1)n\} \right\}^{1/2} \quad (\text{A.II.12})$$

on $\tilde{\theta} = \sum_{i=1}^n \tilde{\theta}_i / n$.

Aquesta fórmula fa pensar en un estimador de l'error estàndard per la mitjana per les dades $\tilde{\theta}_i$, $i=1,2,\dots,n$. La principal idea és que els pseudo-valors poden actuar com a n valors independents.

Què passa si s'intenta portar aquesta idea més enllà utilitzant aquests pseudo-valors per construir un interval de confiança? Una solució raonable podria ser la construcció del següent interval

$$\tilde{\theta} \pm t_{n-1}^{(1-\alpha)} se_{\text{jack}} \quad (\text{A.II.13})$$

on $t_{n-1}^{(1-\alpha)}$ és el percentil $(1-\alpha)$ de la distribució t amb $n-1$ graus de llibertat. Resulta, però, que aquest interval no treballa molt bé: en particular, no és significativament millor que els intervals basats en la teoria normal. Són necessàries, doncs, solucions més bones per la construcció d'intervals de confiança; però en aquest annex no s'explicaran ja que només es volia fer una petita introducció del mètode jackknife.

Referències

Efron B. (1982): *The Jackknife, the Bootstrap and other resampling plans*. SIAM, Philadelphia.

Efron B. (1992): Jackknife-after-bootstrap standard errors and influence functions. *Journal Royal of Statistical*, 54, 83-127.

Quenouille M. (1949): Aproximate tests of correlation in time series. *Journal Royal of Statistical*, 11, 18-44.

Shao J. i Tu D. (1995): *The Jackknife and Bootstrap*. Springer-Verlag, New York.

Tibshirani RJ. (1992): Discussion of Jackknife-after-bootstrap standard errors and influence functions. *Journal Royal of Statistical*, 54, 119-120.

Tukey JW. (1958): Bias and confidence in not quite larg samples. (Abstract). *Annals of Mathematical Statistics*, 29, 614.

Wang J., Sunil Rao J. i Shao J. (1997): Weighted Jackknife-after-bootstrap: a heuristic approach. *Proceedings of the 1997 Winter Simulation Conference*.

ANNEX III. CENTRE D'ATENCIÓ PRIMÀRIA (CAP) I HOSPITALS DE LA REGIÓ SANITÀRIA DE GIRONA.

La Regió Sanitària Girona comprèn els sectors sanitaris Alt Empordà, Baix Empordà, Garrotxa, Gironès, Pla de l'Estany, Ripollès i la Selva que pràcticament coincideixen amb les comarques del mateix nom amb l'excepció d'alguns municipis limítrofs. La població és de 657.948 habitants (tall de població de gener de 2006) i una extensió 5.533 Km², fet que representa el 17.28% del territori de Catalunya.

L'eix vertebrador del sistema sanitari és l'àrea bàsica de salut (ABS), unitat territorial elemental on es presta l'atenció primària de salut d'accés directe de la població. Cada ABS compta, com a mínim, amb un centre d'atenció primària (CAP) i es coordina amb tots els serveis sanitaris i sociosanitaris d'atenció primària del seu àmbit, així com amb els especialistes i hospitals que tingui assignats. Les funcions de l'ABS es desenvolupen quan es crea l'equip d'atenció primària (EAP), format per un conjunt de professionals sanitaris i no sanitaris, que actuen de manera integrada, mitjançant el treball en equip.

A continuació es mostra una llista dels centres d'assistència primària (CAP) que formen la Regió Sanitària de Girona.

Taula A.III.1. Centres d'assistència primària a la Regió Sanitària de Girona

Comarca	Centres d'assistència primària (CAP)
Alt Empordà	CAP Alt Empordà (Figueres) CAP Cadaqués (Cadaqués) CAP Doctor Subirós (La Jonquera) CAP Llançà (Llançà) CAP Dr. Moisès Broggi (L'Escala) CAP Roses (Roses) CAP Vilafant (Vilafant) CAP Bàscara (Bàscara) CAP Peralada (Peralada)
Baix Empordà	CAP La Bisbal d'Empordà (La Bisbal d'Empordà) CAP Doctor Josep Alsina i Bofill (Palafrugell) CAP Catalina Cargol Rigó (Palamós) CAP Sant Feliu de Guíxols (Sant Feliu de Guíxols)

	CAP Torroella de Montgrí (Torroella de Montgrí)
Garrotxa	CAP Besalú (Besalú) CAP Garrotxa (Olot) CAP Sant Joan les Fonts (Sant Joan les Fonts) CAP La Vall d'en Bas (St. Esteve d'en Bas)
Gironès	CAP Cassà de la Selva (Cassà de la Selva) CAP Celrà (Celrà) CAP Güell (Girona) CAP Girona 2 (Can Gibert del Pla) CAP Montilivi (ABS Girona 3) CAP Vila-roja (ABS Girona 3) CAP Santa Clara (ABS Girona 1) CAP Doctor Joan Vilaplana (ABS Girona 4) CAP Salt (Salt) CAP Sarrià de Ter (Sarrià de Ter)
Pla de l'Estany	CAP Banyoles (Banyoles)
Ripollès	CAP Campdevàrol (Campdevàrol) CAP Camprodon (Camprodon) CAP Ribes de Freser (Ribes de Freser) CAP Ripollès (Ripollès)
Selva	CAP Doctor Ramon Vinyes (Anglès) CAP Arbúcies (Arbúcies) CAP Blanes (Blanes) CAP Breda (Breda) CAP Hostalric (Hostalric) CAP Lloret de Mar (Lloret de Mar) CAP Maçanet de la Selva (Maçanet de la Selva) CAP Sant Hilari Sacalm (Sant Hilari Sacalm) CAP La Selva de Santa Coloma de Farners (Sta Coloma de Farners) CAP Sils (Sils) CAP Vidreres (Vidreres)

La xarxa hospitalària d'utilització pública (XHUP) està formada a la Regió Sanitària de Girona pels hospitals que es mostren a continuació:

Taula A.III.2. Hospitals de la Regió Sanitària de Girona

Comarca	Hospitals
Alt Empordà	Hospital de Figueres Fundació Privada
Baix Empordà	Hospital de Palamós
Garrotxa	Hospital Sant Jaume
Gironès	Clínica Girona, S.L. Hospital Universitari de Girona Doctor Josep Trueta Parc Hospitalari Martí i Julià (Hospital Santa Caterina)
Pla de l'Estany	Clínica Salus Infirmorum
Ripollès	Hospital de Campdevàrol
Selva	Hospital Comarcal de Blanes

ANNEX IV. SINTAXIS I MODELS

A.IV.1. ANÀLISI ANDERSEN.

A.IV.1.1. Sintaxis

###IMPORTAR DADES

```
library(foreign)
datos<-read.dbf("dadesMarc.dbf")
names(datos)<-casefold(names(datos))
attach(datos)
```

###Construcció d'una variable categòrica (per quintils)

```
edatc<-cut(edat,quantile(edat,c(0,0.2,0.4,0.6,0.8,1),na.rm=T))
levels(edatc)<-c(0,1,2,3,4)
datos$edatc<-edatc
attach(datos)
options(constrasts=c("contr.treatment","contr.poly"))
```

```
pseudo<-gemma
pseudo[is.na(pseudo)]<-0
obs<- rbinom(length(pseudo),1,pseudo)
```

```
Sexe<-as.numeric(sexe)
edatc1<-as.numeric(edatc)*0
edatc1[edatc==1]<-1
edatc1[is.na(edatc1)]<-0
edatc2<-as.numeric(edatc)*0
edatc2[edatc==2]<-1
edatc2[is.na(edatc2)]<-0
edatc3<-as.numeric(edatc)*0
edatc3[edatc==3]<-1
edatc3[is.na(edatc3)]<-0
edatc4<-as.numeric(edatc)*0
edatc4[edatc==4]<-1
edatc4[is.na(edatc4)]<-0
```

Construcció variables fictícies dicotòmiques per enum

```
enum[enum==0]<-1
enum[enum>6]<-6
```

```
enum2<-enum*0
enum2[enum==2]<-1
enum3<-enum*0
enum3[enum==3]<-1
enum4<-enum*0
enum4[enum==4]<-1
enum5<-enum*0
enum5[enum==5]<-1
enum6<-enum*0
enum6[enum==6]<-1
```

```

Valcohol<-as.numeric(valcohol)
Vmpoc<-as.numeric(vmpoc)
Insufcar<-as.numeric(insufcar)
Hta<-as.numeric(hta)
Dislip<-as.numeric(dislip)
Dm2<-as.numeric(dm2)
Tab<-as.numeric(tab)
Extab<-as.numeric(extab)

```

```

Nvisites<-
nvisites12+nvisites31+nvisites32+nvisites41+nvisites42+nvisites43+nvisites51+nvisites
52+nvisites53+nvisites54+nvisites61+nvisites62+nvisites63+nvisites64+nvisites65

```

ESTIMACIÓ BAYESIANA (SENSE EFECTES ESPACIALS)

```
library(R2WinBUGS)
```

```

dataRSG <- list(N=length(obs),nivell=id,K= max(id), Y=obs, x12=enum2,x13=enum3,
x14=enum4, x15=enum5, x16=enum6,x1=Sexe,
x21=edatc1,x22=edatc2,x23=edatc3,x24=edatc4,x3=Valcohol, x5=Insufcar, x6=Hta,
x7=Dislip, x9=Tab,x10=Extab)

```

```

parametersRSG <- c("b0","c12","c13","c14","c15","c16",
"b1","b21","b22","b23","b24","b3","b5","b6","b7","b9","b10","sigma.beta0")

```

```

initsRSG <- function(){list(beta0= runif(max(id),0,1), b0=rnorm(1), c12=rnorm(1),
c13=rnorm(1), c14=rnorm(1), c15=rnorm(1), c16=rnorm(1), b1= rnorm(1), b21=
rnorm(1), b22= rnorm(1), b23= rnorm(1), b24= rnorm(1), b3= rnorm(1), b5= rnorm(1),
b6= rnorm(1), b7= rnorm(1), b9= rnorm(1), b10= rnorm(1), sigma.beta0=runif(1,0,5))}

```

```
memory.limit(4000)
```

```

simu.hierar <- bugs(dataRSG, initsRSG, parametersRSG,model.file =
"Modelo.txt",n.chains = 2, n.iter =5000 , n.burnin = 500, n.thin = 30,
debug = FALSE, DIC = FALSE, digits = 5, codaPkg = FALSE,
bugs.directory = "C:/Archivos de programa/WinBUGS14",working.directory = NULL)

```

```
print(simu.hierar,digits=5)
```

ESTIMACIÓ BAYESIANA AMB EFECTES ESPACIALS (BYM)

```

library(maptools)
library(tripack)
library(spdep)
library(SparseM)
library("RColorBrewer")

```

Mapes inicials

```
x1 <- read.shape("Rsg.shp")
```

```
plot.Map(x1,xlab="Lon",ylab="Lat")
```

Matriu d'adjacència

```
x.2 <- Map2poly(x1)
rsg <- poly2nb(x.2)
```

Ordre de los dades (en funció del mapa en format shape)

```
CODIGO<- x1$att$ CODI_INE
orden4<-match(datos$muni,CODIGO)

Recode<-cbind(1:length(unique(orden4)),sort(unique(orden4)))
Muni<-orden4
for(i in 1:length(unique(orden4))){
  Muni[Muni==Recode[i,2]]<-Recode[i, 1]}

```

Anàlisis

```
dataRSG <- list(N=length(obs),nivell=idindv,K= max(idindv),
N1=length(rsg),level=Muni,Y=obs, x1=Sexe,x11=Tempmod1, x12=Tempmod2,
x13=Tempmod3, x14=Tempmod4,
x21=edatc1,x22=edatc2,x23=edatc3,x24=edatc4,x3=Valcohol,x4=Vmpoc,
x5=Insufcar,x6=Hta,x7=Dislip,x8=Dm2,x9=Tab,x10=Extab,x11=Nvisites,
num=sapply(rsg,length), adj=unlist(rsg),sumNumNeigh=length(unlist(rsg)))
```

```
parametersRSG <- c("b0","b1","b21","b22","b23","b24","c11","c12",
"b3","b4","b5","b6","b7","b8","b9","b10","b11", "sigma.beta0","sdS")
```

```
initsRSG <- function(){list(beta0= runif(max(idindv),0,1), b0=rnorm(1), c11=rnorm(1),
c12=rnorm(1), c13=rnorm(1), c14=rnorm(1),b1= rnorm(1), b21= rnorm(1), b22=
rnorm(1), b23= rnorm(1), b24= rnorm(1), b3= rnorm(1),b4= rnorm(1),b5=rnorm(1),
b6=rnorm(1), b7=rnorm(1), b8=rnorm(1), b9=rnorm(1),
b10=rnorm(1),b11=rnorm(1),sigma.beta0=runif(1,0,5),
S=rnorm(sapply(rsg,length)),sdS =runif(1,0,5))}
```

```
memory.limit(4000)
```

```
simu.hiarar <- bugs(dataRSG, initsRSG, parametresRSG,model.file =
"ModeloBYMvarexpSIN.txt",n.chains = 2, n.iter =10000 , n.burnin = 1000,n.thin = 30,
debug = FALSE, DIC = FALSE, digits = 5, codaPkg = FALSE,bugs.directory =
"C:/Archivos de programa/WinBUGS14", working.directory = NULL)
```

```
print(simu.hiarar,digits=5)
```


A.IV. 1.2. Model Andersen sense efectes espacials.

```
model
{
  for (i in 1 : N) {
    enum[i]<-c12*x12[i]+c13*x13[i]+c14*x14[i]+c15*x15[i]+c16*x16[i]
    x2[i]<-b21*x21[i]+b22*x22[i]+b23*x23[i]+b24*x24[i]
    Y[i]~dbern(p[i])
    cloglog(p[i])<-
beta0[nivell[i]]+enum[i]+b1*x1[i]+x2[i]+b3*x3[i]+b5*x5[i]+b6*x6[i]+b7*x7[i]+b9*x9[i]+b10*
x10[i]

    proba[i]<-exp(-
exp(beta0[nivell[i]]+enum[i]+b1*x1[i]+x2[i]+b3*x3[i]+b5*x5[i]+b6*x6[i]+b7*x7[i]+b9*x9[i]+
b10*x10[i]))
  }

  # Priors:

  for(i in 1:K){beta0[i] ~ dnorm(b0,tau.beta0)}
  tau.beta0<-pow(sigma.beta0,-2)
  sigma.beta0 ~ dunif(0,5)
  b0 ~ dnorm(0,0.005)
  c12 ~ dnorm(0,1.0E-6)
  c13 ~ dnorm(0,1.0E-6)
  c14 ~ dnorm(0,1.0E-6)
  c15 ~ dnorm(0,1.0E-6)
  c16 ~ dnorm(0,1.0E-6)
  b1 ~ dnorm(0,0.005)
  b21 ~ dnorm(0,0.005)
  b22 ~ dnorm(0,0.005)
  b23 ~ dnorm(0,0.005)
  b24 ~ dnorm(0,0.005)
  b3 ~ dnorm(0,0.05)
  b5 ~ dnorm(0,0.05)
  b6 ~ dnorm(0,0.005)
  b7 ~ dnorm(0,0.005)
  b9 ~ dnorm(0,0.005)
  b10 ~ dnorm(0,0.005)
}
```

A.IV.1.3. Model Andersen amb efectes espacials.

```
model
{
  for (i in 1 : N) {

    Y[i] ~ dbern(p[i])
    cloglog(p[i])<-b0+S[level[i]]
  }

# CAR prior distribution for spatial random effects:
S[1:N1] ~ car.normal(adj[], weights[], num[], tau.S)
for(k in 1:sumNumNeigh) {
  weights[k] <- 1 }

# Priors:

b0 ~ dnorm(0,0.005)
tau.S <- pow(sdS,-2)
sdS ~ dunif(0,5)
}
```

A.IV.2. ANÀLISI GOLDSTEIN.

A.IV.2.1. Sintaxis

###Importar dades

```
library(foreign)
datos<-read.spss("dadesGoldstein.sav",use.value.labels=FALSE,
to.data.frame=TRUE)
names(datos)<-casefold(names(datos))
attach(datos)
```

###Construcció d'una variable categòrica (per quintils)

```
edatc<-cut(edat,quantile(edat,c(0,0.2,0.4,0.6,0.8,1),na.rm=T))
levels(edatc)<-c(0,1,2,3,4)
datos$edatc<-edatc
attach(datos)
options(constrasts=c("contr.treatment","contr.poly"))
```

```
obs<-y
```

```
Sexe<-as.numeric(sexe)
edatc1<-as.numeric(edatc)*0
edatc1[edatc==1]<-1
edatc1[is.na(edatc1)]<-0
edatc2<-as.numeric(edatc)*0
edatc2[edatc==2]<-1
edatc2[is.na(edatc2)]<-0
edatc3<-as.numeric(edatc)*0
edatc3[edatc==3]<-1
```

```

edatc3[is.na(edatc3)]<-0
edatc4<-as.numeric(edatc)*0
edatc4[edatc==4]<-1
edatc4[is.na(edatc4)]<-0

```

tempmod s'introdueix no paramètricament

Primer s'estandarditza

```
Tempmod<-(tempmod-mean(tempmod,na.rm=T))/sd(tempmod,na.rm=T)
```

Restricted cubic splines

```
library(splines)
```

```
Tempmod1<-ns(Tempmod,4)[,1]
Tempmod2<-ns(Tempmod,4)[,2]
Tempmod3<-ns(Tempmod,4)[,3]
Tempmod4<-ns(Tempmod,4)[,4]

```

```

Valcohol<-as.numeric(valcohol)
Vmpoc<-as.numeric(vmpoc)
Insufcar<-as.numeric(insufcar)
Hta<-as.numeric(hta)
Dislip<-as.numeric(dislip)
Dm2<-as.numeric(dm2)
Tab<-as.numeric(tab)
Extab<-as.numeric(extab)
Nvisites<-as.numeric(nvisites)

```

ESTIMACIÓ BAYESIANA (SENSE EFECTES ESPACIALS)

```
library(R2WinBUGS)
```

```

dataRSG <- list(N=length(obs),nivell=idindv,K= max(idindv), Y=obs,
x1=Sexe,x11=Tempmod1, x12=Tempmod2, x13=Tempmod3, x14=Tempmod4,
x21=edatc1,x22=edatc2,x23=edatc3,x24=edatc4,x3=Valcohol,x4=Vmpoc,
x5=Insufcar,x6=Hta,x7=Dislip,x8=Dm2,x9=Tab,x10=Extab,x11=Nvisites)

```

```

parametersRSG <- c("b0","b1","b21","b22","b23","b24","c11","c12",
"b3","b4","b5","b6","b7","b8","b9","b10","b11",
"sigma.beta0")

```

```

initsRSG <- function(){list(beta0= runif(max(idindv),0,1), b0=rnorm(1), c11=rnorm(1),
c12=rnorm(1), c13=rnorm(1), c14=rnorm(1),b1= rnorm(1), b21= rnorm(1), b22=
rnorm(1), b23= rnorm(1), b24= rnorm(1), b3= rnorm(1),b4= rnorm(1),b5=rnorm(1),
b6=rnorm(1), b7=rnorm(1), b8=rnorm(1), b9=rnorm(1),
b10=rnorm(1),b11=rnorm(1),sigma.beta0=runif(1,0,5))}

```

```
memory.limit(4000)
```

```
simu.hierar <- bugs(dataRSG, initsRSG, parametersRSG,model.file =
"Modelo.txt",n.chains = 2, n.iter =10000 , n.burnin = 1000,n.thin = 30, debug = FALSE,
DIC = FALSE, digits = 5, codaPkg = FALSE,bugs.directory = "C:/Archivos de
programa/WinBUGS14",working.directory = NULL)
```

```
print(simu.hierar,digits=5)
```

ESTIMACIÓ BAYESIANA AMB EFECTES ESPACIALS (BYM)

```
library(maptools)
library(tripack)
library(spdep)
library(SparseM)
library("RColorBrewer")
```

Mapes inicials

```
x1 <- read.shape("Rsg.shp")

plot.Map(x1,xlab="Lon",ylab="Lat")
```

Matriu d'adjacència

```
x.2 <- Map2poly(x1)
rsg <- poly2nb(x.2)
```

Ordre de les dades (en funció del mapa en format shape)

```
CODIGO<- x1$att$ CODI_INE
orden4<-match(datos$muni,CODIGO)

Recode<-cbind(1:length(unique(orden4)),sort(unique(orden4)))
Muni<-orden4
for(i in 1:length(unique(orden4))){
  Muni[Muni==Recode[i,2]]<-Recode[i, 1]}
```

Anàlisis

```
dataRSG <- list(N=length(obs),nivell=idindv,K= max(idindv),
N1=length(rsg),level=Muni,Y=obs, x1=Sexe,x11=Tempmod1, x12=Tempmod2,
x13=Tempmod3, x14=Tempmod4,
x21=edatc1,x22=edatc2,x23=edatc3,x24=edatc4,x3=Valcohol,x4=Vm poc,
x5=Insufcar,x6=Hta,x7=Dislip,x8=Dm2,x9=Tab,x10=Extab,x11=Nvisites,
num=sapply(rsg,length), adj=unlist(rsg),sumNumNeigh=length(unlist(rsg)))
```

```
# parametersRSG <- c("b0","b1","b21","b22","b23","b24","c11","c12",
# "c13","c14","b3","b4","b5","b6","b7","b8","b9","b10","b11",
# "sigma.beta0","sdS")
```

```

initsRSG <- function(){list(beta0= runif(max(idindv),0,1), b0=rnorm(1), c11=rnorm(1),
c12=rnorm(1), c13=rnorm(1), c14=rnorm(1),b1= rnorm(1), b21= rnorm(1), b22=
rnorm(1), b23= rnorm(1), b24= rnorm(1), b3= rnorm(1),b4= rnorm(1),b5=rnorm(1),
b6=rnorm(1), b7=rnorm(1), b8=rnorm(1), b9=rnorm(1),
b10=rnorm(1),b11=rnorm(1),sigma.beta0=runif(1,0,5),
S=rnorm(sapply(rsg,length)),sdS =runif(1,0,5))}

```

```
memory.limit(4000)
```

```

simu.hierar <- bugs(dataRSG, initsRSG, parametersRSG,model.file =
"ModeloBYMvarexpSIN.txt",n.chains = 2, n.iter =100 , n.burnin = 10,n.thin = 30, debug
= FALSE, DIC = FALSE, digits = 5, codaPkg = FALSE,bugs.directory = "C:/Archivos de
programa/WinBUGS14", working.directory = NULL)

```

```
print(simu.hierar,digits=5)
```

A.IV.2.2. Model Goldstein sense efectes espacials

```

model
{
  for (i in 1 : N) {
    rcubic[i]<-c11*x11[i]+c12*x12[i]+c13*x13[i]+c14*x14[i]
    x2[i]<-b21*x21[i]+b22*x22[i]+b23*x23[i]+b24*x24[i]
    Y[i] ~ dbern(p[i])
    logit(p[i])<-
beta0[nivell[i]]+rcubic[i]+b1*x1[i]+x2[i]+b3*x3[i]+b4*x4[i]+b5*x5[i]+b6*x6[i]+b7*x7[i]+b8*
x8[i]+b9*x9[i]+b10*x10[i]+b11*x11[i]

    proba[i]<-
exp(beta0[nivell[i]]+rcubic[i]+b1*x1[i]+x2[i]+b3*x3[i]+b4*x4[i]+b5*x5[i]+b6*x6[i]+b7*x7[i]+
b8*x8[i]+b9*x9[i]+
b10*x10[i]+b11*x11[i])/(1+exp(beta0[nivell[i]]+rcubic[i]+b1*x1[i]+x2[i]+b3*x3[i]+b4*x4[i]+
b5*x5[i]+b6*x6[i]+b7*x7[i]+b8*x8[i]+b9*x9[i]+
b10*x10[i]+b11*x11[i]))
  }

# Priors:

for(i in 1:K){beta0[i] ~ dnorm(b0,tau.beta0)}
tau.beta0<-pow(sigma.beta0,-2)
sigma.beta0 ~ dunif(0,5)
b0 ~ dnorm(0,0.005)
c11 ~ dnorm(0,1.0E-6)
c12 ~ dnorm(0,1.0E-6)
c13 ~ dnorm(0,1.0E-6)
c14 ~ dnorm(0,1.0E-6)
b1 ~ dnorm(0,0.005)
b21 ~ dnorm(0,0.005)
b22 ~ dnorm(0,0.005)
b23 ~ dnorm(0,0.005)
b24 ~ dnorm(0,0.005)
b3 ~ dnorm(0,0.005)

```

```

b4 ~ dnorm(0,0.005)
b5 ~ dnorm(0,0.005)
b6 ~ dnorm(0,0.005)
b7 ~ dnorm(0,0.005)
b8 ~ dnorm(0,0.005)
b9 ~ dnorm(0,0.005)
b10 ~ dnorm(0,0.005)
b11 ~ dnorm(0,0.005)
for(k in 1:5){beta[k]~dnorm(0,0.005)}
}

```

A.IV.2.3. Model Goldstein amb efectes espacials.

```

model
{
  for (i in 1 : N) {
    rcubic[i]<-c11*x11[i]+c12*x12[i]+c13*x13[i]+c14*x14[i]
    x2[i]<-b21*x21[i]+b22*x22[i]+b23*x23[i]+b24*x24[i]
    Y[i] ~ dbern(p[i])
    logit(p[i])<-
beta0[nivell[i]]+rcubic[i]+b1*x1[i]+x2[i]+b3*x3[i]+b4*x4[i]+b5*x5[i]+b6*x6[i]+b7*x7[i]+b8*
x8[i]+b9*x9[i]+
      b10*x10[i]+b11*x11[i]+S[level[i]]
    proba[i]<-
exp(beta0[nivell[i]]+rcubic[i]+b1*x1[i]+x2[i]+b3*x3[i]+b4*x4[i]+b5*x5[i]+b6*x6[i]+b7*x7[i]+
b8*x8[i]+b9*x9[i]+
      b10*x10[i]+b11*x11[i]+S[level[i]])/(1+exp(beta0[nivell[i]]+rcubic[i]+b1*x1[i]+x2[i]+b3*x3[i]
+b4*x4[i]+b5*x5[i]+b6*x6[i]+b7*x7[i]+b8*x8[i]+b9*x9[i]+
      b10*x10[i]+b11*x11[i]+S[level[i]]))
  }
}

# CAR prior distribution for spatial random effects:
S[1:N1] ~ car.normal(adj[], weights[], num[], tau.S)
for(k in 1:sumNumNeigh) {
  weights[k] <- 1 }

# Priors:

for(i in 1:K){beta0[i] ~ dnorm(b0,tau.beta0)}
tau.beta0<-pow(sigma.beta0,-2)
sigma.beta0 ~ dunif(0,5)
b0 ~ dnorm(0,0.005)
c11 ~ dnorm(0,1.0E-6)
c12 ~ dnorm(0,1.0E-6)
c13 ~ dnorm(0,1.0E-6)
c14 ~ dnorm(0,1.0E-6)
b1 ~ dnorm(0,0.005)
b21 ~ dnorm(0,0.005)
b22 ~ dnorm(0,0.005)
b23 ~ dnorm(0,0.005)
b24 ~ dnorm(0,0.005)
b3 ~ dnorm(0,0.005)
b4 ~ dnorm(0,0.005)

```

```
b5 ~ dnorm(0,0.005)
b6 ~ dnorm(0,0.005)
b7 ~ dnorm(0,0.005)
b8 ~ dnorm(0,0.005)
b9 ~ dnorm(0,0.005)
b10 ~ dnorm(0,0.005)
b11 ~ dnorm(0,0.005)
for(k in 1:5){beta[k]~dnorm(0,0.005)}
tau.S <- pow(sdS,-2)
sdS ~ dunif(0,5)
}
```

ANNEX V: MODELLING OF MULTIVARIATE SURVIVAL DATA BY MEANS OF GENERALIZED LINEAR MIXED MODELS. THE CASE OF HOSPITAL ADMISSIONS DUE TO CHRONIC OBSTRUCTIVE PULMONARY DISEASE.

Modelling of multivariate survival data by means of Generalized Linear Mixed Models. The case of hospital admissions due to chronic obstructive pulmonary disease

Gemma Renart^{1,2}, Maria Antònia Barceló^{1,2}, Marc Saez^{1,2}

1 Research Group on Statistics, Applied Economics and Health, GRECS, University of Girona, Spain

2 CIBER of Epidemiology and Public Health (CIBERESP), Spain

Corresponding author:

Prof. Marc Saez, PhD, CStat
Research Group on Statistics, Applied Economics and Health (GRECS) and CIBER
Epidemiología y Salud Pública (CIBERESP)
University of Girona
Campus de Montilivi, 17071 Girona
Tel 34-972-418736, Fax 34-972-418032
<http://www3.udg.edu/fcee/economia/english/grecs.htm>
e-mail: marc.saez@udg.edu

Acknowledgements

We would like to thank Meia Faixedas, Responsible of Epidemiology and Programmes, Girona Health Region (*Regió Sanitària Girona, CatSalut*), and Delfi Faixedas, Girona Primary Health Care, Catalan Health Institut (*ICS*) for providing the data.

Abstract

In multivariate survival analysis the individuals can present more than one event (or recurrences) in the observation period. This fact demands a different methodology from that used in univariate survival analysis because duration between recurrences could be correlated due to the presence of unobserved individual factors.

Until recently, the most common way of dealing with this type of situation in survival analysis was by using marginal models or using the conditional models. This article explains a specific analysis method for this situation.

The aim of this study is to model multivariate survival data, based on generalised linear mixed models (GLMM) applied to a database of hospital admissions due to chronic obstructive pulmonary disease (COPD). Specifically, the method that will be used is a special case of the nonlinear multivariate model proposed by Goldstein (1991).

Keywords: Multivariate survival data, conditional models, generalised linear mixed models, chronic obstructive pulmonary disease.

1. Introduction

Survival research is interested in the time that passes from the beginning of the study until the event of interest occurs. However, it is very common to find individuals who experience this event more than once during the period of study. In this case, a different methodology needs to be used to that of the standard univariate survival analysis.

This article explains a specific analysis method for this situation, applied to a database of hospital admissions due to chronic obstructive pulmonary disease (COPD), where the variable to be modelled is the risk that an individual has of being readmitted to hospital due to COPD. Therefore, it is about studying different recurrences of the same event.

In this case, the duration between recurrences could be correlated due to the presence of unobserved individual factors. This type of event is normally dealt with by introducing individual random effects in the model, resulting in a multivariate model. The random effects represent the individual "frailty" and the variance of these effects measures the unobserved heterogeneity between individuals (Goldstein *et al.*, 2004b; Barceló, 2001).

Until recently, the most common way of dealing with this type of situation in survival analysis was by using marginal models such as the robust covariance matrix estimation in the Andersen-Gill approximation (Therneau and Hamilton, 1997); the Wei, Lin and Weissfeld method (Wei *et al.*, 1989) or the Prentice, Williams and Peterson method (Prentice *et al.*, 1981); or using the conditional models such as the frailty models (EM algorithm) (Therneau and Grambsch, 2000). Barceló (2001) and Barceló and Saez (2004) proposed a modification to the frailty model that also enabled not only the conditional mean but also the variance of the frailty to be modelled simultaneously.

The aim of this study is to model multivariate survival data, based on generalised linear mixed models (GLMM). Specifically, the method that will be used is a special case of the nonlinear multivariate model proposed by Goldstein (1991), and developed by Steele *et al.* (1996) in other context, to work with competing risk models and by Goldstein himself (Goldstein *et al.*, 2002) to work with competing risk models with multiple statuses of origin. Finally, Steele, Goldstein and Browne (Goldstein *et al.*, 2004b; Steele, 2008; Goldstein *et al.*, 2004a) combined all these extensions to obtain a more general multilevel competing risk model.

Currently one of the most common respiratory diseases is chronic obstructive pulmonary disease (COPD) which has become one of the main causes of morbidity and mortality in the world. This disease affects more than 52 million people worldwide (Pauwels *et al.*, 2001) and according to the World Health Organisation (WHO), each year it causes the death of at least 2.9 million people. In Spain this figure is at more than 18,000 people (INE, 1998).

Various epidemiological population-based studies have been carried out in Spain, the most important of which is the IBERPOC study (Sobradillo *et al.*, 1999; Sobradillo *et al.*, 2000), designed to measure the prevalence and variation in the distribution of COPD. This study, carried out during 1996-1997, estimated the prevalence for the adult population between 40 and 69 years of age to be 9.1%. Unfortunately this data has not improved in recent studies (Izquierdo, 2003; Halbert *et al.*, 2006) which affirm that the prevalence of COPD in the general population is estimated at around 1%, and increases to 8-10% or more in adults over 40.

Around 8% of Spaniards between 20 and 44 years of age suffer the first symptoms of the disease, prevalence which is the highest on the European continent and which predicts a worsening of the problem in the future (Cerveri *et al.*, 2001; Miravitlles, 2005). This prevalence is related to the high level of tobacco smoking in the Spanish population, that has remained stable or has decreased amongst men, but that has increased considerably amongst women (Peris *et al.*, 2004).

Global estimations of mortality carried out in 1990 (Murray and Lopez, 1996; Murray and Lopez, 1997) have recently been updated (Lopez *et al.*, 2006) and reiterate that COPD, which was the fifth cause of death in 1990, has since the year 2000 become the fourth and will become the third cause by 2020.

Given this situation it is not surprising that the economic repercussions and the use of health resources for COPD are considerable. In Spain it is calculated that there are 280,000 (prevalence) diagnosed and treated cases of this disease (Sobradillo *et al.*, 2000) and that each diagnosed patient generates an annual average expense of between 900 € and 3,000 € (Miravitlles *et al.*, 2003; Masa *et al.*, 2004).

Chronic obstructive pulmonary disease is however an illness that should originally be diagnosed and treated by the primary healthcare services. If this were the case, hospital readmissions should not be found due to this cause and both the economic cost and the use of resources generated by this disease would decrease. At present there are various studies that confirm this fact, given that they demonstrate that an adequate control of the disease by the primary healthcare services would reduce the number of readmissions to hospital and would improve the quality of life of those with the disease (Kaag *et al.*, 1996; Hermiz *et al.*, 2002; Wilkinson *et al.*, 2004; Gadorury *et al.*, 2005; Casas *et al.*, 2006). On the other hand, the study of these readmissions could also be useful to evaluate the treatment given within the hospital for this type of patient, as well as to try and analyse the factors associated with the duration of the stay in hospital and to find out the hospital social security impact that this diseases causes (Librero *et al.*, 1999; Soler *et al.*, 2001; Verdager *et al.*, 2003).

The objective of this study is twofold. First, to assess which factors influence the fact that an individual returns to hospital. Second, because readmissions could be actually considered multivariate survival data, mixed models are used to approach multivariate survival analysis.

Section 2 explains the method used whilst Section 3 presents the results obtained and finally Section 4 will discuss the same.

2. Method

Data.

The database used was created based on two different sources of information: on one hand, the *Conjunt Mínim Bàsic de Dades de l'Alta Hospitalària, CMBDAH* (Basic Minimum Group of Hospital Admission Data), for 2003, which provides information regarding all the hospital admissions that occurred in hospitals in the Healthcare Region of Girona (RSG) (province of Girona excluding the administrative division of Cerdanya) during 2003 of both RSG residents and non-residents. And, on the other hand, information was gathered from the *Institut Català de la Salut, ICS* (Catalan Healthcare Institute), with regard to all those users attended to during 2003 in any of their primary healthcare centres, such as for example the number and type of visits made by each user or their health problems.

After crossing the information provided by these two sources, the admissions of RSG residents, diagnosed with COPD in one of the main or secondary diagnosis, classified as medical episodes and corresponding to the higher respiratory or cardiac diagnostic category, were selected, eliminating admissions corresponding to the type of day hospital activity. This was the way in which the definitive database was obtained, which included all the individuals resident in the RSG who were admitted at least once during the first quarter of 2003, with a main or secondary diagnosis at admission to hospital of chronic obstructive pulmonary disease. Some of the variables included are gender, town and age of the patient, number of admissions of each individual and the different diagnosis for each admission, variables relating to health problems that the patient could suffer (alcoholism, COPD, cardiac insufficiency, arterial hypertension, dyslipemia, diabetes mellitus type 2 and smoking) and finally the number of visits made by the patient to the corresponding basic healthcare area during the first quarter of 2003. It should be taking into account that health problems were diagnosed at the primary health care level. In this sense, COPD diagnosis at admission, used as inclusion criteria, did not necessarily imply that the individual was also diagnosed at the primary health care level.

Methods

The method proposed consists of a general model for multivariate survival data analysis. The principal idea of this model is based on using a logistic regression model after restructuring the database, converting each survival time into a binary sequence called response.

In the present study, the database is structured on two levels: the individual represented by the subindex i , and the episode or recurrence corresponding to subindex k . Finally, with the subindex h , it is represented the status of the patient who could be or not be admitted.

The objective is to build a model for the probability of a change in status that is from no admission to hospital admission, occurring at any interval. It is assumed that this probability depends on the status of the individual at the time of the study, the permanence in this status, of some covariables and of the random effects. This probability (λ) is known as risk function and in this study is defined as:

$$\lambda_{hik(t)} = P(y_{hik(t)}=1 | y_{hik(t-1)}=0) \quad (1)$$

where, as previously mentioned, the subindex i refers to the individual, k to the episode and h to the status; and y is a binary response variable with a value of 1 if the status has changed and 0 if it has not. Presumably this variable is distributed as a Binomial, $y_{hik(t)} \sim \text{Binomial}(1, \lambda_{hik(t)})$.

Table 1 shows the construction of this response variable y using a practical case of two individuals, a woman and a man, that have been admitted to hospital once and twice respectively in a period of ten days.

Subsequently, to predict this probability a logit model can be used working with the logarithm of the odds ratios of the change in status, in other words, the ratio of the probability of change in status, with respect to the probability of no change. The general model used is shown below:

$$\log\left(\frac{\lambda_{hik(t)}}{(1-\lambda_{hik(t)})}\right) = \text{logit}(\lambda_{hik(t)}) = \beta_0 + \sum_{j=1}^p \beta_j x_{jhik(t)} + f(t) + u_i + e_{ik} \quad (2)$$

where the terms u_i and e_{ik} refer to the individual variation and that of the episodes respectively, $x_{jhik(t)}$ are the explicative variables and $f(t)$ is a function of time of permanence in the current status. For this function a set of dummy variables (blocking factors) could be used for the different intervals 1, 2, ..., n , where n is the maximum number of intervals for any status. However a problem arises when there are few episodes of long duration, given that this would imply the need of many dummy variables and therefore some of them could have an inefficient estimation. For this study an attempt has been made to model this, by means of a smoothing function using cubic splines (continuous mathematical functions defined piecewise by polynomials) (Remontet *et al.*, 2007). However, given that sometimes these functions can perform badly at the tail end, the smoothing function $f(t)$ of the model has been approximated based on restricted cubic splines (Harrell *et al.*, 1988). This type of spline is similar to the cubic

splines with the difference that the part prior to the first breakpoint and after the last is linear, a first degree polynomial.

In general, if the breakpoints are defined as k, t_1, t_2, \dots, t_k , the restricted cubic splines can be adjusted creating $k-2$ derived variables (Harrell *et al.*, 1988; Durrleman and Simon, 1989) and its representation is:

$$f(x) = \beta_0 + \beta_1 x + \sum_{j=1}^{k-2} \beta_{j2} \left[(x - t_j)^3 + \frac{t_k - t_j}{t_k - t_{k-1}} (x - t_{k-1})^3 + \frac{t_{k-1} - t_j}{t_k - t_{k-1}} (x - t_k)^3 \right] \quad (3)$$

where $\theta = (\beta_0, \beta_1, \beta_{12}, \dots, \beta_{1_{k-2}})$ is the vector of coefficients.

In this study four breakpoints have been used corresponding to the quartiles of the set of data used.

Finally, therefore, the model used is:

$$\text{logit}(\lambda_{hik(t)}) = \beta_{0i} + \sum_{j=1}^p \beta_j x_{jhik(t)} + f(t) + u_i^{(h)} \quad (4)$$

where $u_i^{(h)}$ represents the random effects for the individual i in the status h .

The random effects associated with the episode (e_{ik}) have been excluded from the model, given that no differences are shown in any of them. However, the independent term β_{0i} is a random effect specific to each individual.

Estimation model

In this case a Bayesian approximation (Saez and Saurina, 2007) has been chosen for use because it works with very large dimension matrixes and with the frequentist approximation this is an unapproachable task using computational means.

In order to carry out the estimation from the Bayesian point of view, the distributions must be fixed a priori for the random effects and for the rest of the parameters (both of the constraint function, of the variance function, and for the parameters of the probability distributions of the random effects) and estimate the distribution a posteriori.

The selection of the priors that will be used in each case is subjective and often based on the judgement of the experts and on the type of data available. When a posterior distribution is of the same family as that of the prior distribution used, they are referred to as conjugate

distributions. The use of these distributions implies that the prior parameters have good mathematical properties for the calculation of the posterior distributions. The calculation of the posterior probabilities in the case of not using conjugated priors is complicated, but currently it is possible using the Markov chain Monte Carlo.

In any case the sensibility analysis has to play an important role in this type of inference, in order to investigate the impact that the choice of priors has on the results and conclusions of the analysis. Ultimately, using Bayesian methods of Markov chain Monte Carlo (Gilks *et al.*, 1996; Spiegelhalter *et al.*, 2003), the unknown parameters are estimated.

As previously mentioned, the complete Bayesian estimation requires the “Likelihood” to be provided, in other words, that the model and the prior distribution of the parameters of interest be specified. Specifically, prior distributions for the random effects (normal interchangeable for heterogeneity), for the parameters of the model, as well as for the standard deviations of the random effects.

In the case of the standard deviations of the random effects, vague or weak informative prior distributions can be selected. The problem in this case is that the selection of an uninformative and “proper” prior (in the sense that its integral is equal to the unit), implies that the posterior inference depends on the parameters of the prior distribution. This (excessive) sensibility is particularly problematic if there is not sufficient evidence of the existence of heterogeneity between units. For this reason a prior has been chosen for use, that does not present this problem, a uniform distribution in the finite range (from 0 to 5) for the standard deviation of the random effects (Gelman, 2005).

Once the model (likelihood) and the prior distributions have been specified, the posterior distribution is obtained. When various parameters have to be estimated, as in this study (standard deviations of the random effects, parameters), multiple integrals are used which makes the use of simulation methods necessary given that the results cannot be obtained directly. Specifically, the dependent sampling is used based on the Markov chain, MCMC (Gilks *et al.*, 1996; Elliot and Best, 2005). There are various algorithms for designing Markov chains. The most well-known is Gibbs sampling (Geman and Geman, 1984; Tanner and Wong, 1987; Gelfand and Smith, 1990) which is used by WinBUGS (Lunn *et al.*, 2000).

Given its iterative form, the convergence of the simulations is crucial for validating the results obtained using these complete Bayesian methods. This is the reason why in this study, for each parameter to be estimated, three chains are used and a large number of simulations are carried out (50,000 minimum), of which 10% of the initial ones are ignored (burn-in procedure) and of which each certain number of iterations the estimated parameters are saved, rejecting the

remaining iterations with the aim of avoiding the autocorrelation that occurs in chains (thinning procedure).

After obtaining the results of the model, the convergence is verified using the Brooks-Gelman-Rubin (Brooks and Gelman, 1998) statistic (called R-hat) and the sampling size of the chains (n.eff) (Carlin *et al.*, 1998; Gelman *et al.*, 2004). The agreed convergence criteria were that the R-hat was to be less than 1.05 and the effective number (n.eff) higher than 100 in the estimations of the typical deviations of the random effects (Barceló *et al.*, 2008), given that they are the most problematic parameters due to the convergence. If these criteria were not met the number of simulations was increased until the convergence was reached.

The free software R (version 2.4.1) (R Development Core Team, 2007) and WinBUGS (Lunn *et al.*, 2000; Spiegelhalter *et al.*, 2003), and the R, R2WinBUGS (Gelman *et al.*, 2007) libraries were also used, enabling the two environments to be connected.

3. Results

During the first quarter of 2003, 753 individuals were recorded for having been admitted at least once due to COPD, of which 72.6% were men and the rest women.

As shown in table 2, the majority of the patients, 83.7%, were admitted only once and the rest between two and five times. The number of admissions presents statistically significant differences between both genders (p-value: 0.008), observing that the group of men show a higher number of readmissions than that of women. With regard to the problems that the individuals suffer, the most common is COPD, with 14.2% of the cases, followed by arterial hypertension with 10.9%. By gender, the men follow the global trend given that their most common problem is also COPD (15.5%), however, in women the most common is arterial hypertension (12.6%). It must also be highlighted that the problem of tobacco is much more common in men (7.9%) than in women (1.0%).

On average men were admitted 1.24 times (confidence interval at 95%, CI 1.19-1.29) and women 1.13 (CI 1.08-1.19) (table 3). The overall average age is 71.04 years (CI 69.90-72.19) and the average number of visits made during this period is 0.4 (CI 0.28-0.51). By gender, the average age of women is 70.9 years (CI 68.56-73.25) and of men 71.1 (CI 69.79-72.41), and with regard to the number of visits, the average for women is 0.48 (CI 0.24-0.72) and for men 0.37 (CI 0.24-0.49) (table 3). These differences between gender are also statistically significant, observing that the average age of men is higher than that of women (p-value: 0.000) and that women make more visits to the primary healthcare services than men (p-value: 0.001).

Table 4 shows the results obtained in the final model where only the variables associated to

time of permanence within each interval resulted in being significant. This variable has been treated in four parts, obtaining a non linear relationship. In this sense, if the time of permanence is within the first, the third or the fourth part, this variable positively influences the response variable, meaning more probability of readmission. On the other hand if the time of permanence is within the second part, this variable influences negatively. Despite not being statistically significant, a positive relationship has been detected in the event of suffering one of the following problems: cardiac insufficiency, dyslipemia or being an ex-smoker. In other words, if the individual suffers any of these problems there is more probability of being readmitted to hospital than if they do not suffer them. On the other hand, a negative relationship with the probability of being admitted has been discovered for the rest of the variables. In other words, women are less likely to be readmitted than the men, the older the individual the less the probability of being readmitted, if the individual suffers any of the following problems: alcoholism, chronic obstructive pulmonary disease (COPD), arterial hypertension, diabetes mellitus 2 or smoking they are less likely to be readmitted to hospital than those that do not suffer, and finally, the more visits made to primary, the lower the probability of the individual being readmitted.

With the aim of finding the probabilities of change of status for each unit of time the following formula was used:

$$S(t) = \prod_{j=1}^{J_t} (1 - \lambda_{hik}(t_j)) \quad (5)$$

where t_j indicates the duration of the interval j , J_t is the j value corresponding to the duration of the interval j and λ_{hik} is the estimated risk function with either explicative variables or without taking them into account.

Figure 1 shows the probability of a change in status occurring, as in, of being admitted or not being admitted or the reverse, according to the number of admission. First of all it must be noted that for the first three admissions this probability exists for the whole study period (90 days) whilst for the fourth and fifth admission there is only data available for up to the middle of the period. It can also be observed that during the first thirty days, the probability of the change in status of the first admission is higher than the rest, and that the fourth admission has the lowest probability. During the rest of the period, the probability of the change of status of the first admission and of the third, are very close and that of the second admission is the lowest.

Figure 2 shows the probability of a change in status occurring according to gender for the first admission and it can be observed how women have a higher probability than men during the whole period of study. The other graphs (figures 3-7) show how the probability of a change of status occurring for the first admission according to the problems that the individual could suffer (alcohol, COPD, arterial hypertension, cardiac insufficiency and tobacco) follows the same trend

in all. If the individual suffers from the problem they are more likely to show a change in status compared to not suffering (although no relationship was statistically significant).

4. Discussion

For this study a method for modelling multivariate survival data has been proposed. Specifically, and based on the proposals of Goldstein, Steele and others (Goldstein, 1991; Steele *et al.*, 1996; Goldstein *et al.*, 2002; Steele, 2008; Gelman *et al.*, 2004), a mixed generalised linear model has been used with a binomial response with the following characteristics: in the first place, the individual heterogeneity has been modelled using random effects in the constant on the level of the individual; in second place, the time of permanence in the current status has been modelled using a smoothing function that included restricted cubic splines and finally, the model has been estimated using a Bayesian approximation.

It must also be noted that the model excludes the random effects associated with the episode given that significant differences are not shown in any of them, which could be a result of a low proportion of women that experience more than one change of status.

As an application the risk has been used that each individual has of being readmitted to hospital due to COPD, and the model aims to explain the probability of a change in status occurring, in other words, of changing from being admitted to not being admitted or the reverse. The database used has been structured on three levels, the individual, the episode or recurrence and the status of the patient.

Other published studies that try to identify the risk factors associated to hospital readmission due to COPD, have related pulmonary hypertension (*cor pulmonale*) (Kessler *et al.*, 1999; González *et al.*, 2004), a high pressure-time index (PTI) (González *et al.*, 2004), the number of previous admissions and a low forced expiratory volume in one second (FEV₁) (Garcia-Aymerich *et al.*, 2001; Garcia-Aymerich *et al.*, 2003), with a higher risk of hospitalisation.

In this sense, it has also been observed that various factors related to medical attention and prescription, such as being controlled by a doctor and taking anticholinergics or oral corticosteroids, increase the risk of readmission, contrary to what would be expected (Garcia-Aymerich *et al.*, 2001; Garcia-Aymerich *et al.*, 2003).

Burge *et al.* (2000), showed that inhaled corticoids reduced the number of hospitalisations and in the same sense, Paggiaro *et al.* (1998) found that the seriousness of the disease was also reduced, as opposed to Calverly *et al.* (2003) who found no differences between a group of patients treated with inhaled corticoids and a placebo group.

It has also been found that a factor that reduced the risk of readmission to hospital of sufferers of COPD is the practice of physical exercise by the patient (Osman *et al.*, 1997; Garcia-Aymerich *et al.*, 2001; Garcia-Aymerich *et al.*, 2003).

With regard to personal variables of the patient, such as age, gender or chronic problems such as hypertension, diabetes mellitus, etc., it has not been found in any of these studies that they influence in the readmission to hospital of this type of patient.

Therefore, the results obtained in the present study do not differ from those published to date, given that no influence has been shown by these variables and additionally, the only variable that could be significantly associated with the probability of a change in status occurring, was the time of permanence within each status, and this variable has not been considered in any previous study.

Table 1. Example of the construction of the response variable y.

Individual	External time	Internal time	y	Status	Gender	COPD
1	1	1	0	0	F	0
1	2	2	0	0	F	0
1	3	3	0	0	F	0
1	4	4	0	0	F	0
1	5	5	0	0	F	0
1	6	6	1	0	F	0
1	7	1	0	1	F	0
1	8	2	0	1	F	0
1	9	3	1	1	F	0
1	10	1	0	0	F	0
2	1	1	0	0	M	1
2	2	2	1	0	M	1
2	3	1	0	1	M	1
2	4	2	0	1	M	1
2	5	3	1	1	M	1
2	6	1	0	0	M	1
2	7	2	1	0	M	1
2	8	1	0	1	M	1
2	9	2	0	1	M	1
2	10	3	1	1	M	1

External time: calendar time.

Internal time: time within each status.

Y: response variable.

status $\begin{cases} 0 & \text{not admitted} \\ 1 & \text{admitted} \end{cases}$

gender $\begin{cases} F & \text{female} \\ M & \text{male} \end{cases}$

status $\begin{cases} 0 & \text{COPD no} \\ 1 & \text{COPD yes} \end{cases}$

Table 2. Distribution by gender of the variables.

	Men n(%)	Women n(%)
No. readmissions		
1	447(81.7)	183(88.8)
2	78(14.3)	20(9.7)
3	16(2.9)	2(1.0)
4	4(0.7)	1(0.5)
5	2(0.4)	0(0.0)
Alcohol consumption		
Yes	19(3.5)	0(0.0)
COPD		
Yes	85(15.5)	22(10.7)
Cardiac insufficiency		
Yes	12(2.2)	2(1.0)
Arterial hypertension		
Yes	56(10.2)	26(12.6)
Dyslipemia		
Yes	36(6.6)	14(6.8)
Diabetes mellitus 2		
Yes	38(6.9)	10(4.9)
Other diabetes		
Yes	2(0.4)	0(0.0)
Smoker		
Yes	43(7.9)	2(1.0)
Ex-smoker		
Yes	12(2.2)	2(1.0)

Table 3. Descriptives of the numerical variables

	Gender		<i>Total</i> Average (sd); median; [95% CI]
	Male Average (sd); median; [95% CI]	Female Average (sd); median; [95% CI]	
Age	71.10(15.61);75.00; [69.79-72.41]	70.90(17.08);75.50; [68.56-73.25]	71.04(16.01);75.00; [69.90-72.19]
No. of visits	0.37(1.52);0.00; [0.24-0.49]	0.48(1.75);0.00; [0.24-0.72]	0.40(1.58);0.00; [0.28-0.51]
No. of admissions	1.24(0.53);1.00; [1.19-1.29]	1.13(0.40);1.00; [1.08-1.19]	1.21(0.53);1.00; [1.17-1.25]

Table 4. Final model without spatial effects

	Odds Ratio (sd)	[95% Confidence Interval]
Gender (male)		
Female	0.897 (0.061)	[0.806- 1.015]
Time of permanence in each interval		
First break	2.002 (2.043)	[0.094- 40.987]
Second break	0.161 (0.193)	[0.109- 0.229]
Third break	3.850(0.325)	[1.914- 6.568]
Fourth break	5.562 (0.167)	[3.908- 7.706]
Age (< 64 years)		
64-72 years	0.951 (0.096)	[0.780- 1.119]
72-77 years	0.908 (0.098)	[0.738- 1.083]
77-82 years	0.886 (0.095)	[0.730- 1.048]
>82 years	0.916 (0.097)	[0.749- 1.083]
Alcohol (no)		
Yes	0.975 (0.186)	[0.669- 1.364]
Chronic Obstructive Pulmonary Disease, COPD (no)		
Yes	0.999 (0.815)	[0.838- 1.166]
Cardiac insufficiency (no)		
Yes	1.068 (0.199)	[0.717- 1.586]
Arterial hypertension (no)		
Yes	0.989 (0.102)	[0.821- 1.217]
Dyslipemia (no)		
Yes	1.011 (0.121)	[0.794- 1.257]
Diabetes Mellitus 2 (no)		
Yes	0.968 (0.115)	[0.772- 1.206]
Smoker (no)		
Yes	0.838 (0.124)	[0.658- 1.073]
Ex-smoker (no)		
Yes	1.033 (0.195)	[0.724- 1.482]
Number of visits	0.492 (2.041)	[0.025- 10.581]
DIC		16125.300
AIC		16143.136
BIC		16193.446

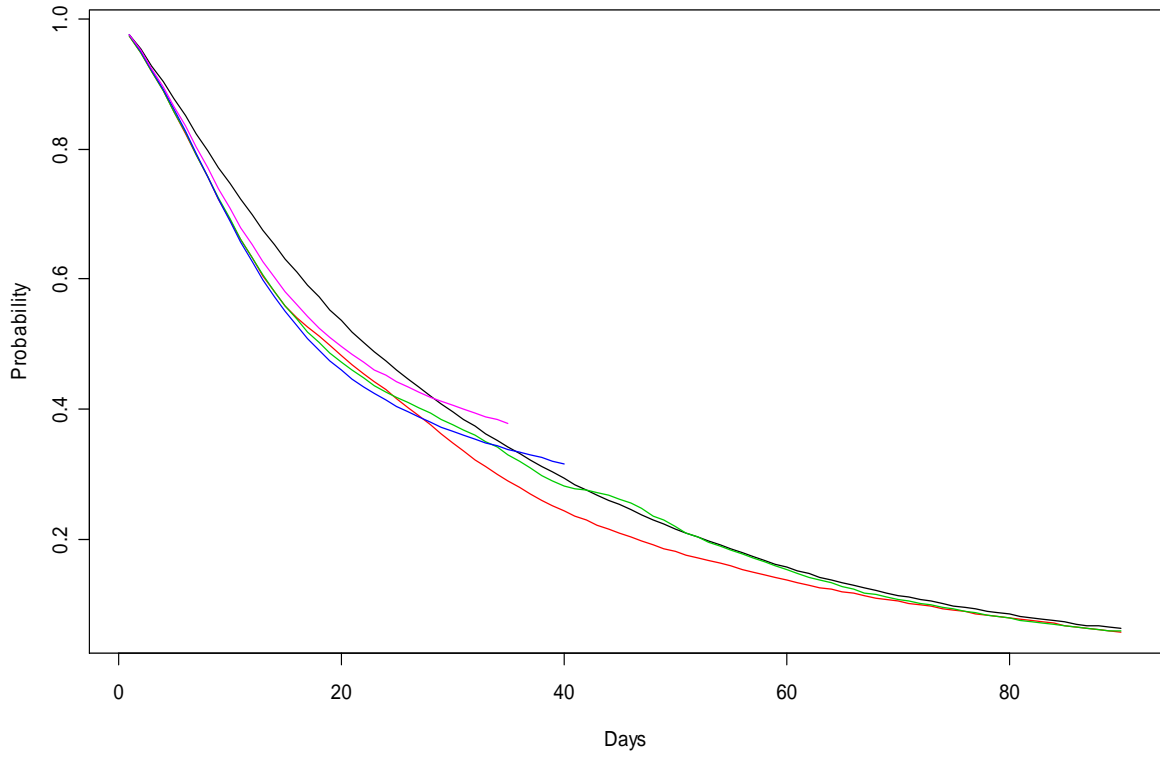
Reference values in brackets. Significant variables in bold.

DIC: deviance information criteria.

AIC: akaike information criteria.

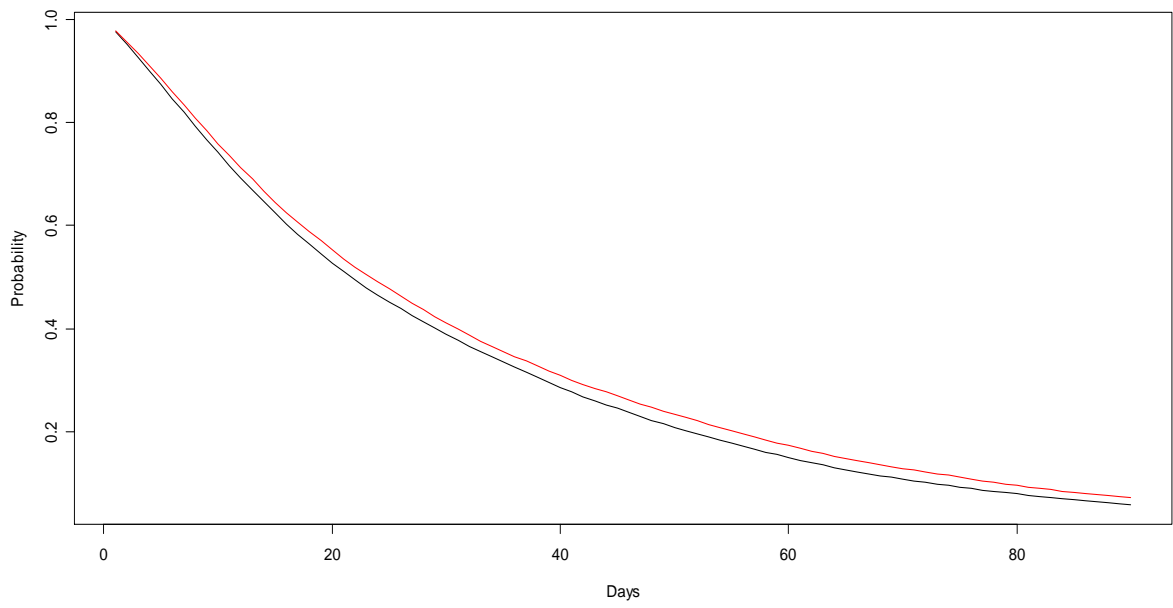
BIC: Bayesian information criteria.

Figure 1. Probability of each admission.



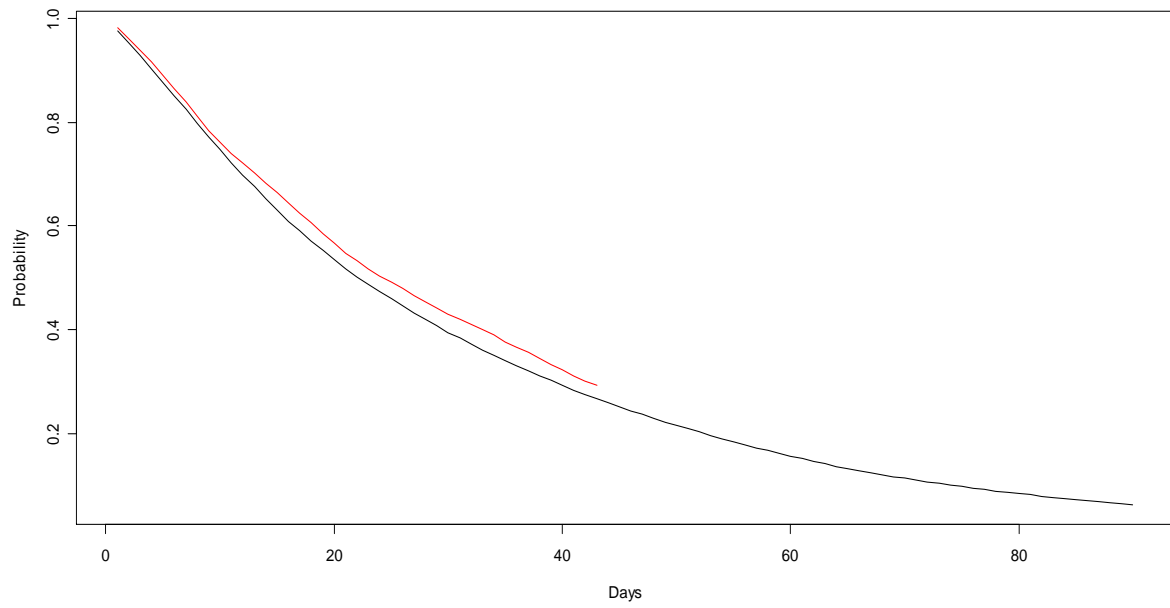
Black: first admission
Red: second admission
Green: third admission
Blue: fourth admission
Pink: fifth admission

Figure 2. Probability of the first admission by gender.



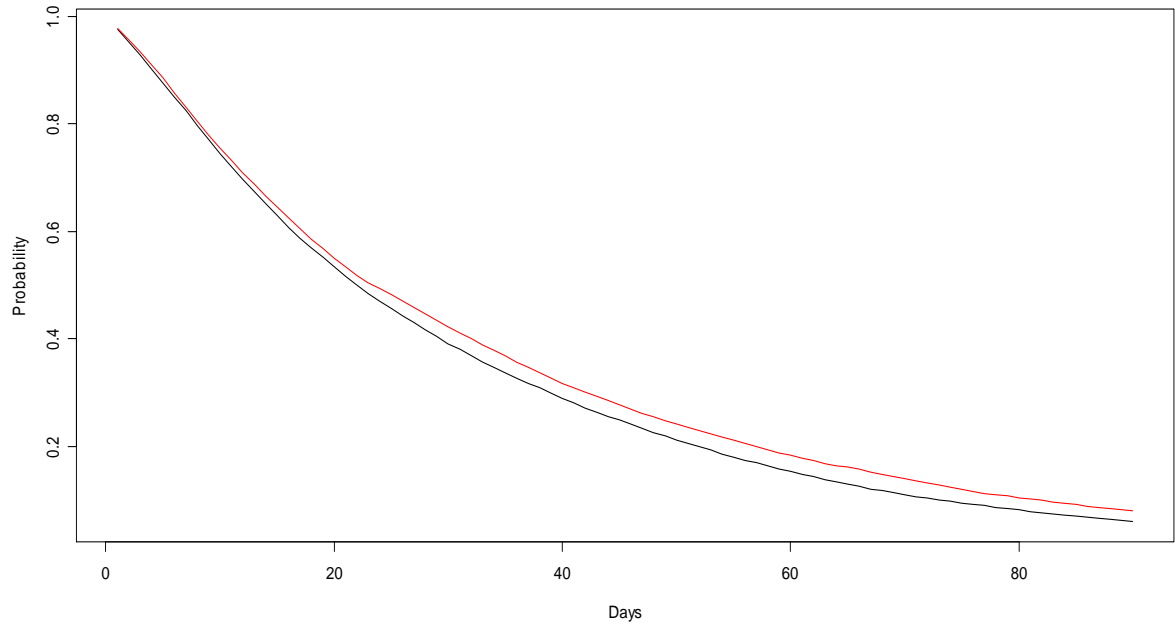
Black: men
Red: women

Figure 3. Probability of first admission according to if suffering an alcohol problem.



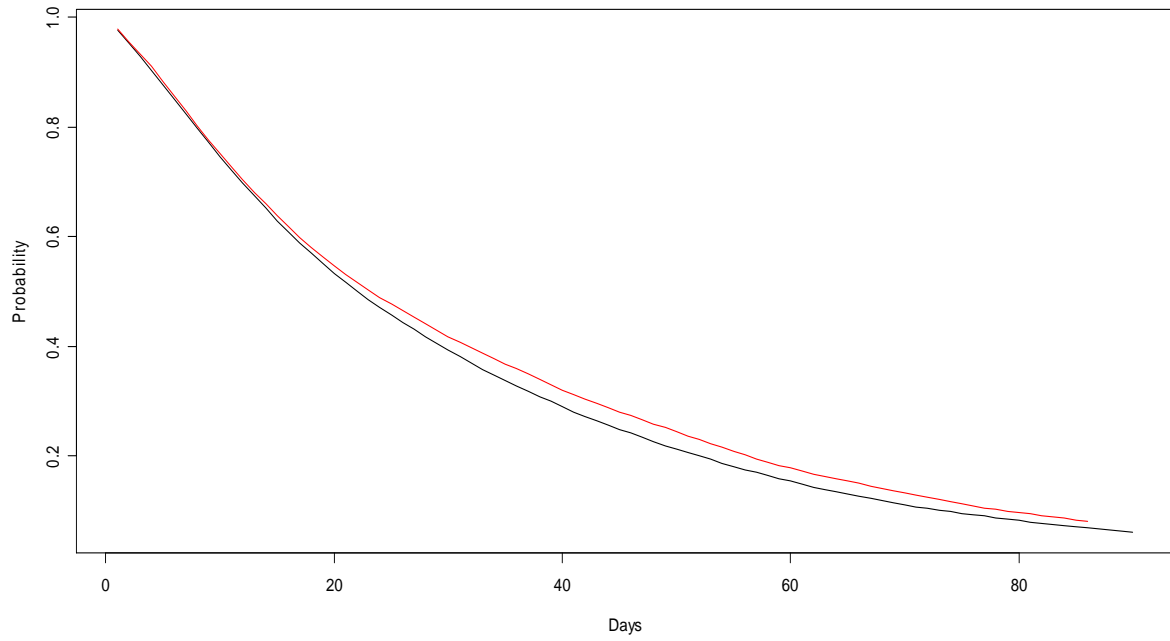
Black: no
Red: yes

Figure 4. Probability of first admission according to if suffering a COPD problem.



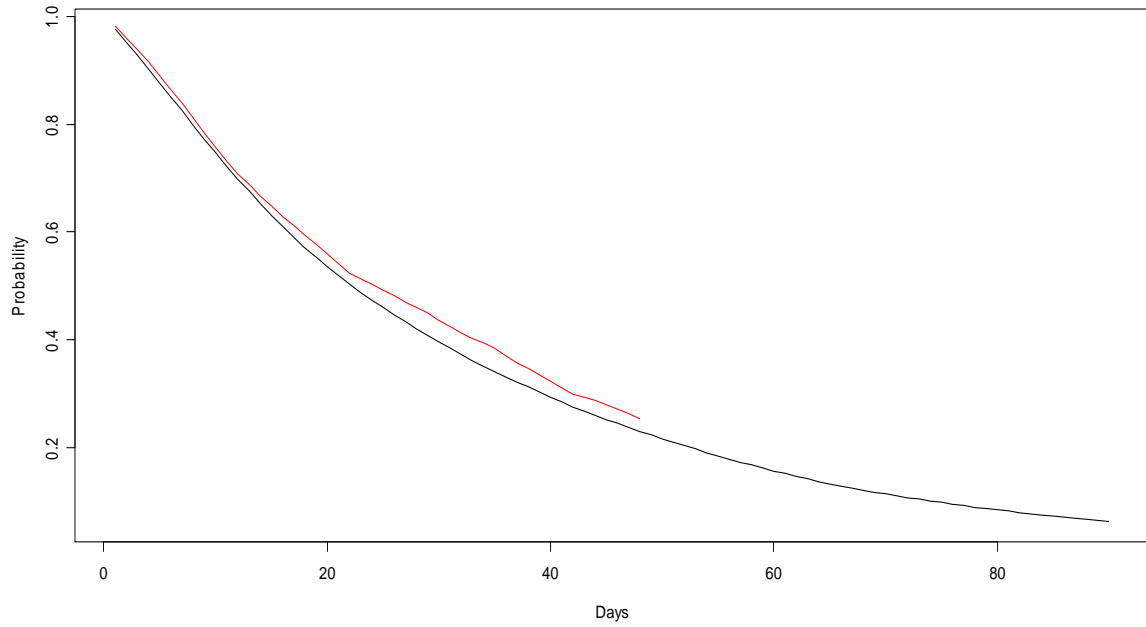
Black: no
Red: yes

Figure 5. Probability of first admission according to if suffering an arterial hypertension problem.



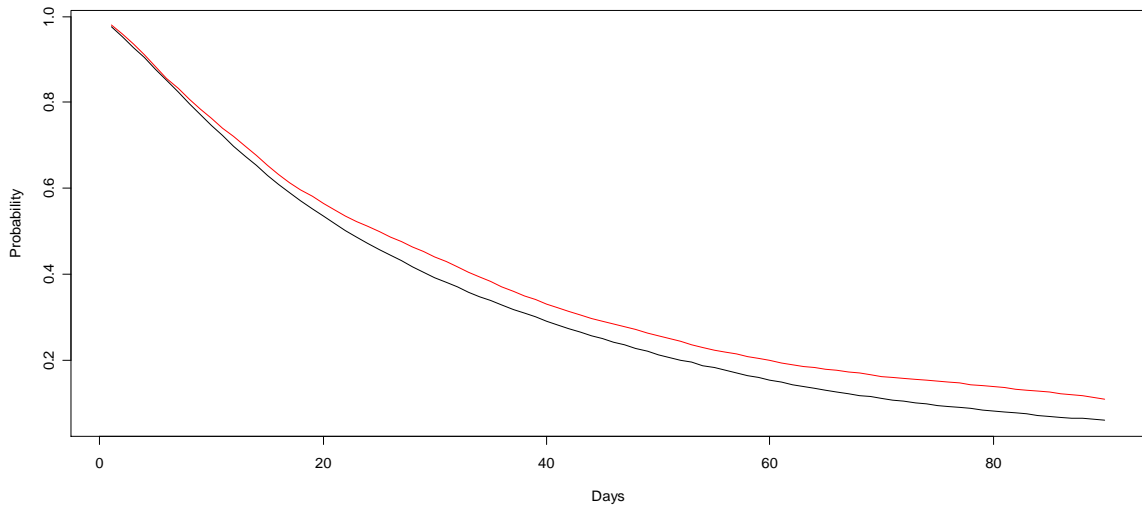
Black: no
Red: yes

Figure 6. Probability of first admission according to if suffering a cardiac insufficiency problem.



Black: no
Red: yes

Figure 7. Probability of first admission according to if suffering a tobacco problem.



Black: no
Red: yes

Bibliography

Barceló MA. (2001). Modelos de fragilidad en el análisis multivariante de supervivencia [tesis doctoral]. Girona: Universitat de Girona.

Barceló MA., Sáez M. (2004). A modifications of the EM algorithm to estimate an Andersen-Gill gamma frailty model for multivariate failure time data. *Journal of Modern Applied Statistical Methods*, 3, 504-17.

Barceló MA., Saez M., Cano-Serral G., Martínez-Beneito MA., Martínez JM., Borrell C., Ocaña-Riola R., Montoya I., Calvo M., López-Abente G., Rodríguez-Sanz M., Toro S., Alcalá JT., Saurina C., Sánchez-Villegas P., Figueiras A. (2008). Métodos para la suavización de indicadores de mortalidad: Aplicación al análisis de desigualdades en mortalidad en ciudades del Estado Español (Proyecto MEDEA). *Gaceta Sanitaria*, 22, 596-608.

Brooks S., Gelman A. (1998). Some issues in monitoring convergence of iterative simulations. In *Proceedings of the Section on Statistical Computing, ASA*.
<http://citeseer.ist.psu.edu/brooks98some.html>.

Burge PS., Calverley PM., Jones PW., Spencer S., Anderson JA., Maslen TK. (2000). Randomised, double blind, placebo controlled study of fluticasone propionate in patients with moderate to severe chronic obstructive pulmonary disease: the ISOLDE trial. *British Medical Journal*, 320, 1297-303.

Calverly P., Pauwles R., Vestbo J., Jones P., Pride N., Gulsvik A., Anderson J., Maden C. (2003). Combined salmeterol and fluticasone in the treatment of chronic obstructive pulmonary disease: a randomised controlled trial. *Lancet*, 361, 449-56.

Carlin BP., Gelman A., Neal RM. (1998). Markov Chain Monte Carlo in Practice: A Roundtable Discussion. *The American Statistician*, 52, 93-100.

Casas A., Troosters T., Garcia-Aymerich J., Roca J., Hernández C., Alonso A., del Pozo F., de Toledo P., Antó JM., Rodríguez-Roisín R., Decramer M. (2006). Integrated care prevents hospitalisations for exacerbations in COPD patients. *European Respiratory Journal*, 28, 123-30.

Cerveri I., Accordini S., Verlato G., Corsico A., Zaoia MC., Casali L., Burney P., de Marco R. (2001). Variations in the prevalence across countries of chronic bronchitis and smoking habits in young adults. *European Respiratory Journal*, 18, 85-92.

Durrleman S., Simon R. (1989). Flexible regression models with cubic splines. *Statistics in Medicine*, 8, 551-61.

Elliot P., Best N. (2005). Geographical patterns of disease. In Armitage P, Colton T. Encyclopedia of Biostatistics, 2nd edition. Chichester, West Sussex, United Kingdom: John Wiley & Sons. Also available at: <http://eu.wiley.com/legacy/wileychi/eob/articles.html> [accessed on 5 February 2009].

Gadorury MA., Schwartzman K., Rouleau M., Maltais R., Julien M., Beupré A., Renzi P., Bégin R., Nault D., Bourbeau J. (2005). Self-management reduces both short- and long- term hospitalisation in COPD. *European Respiratory Journal*, 26, 853-57.

Garcia-Aymerich J., Monsó E., Marrades RM., Escarrabill J., Félez MA., Sunyer J., Antó JM. (2001). Risk factors for hospitalization for a chronic obstructive pulmonary disease exacerbation. *American Journal of Respiratory Crittical Care*, 164, 1002-7.

Garcia-Aymerich J., Farrero E., Félez MA., Izquierdo J., Marrades RM., Antó JM. (2003). Risk factors of readmission to hospital for a COPD exacerbation: a prospective study. *Thorax*, 58, 100-5.

Gelfand AE., Smith AFM. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398-409.

Gelman A., Carlin JB., Stern HS., Rubin DB. (2004). *Bayesian data analysis*. Londres:Chapman & Hall/CRC.

Gelman A. (2005). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1, 1-19.

Gelman A., Sturtz S., Ligges U (2008). The R2WinBUGS Package. Available at: <http://www.stat.columbia.edu/~gelman/bugsR/> [accessed on 5 February 2009].

Geman S., Geman D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12, 609-28.

Gilks WR., Richardson S., Spiegelhalter DJ. (1996). *Markov Chain Monte Carlo in Practice*. Nueva York: Chapman & Hall.

Goldstein H.(1991). Nonlinear multilevel Models, with an application to discrete response data. *Biometrika*, 78, 45-51.

Goldstein H., Browne W., Rasbash J. (2002). Multilevel modelling of medical data. *Statistics in Medicine*, 21, 3291-315.

Goldstein H., Pan H., Bynner J. (2004a). A flexible procedure for analysing longitudinal event histories using a multilevel model. *Understanding Statistics*, 3, 85-99.

Goldstein H., Steele F., Browne W. (2004b). A general multilevel multistate competing risks model for event history data, with an application to a study of contraceptive use dynamics. *Statistical Modelling*, 4, 145-59.

González C., Servera E., Ferris G., Blasco ML., Marín J. (2004). Factores predictivos de reingreso hospitalario en la agudización de la EPOC moderada-grave. *Archivos de Bronconeumología*, 40, 502-7.

Halbert RJ., Natoli JL., Gano A., Badamgarav E., Buist AS., Mannino DM. (2006). Global burden of COPD: systematic review and metaanalysis. *European Respiratory Journal*, 28, 523-32.

Harrell FE., Lee KL., Pollock BG. (1988). Regression models in clinical studies: determining relationships between predictors and response. *Journal of National Cancer Institutes*, 80, 1198-202.

Hermiz O., Comino E., Marks G., Daffurn K., Wilson S., Harris M. (2002). Randomised controlled trial of home based care of patients with chronic obstructive pulmonary disease. *British Medical Journal*, 325, 938.

Instituto Nacional de Estadística (INE) (1998). Natural evolution of the population. Mortality by specific causes [in Spanish]. Madrid: INE. Also available at: <http://www.ine.es> [accessed on 5 February 2009].

Izquierdo JL. (2003). The burden of COPD in Spain: results from the Confronting Survey. *Respiratory Medicine*, 97, S61-9.

Kaag M., Wijkel D., De Jong D. (1996). Primary health care replacing hospital care- the effect on quality of care. *International Journal for quality in health care*, 8, 367-73.

Kessler R., Faller M., Forgaut G., Menecier B., Weitzenblum E. (1999). Predictive factors of hospitalization for acute exacerbation in a series of 64 patients with chronic obstructive pulmonary disease. *American Journal of Respiratory Critical Care*, 159, 158-64.

Librero J., Peiró S., Ordiñana R. (1999). Chronic Comorbidity and Outcomes of Hospital Care: Length of stay, mortality, and readmission at 30 and 365 days. *Journal of Clinical Epidemiology*, 52, 171-79.

Lopez AD., Shibuya K., Rao C., Mathers CD., Hansell AL., Held LS., Schmid V., Buist S.(2006). Chronic obstructive pulmonary disease: current burden and future projections. *European Respiratory Journal*, 27, 397-412.

Lunn DJ., Thomas A., Best N., Spiegelhalter D. (2000) WinBUGS -- a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10:325--337.

Masa JF., Sobradillo V., Villasante C., Jiménez-Ruiz CA., Fernández-Fau L., Viejo JL., Miravittles M. (2004). Costs of chronic obstructive pulmonary disease in Spain. Estimation from a population-based study [in Spanish]. *Archivos de Bronconeumología*, 40, 72-9.

Miravittles M., Murio C., Guerrero T., Gisbert R.(2003). Costs of chronic bronchitis and COPD. A one year follow-up study. *Chest*, 123, 784-91.

Miravittles M. (2005). Enfermedad pulmonar obstructiva crónica. *Med Clin (Barc)*, 125, 65-74.

Murray CJL., Lopez AD., editors. (1996). *The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020*. Cambridge, MA: Harvard University Press.

Murray CJ., Lopez AD. (1997). Alternative projections of mortality and disability by cause 1990-2020: Global Burden of Disease Study. *Lancet*, 349, 1498-1504.

Osman IM., Godden DJ., Friend JA., Legge JS., Douglas G. (1997). Quality of life and hospital re-admission in patients with chronic obstructive pulmonary disease. *Thorax*, 52, 67-71.

Paggiaro PL., Dahle R., Bakran I., Frith L., Hollingworth K., Efthimiou J. (1998). Multicentre randomised placebo-controlled trial of inhaled fluticasone propionate in patients with chronic obstructive pulmonary disease. International COPD Study Group. *Lancet*, 351, 773-80.

Pauwels RA., Buist AS., Calverley PMA., Jenkins CR., Hurd SS. (2001). Global strategy for the diagnosis, management and prevention fo chronic osbtructive pumonary disease: NHLBI/WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop summary. *American Journal of Respiratory and Critical Care*, 163, 1256-76.

Peris Cardells R., Rico Salvador I., Herrera de Pablo P., Pérez Jiménez A., Sánchez-Toril López F., Pérez Fernández JA.(2004). Demographic study of tobacco addiction in Health District 5, Valencia, Spain [in Spanish]. *Archivos de Bronconeumología*, 40, 160-5.

Prentice RL., Williams BJ., Peterson AL. (1981). On the regression analysis of multivariate failure time data. *Biometrika*, 68, 373-9.

R Development Core Team (2007). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2007, ISBN 3-900051-07-0. Available at: <http://www.R-project.org> [accessed on 5 February 2009].

Remontet L., Bossard N., Belot A., Esteve J. (2007). An overall strategy based on regression models to estimate relative survival and model the effects of pronostic factors in cancer survival studies. *Statistics in Medicine*, 26, 2214-28.

Saez M., Saurina C. (2007). *Estadística y Epidemiología Espacial*. Girona: Edicions a Petició.

Sobradillo V., Miravittles M., Jiménez CA., Gabriel R., Viejo JL., Masa JF., Fernández-Fau L., Villasante C. (1999). Epidemiological study of chronic obstructive pulmonary disease in Spain (IBERPOC): prevalence of chronic respiratory symptoms and airflow limitation [in Spanish]. *Archivos de Bronconeumología*, 35, 159-166.

Sobradillo V., Miravittles M., Gabriel R., Jiménez-Ruiz CA., Villasante C., Fernando JF., Viejo JL., Fernández-Pau L. (2000). Geographical variations in prevalence and underdiagnosis of COPD: results of the IBERPOC multicentre epidemiological study. *Chest*, 118, 981-989.

Soler JJ., Sánchez L., Latorre M., Alamar J., Román P., Perpiñá M. (2001). The impact of COPD on hospital resources: the specific burden of COPD patients with high rates of hospitalization [in Spanish]. *Archivos de Bronconeumología*, 37, 375-81.

Spiegelhalter D., Thomas A., Best N., Lunn D. (2003). *WinBUGS User Manual*. Version 1.4 Also available at: <http://www.mrc-bsu.cam.ac.uk/bugs> [accessed on 5 February 2009].

Steele F., Diamond I., Amin S. (1996). Immunization uptake in rural blangladesh: a multilevel analysis. *Journal of the Royal Statistical Society*, 159, 289-99.

Steele F.(2008). Multilevel models for longitudinal data. *Journal of the Royal Statistical Society*, 171, 5-19.

Tanner MA., Wong WH. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82, 528-49.

Therneau TM., Hamilton SA.(1997). rhDNase as an example of recurrent event analysis. *Statistics in Medicine*, 16, 2029-47.

Therneau TM., Grambsch PM. (2000). Modeling survival data. Extending the Cox model. *Statistics for biology and health*. New York: Springer.

Verdaguer A., Peiró S., Librero J. (2003). Variations in the use of hospital resources in treating patients with chronic obstructive pulmonary disease [in Spanish]. *Archivos de Bronconeumología*, 39, 442-8.

Wei LJ., Lin DY., Weissfeld L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association*, 84, 1065-73.

Wilkinson TMA., Donaldson GC., Hurst JR., Seemungal TAR., Wedzicha JA.(2004). Early therapy improves outcomes of exacerbations of chronic obstructive pulmonary disease. *American Journal of Respiratory and Critical Care*, 169, 1298-303.

ANNEX VI: ANÀLISI DELS INDIVIDUS SEGONS EL NOMBRE D'INGRESSOS QUE HAN GENERAT.

A.VI.1. Anàlisi dels individus que han generat un sol ingrés.

Tal com s'observa a la taula A.VI.1, durant l'any 2003 han ingressat una sola vegada degut a la MPOC un total de 1679 individus. D'aquests, el 69.1% són homes i un 30.9% dones.

Taula A.VI.1. Distribució dels individus segons el sexe.

	Nº d'individus	% d'individus
Homes	1160	69.1
Dones	519	30.9
Total	1679	100.0

Analitzant l'edat d'aquests individus s'obté que tenen una mitjana de 67.58 anys i una mediana de 73 anys, per tant, igual que en els descriptius globals, aquests ingressos han estat realitzats per individus amb edat avançada i la distribució corresponent és asimètrica a l'esquerra. La resta d'estadístics calculats prenen valors d'acord amb aquesta idea, com ara els quartils primer i tercer que són 65 i 80 anys respectivament, o la moda que és de 78 anys.

Taula A.VI.2. Descriptius de l'edat del pacient (anys).

Mitjana	67.58	Moda	78.00
Mediana	73.00	Mínim	1.00
Mitjana retallada el 5%	69.69	Màxim	102.00
Desviació típica	20.32	Quartils	
		Q1	63.00
		Q3	80.00

Per sexes s'observa la mateixa característica esmentada (taula A.VI.3). Potser cal destacar que les dones tenen una edat mitjana inferior a la dels homes, 65.22 i 68.64 respectivament, i en canvi, les dones tenen una mediana lleugerament superior. A més, la distribució de les dones és més dispersa.

Taula A.VI.3. Descriptius de l'edat del pacient segons el sexe (anys).

	SEXE	
	Home	Dona
Mitjana	68.64	65.22
Mediana	73.00	74.00
Mitjana retallada el 5%	70.76	67.15
Desviació típica	18.76	23.30
Moda	78.00	78.00
Mínim	1.00	1.00
Màxim	102.00	96.00
Quartils		
Q1	64.00	58.00
Q3	80.00	81.00

Taula A.VI.4. Distribució dels individus segons sexe i problemes de salut que pateixen.

	No presenten el problema			Presenten el problema		
	Homes	Dones	Total	Homes	Dones	Total
Consum d'alcohol	1117 (96.3%)	518 (99.8%)	1635 (97.4%)	43 (3.7%)	1 (0.2%)	44 (2.6%)
MPOC	971 (83.7%)	459 (88.4%)	1430 (85.2%)	189 (16.3%)	60 (11.6%)	249 (14.8%)
Insuficiència cardíaca	1139 (98.2%)	509 (98.1%)	1648 (98.2%)	21 (1.8%)	10 (1.9%)	31 (1.8%)
Hipertensió arterial	1036 (89.3%)	446 (85.9%)	1482 (88.3%)	124 (10.7%)	73 (14.1%)	197 (11.7%)
Dislipèmia	1077 (92.8%)	475 (91.5%)	1552 (92.4%)	83 (7.2%)	44 (8.5%)	127 (7.6%)
Diabetis mellitus 2	1086 (93.6%)	487 (93.8%)	1573 (93.7%)	74 (6.4%)	32 (6.2%)	106 (6.3%)
Fumador	1060 (91.4%)	497 (95.8%)	1557 (92.7%)	100 (8.6%)	22 (4.2%)	122 (7.3%)
Ex-fumador	1142 (98.4%)	514 (99.0%)	1656 (98.6%)	18 (1.6%)	5 (1.0%)	23 (1.4%)

Analitzant els problemes que pateixen aquests 1679 individus s'observa que el problema més comú és MPOC (14.8%) seguit de la hipertensió arterial (11.7%), tendència ja observada en els descriptius globals (taula A.VI.4). A més, tal com passava també en el cas global, els homes segueixen la mateixa classificació, és a dir, el problema més comú és MPOC (16.3%) seguit de la hipertensió arterial (10.7%), en canvi les dones tenen com a problema més comú la hipertensió arterial (14.1%) seguit de MPOC (11.6%).

El 79.1% del total d'ingressos que han generat aquests individus han estat deguts a una urgència (taula A.VI.5). Per sexes els percentatges no varien massa, només en el cas de les dones és lleugerament inferior (77.4%).

Taula A.VI.5. Distribució dels individus segons sexe i circumstància d'ingrés.

	SEXE		Total
	Home	Dona	
Admissió urgent	927 (79.9%)	401 (77.4%)	1328 (79.1%)
Admissió programada	233 (20.1%)	117 (22.6%)	350 (20.9%)
Total	1160 (69.1%)	518 (30.9%)	1678 (100.0%)

La circumstància d'alta més comú, amb un 89.7% dels casos és la destinació al propi domicili seguida de la defunció amb un 5.8% dels casos. Per sexes es manté aquest mateix ordre amb uns percentatges molt similars als globals esmentats (taula A.VI.6).

La següent variable analitzada és la procedència del pacient en el moment d'ingressar (taula A.VI.7). S'observa com la majoria dels ingressos (80.4%) provenien del propi domicili i el 15.8% han estat traslladats de la consulta externa del mateix hospital.

A la taula A.VI.8 es mostren els 10 GRD's més freqüents d'aquests 1679 individus. Tal com passava en l'anàlisi global, el GRD més freqüent és la MPOC ja que el 37.6% dels ingressos estan catalogats en aquest GRD. En els homes, el percentatge de casos que tenen aquest GRD augmenta fins a un 41.5% i en les dones disminueix fins a un 28.9%.

Taula A.VI.6. Distribució dels individus segons sexe i circumstància d'alta.

	SEXE		Total
	Home	Dona	
Destinació al propi domicili	1031 (88.9%)	474 (91.5%)	1505 (89.7%)
Trasllat a un hospital general d'aguts	25 (2.2%)	8 (1.5%)	33 (2.0%)
Trasllat a un centre de mitjana o llarga durada	22 (1.9%)	4 (0.8%)	26 (1.5%)
Trasllat a una residència social	9 (0.8%)	6 (1.2%)	15 (0.9%)
Defunció	71 (6.1%)	36 (5.0%)	97 (5.8%)
Hospitalització domiciliària	2 (0.2%)	0 (0.0%)	2 (0.1%)
Total	1160 (69.1%)	518 (30.9%)	1678 (100.0%)

Taula A.VI.7. Distribució dels individus segons sexe i procedència de l'ingrés.

	SEXE		Total
	Home	Dona	
Domicili o residència social	932 (80.7%)	413 (79.7%)	1345 (80.4%)
Trasllat des d'un hospital general d'aguts	8 (0.7%)	7 (1.4%)	15 (0.9%)
Trasllat des d'un recurs de la xarxa sociosanitària	5 (0.4%)	5 (1.0%)	10 (0.6%)
Trasllat des d'un centre d'atenció primària	26 (2.3%)	12 (2.3%)	38 (2.3%)
Trasllat des de consulta externa del mateix hospital	184 (15.9%)	81 (15.6%)	265 (15.8%)
Total	1155 (69.0%)	518 (31.0%)	1673 (100.0%)

Taula A.VI.8. Distribució dels individus segons sexe i grup relacionat amb el diagnòstic.

	SEXE		Total
	Home	Dona	
088 MALALTIA PULMONAR OBSTRUCTIVA CRONICA	481 (41.5%)	150 (28.9%)	631 (37.6%)
089 PNEUMONIA SIMPLE/PLEURITIS >17a a/cc	174 (15.0%)	47 (9.1%)	221 (13.2%)
127 INSUFICIENCIA CARDIACA i XOC	72 (6.2%)	43 (8.3%)	115 (6.8%)
096 BRONQUITIS/ASMA >17a a/cc	38 (3.3%)	76 (14.6%)	114 (6.8%)
098 BRONQUITIS/ASMA 0-17a	47 (4.1%)	36 (6.9%)	83 (4.9%)
097 BRONQUITIS/ASMA >17a s/cc	17 (1.5%)	57 (11.0%)	74 (4.4%)
082 NEOPLASIES RESPIRATORIES	57 (4.9%)	4 (0.8%)	61 (3.6%)
087 EDEMA PULMONAR/INSUFICIENCIA RESPIRATORIA	19 (1.6%)	20 (3.9%)	39 (2.3%)
140 ANGOR PECTORIS	30 (2.6%)	5 (1.0%)	35 (2.1%)
122 TRAST CIRCULAT a/IAM s/COMPL CVASC, ALTA VIU	28 (2.4%)	3 (0.6%)	31 (1.8%)

A la taula A.VI.9 es mostra l'anàlisi de la variable pes relatiu associat a cada un dels GRD. Globalment la mitjana del pes relatiu és pràcticament la unitat (0.9749), la qual cosa indica que el cost mitjà global és lleugerament inferior a la mitjana del conjunt global de tots els GRD. Aquesta característica es manté per sexes tot i que per les dones és més baix.

Taula A.VI.9. Descriptius del pes relatiu del GRD globals i segons sexe.

	SEXE		Descriptius Globals
	Home	Dona	
Mitjana	0.9974	0.9245	0.9749
Mediana	0.9317	0.9317	0.9317
Mitjana retallada el 5%	0.9797	0.9025	0.9558
Desviació típica	0.2736	0.2865	0.2796
Moda	0.9317	0.9317	0.9317
Mínim	0.5141	0.0000	0.0000
Màxim	3.6936	3.6936	3.6936
Quartils			
Q1	0.9317	0.7871	0.8768
Q3	1.0647	1.0130	1.0647

Aquest apartat s'acaba amb uns descriptius de la variable estada. L'estada mitjana és de 6.70 dies i la mediana de 6.00. El valor més freqüent és 3 dies i el 50% de les estades es troben entre 3 i 9 dies.

Taula A.VI.10. Descriptius de l'estada globals i per sexe.

	SEXE		Descriptius Globals
	Home	Dona	
Mitjana	6.62	6.88	6.70
Mediana	5.50	6.00	6.00
Mitjana retallada el 5%	6.08	6.30	6.15
Desviació típica	5.20	5.31	5.23
Moda	3.00	5.00	3.00
Mínim	1.00	1.00	1.00
Màxim	64.00	41.00	64.00
Quartils			
Q1	3.00	3.00	3.00
Q3	9.00	9.00	9.00

A.VI.2.. Anàlisi dels individus amb dos ingressos..

Ja s'ha vist que es disposa de 416 individus (taula 19 dels descriptius globals, capítol 6), que han generat dos ingressos durant l'any 2003, per tant, 832 ingressos en total. D'aquests, el 77.9% són homes i el 22.1% dones (taula A.VI.11).

Taula A.VI.11. Distribució dels individus segons el sexe

	Nº d'individus	% d'individus
Homes	648	77.9
Dones	184	22.1
Total	832	100.0

La mitjana d'edat global d'aquests individus és de 71.38 anys. La meitat tenen una edat inferior a 75 anys i l'altra meitat superior (taula A.VI.12). Això indica que aquests ingressos han estat realitzats per individus d'edat avançada i com a conseqüència la distribució d'aquesta variable és asimètrica a l'esquerra. En la mateixa línia es té que el primer quartil es troba a 68 anys i el tercer a 81; i que l'edat més freqüent és de 78 anys.

Taula A.VI.12. Descriptius de l'edat del pacient (anys).

Mitjana	71.38	Moda	78.00
Mediana	75.00	Mínim	1.00
Mitjana retallada el 5%	73.37	Màxim	94.00
Desviació típica	15.73	Quartils	
		Q1	68.00
		Q3	81.00

Per sexes, s'observa la mateixa característica esmentada en el cas global. Comentar només que el valor d'alguns estadístics en les dones, com ara la mitjana, la mediana i la mitjana retallada és lleugerament inferior que el corresponent en els homes (Taula A.VI.13).

Taula A.VI.13. Descriptius de l'edat del pacient segons el sexe (anys).

	SEXE	
	Home	Dona
Mitjana	72.45	67.62
Mediana	75.00	74.00
Mitjana retallada el 5%	74.09	69.68
Desviació típica	13.99	20.34
Moda	78.00	81.00
Mínim	1.00	1.00
Màxim	94.00	92.00
Quartils		
Q1	69.00	59.50
Q3	80.00	81.00

La següent taula (taula A.VI.14) mostra el nombre d'individus que presenten cada un dels problemes de salut que poden influir en la malaltia segons el sexe. El problema més comú continua essent MPOC (15.6%) seguit de la hipertensió arterial (11.8%). Aquests percentatges en els homes es mantenen augmenten lleugerament (MPOC 17.3% i hipertensió arterial 13.0%), en canvi en les dones disminueixen (MPOC 9.8% i hipertensió arterial 7.6%).

Quan a la circumstància d'ingrés més comú continua essent l'admissió urgent en un 80.2% del total dels ingressos. Per sexes el percentatge ronda aquest mateix valor (taula A.VI.15).

A la taula A.VI.16 s'observa que la circumstància d'alta més freqüent és la destinació al propi domicili amb un 87.4% del total dels ingressos, seguida de la defunció amb un percentatge de 5.6%. Per sexes es segueix aquesta mateixa tendència amb uns percentatges similars.

Taula A.VI.14. Distribució dels individus segons sexe i problemes de salut que pateixen

	No presenten el problema			Presenten el problema		
	Homes	Dones	Total	Homes	Dones	Total
Consum d'alcohol	626 (96.6%)	184 (100.0%)	810 (97.4%)	22 (3.4%)	0 (0.0%)	22 (2.6%)
MPOC	536 (82.7%)	166 (90.2%)	702 (84.4%)	112 (17.3%)	18 (9.8%)	130 (15.6%)
Insuficiència cardíaca	636 (98.1%)	182 (98.9%)	818 (98.3%)	12 (1.9%)	2 (1.1%)	14 (1.7%)
Hipertensió arterial	564 (87.0%)	170 (92.4%)	734 (88.2%)	84 (13.0%)	14 (7.6%)	98 (11.8%)
Dislipèmia	604 (93.2%)	178 (96.7%)	782 (94.0%)	44 (6.8%)	6 (3.3%)	50 (6.0%)
Diabetis mellitus 2	592 (91.4%)	176 (95.7%)	768 (92.3%)	56 (8.6%)	8 (4.3%)	64 (7.7%)
Fumador	588 (90.7%)	180 (97.8%)	768 (92.3%)	60 (9.3%)	4 (2.2%)	64 (7.7%)
Ex-fumador	632 (97.5%)	182 (98.9%)	814 (97.8%)	16 (2.5%)	2 (1.1%)	18 (2.2%)

Taula A.VI.15. Distribució dels individus segons sexe i circumstància d'ingrés.

	SEXE		Total
	Home	Dona	
Admissió urgent	520 (80.2%)	147 (79.9%)	667 (80.2%)
Admissió programada	128 (19.8%)	37 (20.1%)	165 (19.8%)
Total	648 (77.9%)	184 (22.1%)	832 (100.0%)

Taula A.VI.16. Distribució dels individus segons sexe i circumstància d'alta.

	SEXE		Total
	Home	Dona	
Destinació al propi domicili	566 (87.3%)	161 (87.5%)	727 (87.4%)
	26	11	37
Trasllat a un hospital general d'aguts	(4.0%) 8	(6.0%) 3	(4.4%) 11
Trasllat a un centre de mitjana o llarga durada	(1.2%) 8	(1.6%) 1	(1.3%) 9
Trasllat a una residència social	(1.2%) 39	(0.5%) 8	(1,1%) 47
Defunció	(6.0%) 1	(4.3%) 0	(5.6%) 1
Evasió o alta administrativa	(0.2%)	(0.0%)	(0.1%)
Total	648 (77.9%)	184 (22.1%)	832 (100.0%)

El 80.7% dels ingressos provenen del propi domicili, seguits del 15% que provenen de consulta externa del mateix hospital (taula A.VI.17).

Taula A.VI.17. Distribució dels individus segons sexe i procedència de l'ingrés.

	SEXE		Total
	Home	Dona	
Domicili o residència social	535 (82.7%)	136 (73.9%)	671 (80.7%)
	9	5	14
Trasllat des d'un hospital general d'aguts	(1.4%) 0	(2.7%) 5	(1.7%) 5
Trasllat des d'un recurs de la xarxa socio sanitària	(0.0%) 11	(2.7%) 5	(0.6%) 16
Trasllat des d'un centre d'atenció primària	(1.7%) 92	(2.7%) 33	(1.9%) 125
Trasllat des de consulta externa del mateix hospital	(14.2%)	(17.9%)	(15.0%)
Total	647 (77.9%)	184 (22.1%)	831 (100.0%)

A la taula A.VI.18 es mostra els 10 GRD més freqüents en aquests 832 ingressos. El més comú continua essent la malaltia pulmonar obstructiva crònica amb un percentatge global del 45.7% dels casos. Aquest percentatge augmenta lleugerament en el cas dels homes (47.2%) i disminueix en les dones (40.2%).

Taula A.VI.18. Distribució dels individus segons sexe i grup relacionat amb el diagnòstic.

	SEXE		Total
	Home	Dona	
088 MALALTIA PULMONAR OBSTRUCTIVA CRONICA	306 (47.2%)	74 (40.2%)	380 (45.7%)
089 PNEUMONIA SIMPLE/PLEURITIS >17a a/cc	77 (11.9%)	11 (6.0%)	88 (10.6%)
127 INSUFICIENCIA CARDIACA i XOC	55 (8.5%)	16 (8.7%)	71 (8.5%)
082 NEOPLASIES RESPIRATORIES	32 (4.9%)	1 (0.5%)	33 (4.0%)
096 BRONQUITIS/ASMA >17a a/cc	7 (1.1%)	26 (14.1%)	33 (4.0%)
098 BRONQUITIS/ASMA 0-17a	12 (1.9%)	8 (4.3%)	20 (2.4%)
138 ARITMIES CARDIAQUES/TRASTORNS CONDUCCIÓ a/cc	16 (2.5%)	4 (2.2%)	20 (2.4%)
140 ANGOR PECTORIS	20 (3.1%)	0 (0.0%)	20 (2.4%)
079 INFECCIONS/INFLAMACIÓ NS RESPIRAT >17 a a/cc	17 (2.6%)	1 (0.5%)	18 (2.2%)
087 EDEMA PULMONAR/INSUFICIENCIA RESPIRATORIA	9 (1.4%)	9 (4.9%)	18 (2.2%)

A la següent taula (taula A.VI.19) es presenta un anàlisi de la variable pes relatiu dels GRD associats a cada ingrés. En aquest cas la mitjana del pes relatiu és lleugerament superior a la unitat (1.0119) la qual cosa indica que el cost mitjà global és superior a la mitjana del conjunt global de tots els GRD. Si s'observen els homes i les dones per separat aquesta característica es manté.

Taula A.VI.19. Descriptius del pes relatiu del GRD globals i segons sexe.

	SEXE		Descriptius Globals
	Home	Dona	
Mitjana	1.0122	1.0106	1.0119
Mediana	0.9317	0.9317	0.9317
Mitjana retallada el 5%	0.9913	0.9407	0.9805
Desviació típica	0.2754	0.4847	0.3329
Moda	0.9317	0.9317	0.9317
Mínim	0.5402	0.5873	0.5402
Màxim	3.6936	3.6936	3.6936
Quartils			
Q1	0.9317	0.8359	0.9317
Q3	1.0647	1.0130	1.0647

Per finalitzar l'anàlisi dels individus que han generat dos ingressos durant l'any 2003, només queda estudiar la durada de l'estada a l'hospital. En la taula A.VI.20 es pot observar com en aquest cas els estadístics prenen uns valors lleugerament superiors als calculats en el cas dels individus amb un sol ingrés (taula A.VI.4), excepte l'estada màxima que ara és de 45 dies i pels individus amb un sol ingrés era de 64.

Taula A.VI.20. Descriptius de l'estada globals i per sexe.

	SEXE		Descriptius Globals
	Home	Dona	
Mitjana	7.08	7.11	7.09
Mediana	6.00	6.00	6.00
Mitjana retallada el 5%	6.48	6.66	6.52
Desviació típica	5.65	4.77	5.47
Moda	6.00	5.00	5.00
Mínim	0.00	0.00	0.00
Màxim	45.00	32.00	45.00
Quartils			
Q1	3.00	4.00	3.00
Q3	9.00	9.00	9.00

A.VI.3. Anàlisi dels individus amb tres ingressos.

En l'arxiu d'estudi es tenen 125 individus que han generat tres ingressos; això vol dir, un total de 375 ingressos. D'aquests, el 79.2% són homes i el 20.8% dones (taula A.VI.21).

Taula A.VI.21. Distribució dels individus segons el sexe.

	Nº d'individus	% d'individus
Homes	297	79.2
Dones	78	20.8
Total	375	100.0

L'edat mitjana global d'aquest 125 individus és 70.90 anys, amb una desviació de 13.75 anys i una mediana de 73. L'edat màxima és 92 anys i la mínima 5. Per tant, la distribució de la variable edat continua essent asimètrica a l'esquerra amb el primer quartil situat als 67 anys i el tercer als 80 (taula A.VI.22).

Taula A.VI.22. Descriptius de l'edat del pacient (anys).

Mitjana	70.90	Moda	65.00
Mediana	73.00	Mínim	5.00
Mitjana retallada el 5%	72.24	Màxim	92.00
Desviació típica	13.75	Quartils	
		Q1	67.00
		Q3	80.00

Per sexes cal destacar que la distribució dels homes és més simètrica i menys dispersa que la de les dones (taula A.VI.23). Els homes tenen una mitjana i una mediana més gran que les dones i una desviació típica força més petita.

El problema més comú continua essent MPOC amb un 11.2% dels casos i el segon problema més comú en aquest cas és la dislipèmia amb un 9.6% (taula A.VI.24). Per sexes els homes presenten més problemes que les dones; un 13.1% dels homes pateixen MPOC i un 11.1% dislipèmia. Quan a les dones, els problemes que pateixen són MPOC, hipertensió arterial, dislipèmia i tabaquisme en un 3.8% dels casos per cada problema.

Taula A.VI.23. Descriptius de l'edat del pacient segons el sexe (anys).

	SEXE	
	Home	Dona
Mitjana	73.23	62.00
Mediana	73.00	68.00
Mitjana retallada el 5%	73.63	63.64
Desviació típica	9.20	22.16
Moda	67.00	77.00
Mínim	42.00	5.00
Màxim	92.00	90.00
Quartils		
Q1	67.00	52.00
Q3	80.00	78.00

Taula A.VI.24. Distribució dels individus segons sexe i problemes que pateixen.

	No presenten el problema			Presenten el problema		
	Homes	Dones	Total	Homes	Dones	Total
Consum d'alcohol	288 (97.0%)	78 (100.0%)	366 (97.6%)	9 (3.0%)	0 (0.0%)	9 (2.4%)
MPOC	258 (86.9%)	75 (96.2%)	333 (88.8%)	39 (13.1%)	3 (3.8%)	42 (11.2%)
Insuficiència cardíaca	285 (96.0%)	78 (100.0%)	363 (96.8%)	12 (4.0%)	0 (0.0%)	12 (3.2%)
Hipertensió arterial	273 (91.9%)	75 (96.2%)	348 (92.8%)	24 (8.1%)	3 (3.8%)	27 (7.2%)
Dislipèmia	264 (88.9%)	75 (96.2%)	339 (90.4%)	33 (11.1%)	3 (3.8%)	36 (9.6%)
Diabetis mellitus 2	285 (96.0%)	78 (100.0%)	363 (96.8%)	12 (4.0%)	0 (0.0%)	12 (3.2%)
Fumador	282 (94.9%)	75 (96.2%)	357 (95.2%)	15 (5.1%)	3 (3.8%)	18 (4.8%)
Ex-fumador	282 (94.9%)	78 (100.0%)	360 (96.0%)	15 (5.1%)	0 (0.0%)	15 (4.0%)

A la taula A.VI.25 s'observa que el 78.3% dels ingressos han estat per una urgència deixant la resta (21.7%) per ingressos programats. Per sexes, els percentatges corresponents als homes i a les dones ronden aquests mateixos valors.

Taula A.VI.25. Distribució dels individus segons sexe i circumstància d'ingrés.

	SEXE		Total
	Home	Dona	
Admissió urgent	230 (77.7%)	63 (80.8%)	293 (78.3%)
Admissió programada	66 (22.3%)	15 (19.2%)	81 (21.7%)
Total	296 (79.1%)	78 (20.9%)	374 (100.0%)

Quan a la circumstància d'alta mes comú en un 87.2% dels casos és la destinació al propi domicili. Com a segona causa es té el trasllat a un hospital d'aguts amb un 4.8% seguit de molt a prop de la defunció en un 4.5% dels ingressos (taula A.VI.26).

Taula A.VI.26. Distribució dels individus segons sexe i circumstància d'alta.

	SEXE		Total
	Home	Dona	
Destinació al propi domicili	258 (86.9%)	69 (88.5%)	327 (87.2%)
Trasllat a un hospital general d'aguts	14 (4.7%)	4 (5.1%)	18 (4.8%)
Trasllat a un centre de mitjana o llarga durada	3 (1.0%)	4 (5.1%)	7 (1.9%)
Trasllat a una residència social	6 (2.0%)	0 (0.0%)	6 (1.6%)
Defunció	16 (5.4%)	1 (1.3%)	17 (4.5%)
Total	297 (79.2%)	78 (20.8%)	375 (100.0%)

A continuació s'observa com la majoria dels ingressos (83.4%) provenen del propi domicili o residència social i el 12.8% són trasllats des de la consulta externa del mateix hospital (taula A.VI.27).

Taula A.VI.27. Distribució dels individus segons sexe i procedència de l'ingrés.

	SEXE		Total
	Home	Dona	
Domicili o residència social	249 (83.8%)	63 (81.8%)	312 (83.4%)
Trasllat des d'un hospital general d'aguts	5 (1.7%)	2 (2.6%)	7 (1.9%)
Trasllat des d'un recurs de la xarxa sociosanitària	1 (0.3%)	1 (1.3%)	2 (0.5%)
Trasllat des d'un centre d'atenció primària	4 (1.3%)	1 (1.3%)	5 (1.3%)
Trasllat des de consulta externa del mateix hospital	38 (12.8%)	10 (13.0%)	48 (12.8%)
Total	297 (79.4%)	77 (20.6%)	374 (100.0%)

A la taula A.VI.28 es mostra quin són els 10 GRD a cada un dels ingressos més freqüents. El més freqüent continua essent la malaltia pulmonar obstructiva crònica en un 54.9% dels ingressos, percentatge més elevat que els trobats fins ara.

A la següent taula (taula A.VI.28) es mostra l'anàlisi de la variable pes relatiu dels GRD's associats a cada un dels ingressos. En aquest cas es pot dir que globalment és pràcticament igual a la mitjana de tots els GRD's i per tant, que aquests ingressos no han generat més o menys cost que la mitjana del conjunt global de tots els GRD's.

Taula A.VI.28. Distribució dels individus segons sexe i grup relacionat amb el diagnòstic.

	SEXE		Total
	Home	Dona	
088 MALALTIA PULMONAR OBSTRUCTIVA CRONICA	178 (59.9%)	28 (35.9%)	206 (54.9%)
089 PNEUMONIA SIMPLE/PLEURITIS >17a a/cc	29 (9.8%)	4 (5.1%)	33 (8.8%)
127 INSUFICIENCIA CARDIACA i XOC	18 (6.1%)	6 (7.7%)	24 (6.4%)
082 NEOPLASIES RESPIRATORIES	16 (5.4%)	2 (2.6%)	18 (4.8%)
096 BRONQUITIS/ASMA >17a a/cc	3 (1.0%)	15 (19.2%)	18 (4.8%)
079 INFECCIONS/INFLAMACIÓ NS RESPIRAT >17 a a/cc	11 (3.7%)	1 (1.3%)	12 (3.2%)
138 ARITMIES CARDIAQUES/TRASTORNS CONDUCCIÓ a/cc	9 (3.0%)	1 (1.3%)	10 (2.7%)
097 BRONQUITIS/ASMA >17a s/cc	3 (1.0%)	4 (5.1%)	7 (1.9%)
140 ANGOR PECTORIS	6 (2.0%)	1 (1.3%)	7 (1.9%)
098 BRONQUITIS/ASMA 0-17a	0 (0.0%)	6 (7.7%)	6 (1.6%)

Taula A.VI.29. Descriptius del pes relatiu del GRD globals i segons sexe.

	SEXE		Descriptius Globals
	Home	Dona	
Mitjana	1.0013	0.9984	1.0007
Mediana	0.9317	0.9317	0.9317
Mitjana retallada el 5%	0.9808	0.9287	0.9704
Desviació típica	0.2529	0.4902	0.3163
Moda	0.9317	0.9317	0.9317
Mínim	0.5402	0.5402	0.5402
Màxim	3.6936	3.6936	3.6936
Quartils			
Q1	0.9317	0.7871	0.9317
Q3	1.0130	1.0130	1.0130

L'estada mitjana és de 7.32 dies i la mediana de 7 (taula A.VI.30). Per sexes aquests valors es mantenen pràcticament iguals. L'estada màxima és de 29 dies i la mínima de 0; això pot ser degut a individus que han estat ingressats i traslladats abans de produir un ingrés, o també individus que s'hagin mort.

Taula A.VI.30. Descriptius de l'estada globals i per sexe.

	SEXE		Descriptius Globals
	Home	Dona	
Mitjana	7.32	7.31	7.32
Mediana	6.00	7.00	7.00
Mitjana retallada el 5%	6.91	7.07	6.93
Desviació típica	4.94	4.46	4.84
Moda	4.00	3.00	6.00
Mínim	0.00	0.00	0.00
Màxim	29.00	22.00	29.00
Quartils			
Q1	4.00	4.00	4.00
Q3	9.00	10.00	10.00

A.VI.4. Anàlisi dels individus amb quatre ingressos

A l'arxiu de treball s'hi troben 48 individus que han generat 4 ingressos, això equival a 192 ingressos en total. D'aquests, el 72.9% són homes i el 27.1% dones.

Taula A.VI.31. Descriptius de l'estada globals i per sexe.

	Nº d'individus	% d'individus
Homes	140	72.9
Dones	52	27.1
Total	192	100.0

La mitjana d'edat global és de 73.04 anys, amb una desviació de 9.157 i una mediana de 72.5 anys (taula A.VI.32). L'edat màxima és de 90 anys i la mínima de 52. La distribució de l'edat continua essent asimètrica a l'esquerra, tot i que el valor dels estadístics calculats ha disminuït respecte els càlculs anteriors.

Taula A.VI.32. Descriptius de l'edat del pacient (anys).

Mitjana	73.04	Moda	71.00
Mediana	72.50	Mínim	52.00
Mitjana retallada el 5%	73.29	Màxim	90.00
Desviació típica	9.16	Quartils	
		Q1	69.25
		Q3	79.00

Per sexes, els homes continuen tenint una mitjana d'edat més elevada (74.2 anys) que les dones (69.92). La resta d'estadístics tenen uns valors similars en ambdós sexes.

Taula A.VI.33. Descriptius de l'edat del pacient segons el sexe (anys).

	SEXE	
	Home	Dona
Mitjana	74.20	69.92
Mediana	75.00	71.00
Mitjana retallada el 5%	74.47	70.08
Desviació típica	8.82	9.40
Moda	72.00	71.00
Mínim	54.00	52.00
Màxim	90.00	85.00
Quartils		
Q1	70.00	65.00
Q3	80.00	74.00

Si s'analitzen els problemes de salut que pateixen els individus que han ingressat quatre vegades durant l'any 2003 es té que MPOC i la hipertensió arterial són els dos problemes més comuns ja que es presenten en un 10.4% dels casos tots dos (taula A.VI.34). En aquest cas les dones presenten més problemes de salut que els homes essent el més comú la hipertensió arterial (30.8%) seguit de la dislipèmia (23.1%); en canvi els homes segueixen la tendència general que el problema més comú és MPOC (8.6%).

Taula A.VI.34. Distribució dels individus segons sexe i problemes de salut que pateixen.

	No presenten el problema			Presenten el problema		
	Homes	Dones	Total	Homes	Dones	Total
Consum d'alcohol	140 (100.0%)	52 (100.0%)	192 (100.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
MPOC	128 (91.4%)	44 (84.6%)	172 (89.6%)	12 (8.6%)	8 (15.4%)	20 (10.4%)
Insuficiència cardíaca	136 (97.1%)	52 (100.0%)	188 (97.9%)	4 (2.9%)	0 (0.0%)	4 (2.1%)
Hipertensió arterial	136 (97.1%)	36 (69.2%)	172 (89.6%)	4 (2.9%)	16 (30.8%)	20 (10.4%)
Dislipèmia	140 (100.0%)	40 (76.9%)	180 (93.8%)	0 (0.0%)	12 (23.1%)	12 (6.3%)
Diabetis mellitus 2	140 (100.0%)	48 (92.3%)	188 (97.9%)	0 (0.0%)	4 (7.7%)	4 (2.1%)
Fumador	136 (97.1%)	48 (92.3%)	184 (95.8%)	4 (2.9%)	4 (7.7%)	8 (4.2%)
Ex-fumador	140 (100.0%)	52 (100.0%)	192 (100.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)

Quan a la circumstància d'ingrés més comú és l'admissió urgent amb un 80.2% dels casos i la resta han estat ingressos programats (taula A.VI.35). Els homes es mantenen en aquests mateixos percentatges mentre que en les dones el percentatge d'ingressos urgents disminueix, és d'un 76.9% sobre el total d'ingressos fets per dones.

Taula A.VI.35. Distribució dels individus segons sexe i circumstància d'ingrés.

	SEXE		Total
	Home	Dona	
admissió urgent	114 (81.4%)	40 (76.9%)	154 (80.2%)
admissió programada	26 (18.6%)	12 (23.1%)	38 (19.8%)
Total	140 (72.9%)	52 (27.1%)	192 (100.0%)

I la destinació després de rebre l'alta més comú és al propi domicili amb un 90.6% dels casos.

Taula A.VI.36. Distribució dels individus segons sexe i circumstància d'alta.

	SEXE		Total
	Home	Dona	
Destinació al propi domicili	123 (87.9%)	51 (98.1%)	174 (90.6%)
	5	1	6
Trasllat a un hospital general d'aguts	(3.6%) 3	(1.9%) 0	(3.1%) 3
Trasllat a un centre de mitjana o llarga durada	(2.1%) 3	(0.0%) 0	(1.6%) 3
Trasllat a una residència social	(2.1%) 6	(0.0%) 0	(1.6%) 6
Defunció	(4.3%)	(0.0%)	(3.1%)
Total	140 (72.9%)	52 (27.1%)	192 (100.0%)

Quan a la procedència del pacient en el moment d'ingressar s'observa que el 82.8% dels ingressats provenien del propi domicili o d'una residència social i el 14.6% de consultes externes del mateix hospital (taula A.VI.37).

Taula A.VI.37. Distribució dels individus segons sexe i procedència de l'ingrés.

	SEXE		Total
	Home	Dona	
Domicili o residència social	119 (85.0%)	40 (76.9%)	159 (82.8%)
Trasllat des d'un hospital general d'aguts	2 (1.4%)	0 (0.0%)	2 (1.0%)
Trasllat des d'un recurs de la xarxa sociosanitària	2 (1.4%)	0 (0.0%)	2 (1.0%)
Trasllat des d'un centre d'atenció primària	1 (0.7%)	0 (0.0%)	1 (0.5%)
Trasllat des de consulta externa del mateix hospital	16 (11.4%)	12 (23.1%)	28 (14.6%)
Total	140 (72.9%)	52 (27.1%)	192 (100.0%)

A la següent taula (taula A.VI.38) es mostren els 10 GRD's més freqüents dels 832 ingressos estudiats per sexe. El més comú continua essent MPOC amb un percentatge del 56.8% dels ingressos.

Taula A.VI.38. Distribució dels individus segons sexe i grup relacionat amb el diagnòstic.

	SEXE		Total
	Home	Dona	
088 MALALTIA PULMONAR OBSTRUCTIVA CRONICA	81 (57.9%)	28 (53.8%)	109 (56.8%)
089 PNEUMONIA SIMPLE/PLEURITIS >17a a/cc	14 (10.0%)	4 (7.7%)	18 (9.4%)
127 INSUFICIENCIA CARDIACA i XOC	11 (7.9%)	7 (13.5%)	18 (9.4%)
096 BRONQUITIS/ASMA >17a a/cc	4 (2.9%)	8 (15.4%)	12 (6.3%)
082 NEOPLASIES RESPIRATORIES	9 (6.4%)	0 (0.0%)	9 (4.7%)
087 EDEMA PULMONAR/ INSUFICIÈNCIA RESPIRATÒRIA	4 (2.9%)	1 (1.9%)	5 (2.6%)
099 SIGNES/SIMPTOMES RESPIRATORIS a/cc	3 (2.1%)	0 (0.0%)	3 (1.6%)
130 TRASTORNS VASCULARS PERIFÈRICS a/cc	3 (2.1%)	0 (0.0%)	3 (1.6%)
140 ANGOR PECTORIS	3 (2.1%)	0 (0.0%)	3 (1.6%)

A la taula A.VI.39 es presenten uns descriptius del pes relatiu del total de GRD's. En aquest cas, tant la mitjana global com la mitjana per sexes es troba lleugerament per sota de 1 la qual cosa indica que el cost mitjà global i per sexes és inferior al cost mig del conjunt global de tots els GRD's.

L'anàlisi de la durada de l'estada a l'hospital mostra com el valor dels estadístics calculats han tornat a augmentar respecte els valors dels apartats anteriors cosa que sembla indicar que com més ingressos més llarga és l'estada (taula A.VI.40).

Taula A.VI.39. Descriptius del pes relatiu del GRD globals i segons sexe.

	SEXE		Descriptius Globals
	Home	Dona	
Mitjana	0.9881	0.9322	0.9729
Mediana	0.9317	0.9317	0.9317
Mitjana retallada el 5%	0.9824	0.9299	0.9654
Desviació típica	0.1678	0.1123	0.1564
Moda	0.9317	0.9317	0.9317
Mínim	0.5740	0.5873	0.5740
Màxim	1.6501	1.3781	1.6501
Quartils			
Q1	0.9317	0.9317	0.9317
Q3	1.0130	0.9927	1.0130

Taula A.VI.40. Descriptius de l'estada globals i per sexe.

	SEXE		Descriptius Globals
	Home	Dona	
Mitjana	7.72	7.98	7.79
Mediana	6.00	7.00	6.50
Mitjana retallada el 5%	7.33	7.67	7.42
Desviació típica	5.49	5.05	5.36
Moda	5.00	7.00	4.00
Mínim	0.00	0.00	0.00
Màxim	27.00	23.00	27.00
Quartils			
Q1	4.00	4.00	4.00
Q3	11.00	11.00	11.00

A.VI.5. Anàlisi dels individus amb cinc ingressos.

Del total d'individus de l'arxiu de treball, 21 han ingressat cinc vegades durant l'any 2003, generant un total de 105 ingressos. D'aquests, el 85% són homes i el 15% dones.

Taula A.VI.41. Distribució dels individus segons el sexe.

	Nº d'individus	% d'individus
Homes	85	81.0
Dones	20	19.0
Total	105	100.0

L'edat mitjana global és de 69.15 anys amb una desviació de 17.736 i una mediana de 75.5 anys. L'edat màxima és de 86 anys i la mínima de 3 (taula A.VI.42).

Taula A.VI.42. Descriptius de l'edat del pacient (anys).

Mitjana	69.81	Moda	62.00
Mediana	76.00	Mínim	3.00
Mitjana retallada el 5%	72.49	Màxim	86.00
Desviació típica	17.56	Quartils	
		Q1	64.00
		Q3	80.00

En aquest cas per sexes, els homes tenen una mitjana d'edat inferior a les dones, 68.65 anys versus 72, i si observem la mediana i la mitjana retallada, també tenim que per les dones prenen valors més alts (taula A.VI.43). Per tant es pot afirmar que les dones que han ingressat durant el 2003, cinc vegades tenen una edat més alta que els homes.

Taula A.VI.43. Descriptius de l'edat del pacient segons el sexe (anys).

	SEXE	
	Home	Dona
Mitjana	68.65	74.75
Mediana	75.00	79.00
Mitjana retallada el 5%	71.33	75.22
Desviació típica	18.74	10.17
Moda	62.00	58.00
Mínim	3.00	58.00
Màxim	86.00	83.00
Quartils		
Q1	64.00	62.75
Q3	78.00	82.50

El problema més comú torna a ser MPOC amb un 15% dels casos, seguit de la dislipèmia en un 10% (taula A.VI.44). Per la seva banda, els homes segueixen aquesta mateixa classificació amb un 17.6% i un 11.8% dels casos respectivament; en canvi, les dones només pateixen dues malalties, MPOC i insuficiència cardíaca les dues amb un 25%.

Taula A.VI.44. Distribució dels individus segons sexe i problemes de salut que pateixen.

	No presenten el problema			Presenten el problema		
	Homes	Dones	Total	Homes	Dones	Total
Consum d'alcohol	80 (94.1)	20 (100.0%)	95 (95.0%)	5 (5.9%)	0 (0.0%)	5 (5.0%)
MPOC	70 (82.4%)	15 (75.0%)	85 (85.0%)	15 (17.6%)	5 (25.0%)	15 (15.0%)
Insuficiència cardíaca	85 (100.0%)	15 (75.0%)	100 (100.0%)	0 (0.0%)	5 (25.0%)	0 (0.0%)
Hipertensió arterial	80 (94.1%)	20 (100.0%)	95 (95.0%)	5 (5.9%)	0 (0.0%)	5 (5.0%)
Dislipèmia	75 (88.2%)	20 (100.0%)	90 (90.0%)	10 (11.8%)	0 (0.0%)	10 (10.0%)
Diabetis mellitus 2	85 (100.0%)	20 (100.0%)	100 (100.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Fumador	80 (94.1%)	20 (100.0%)	95 (95.0%)	5 (5.9%)	0 (0.0%)	5 (5.0%)
Ex-fumador	85 (100.0%)	20 (100.0%)	100 (100.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)

El 84% dels ingressos han estat deguts a una urgència i la resta (16%) eren admissions programades (taula A.VI.45). Quan a les dones el percentatge d'ingressos per urgència puja fins a un 95% dels casos, en canvi en els homes es manté sobre la línia global, un 82.4%.

A la taula A.VI.46 s'observa que el 89% de les altes s'han enviat al propi domicili, i només el 6% han mort. En el cas de les dones aquest percentatge augmenta fins a un 95% del total, en canvi en els homes es manté amb un 88.2% dels casos.

Taula A.VI.45. Distribució dels individus segons sexe i circumstància d'ingrés.

	SEXE		Total
	Home	Dona	
Admissió urgent	70 (82.4%)	19 (95.0%)	89 (84.8%)
Admissió programada	15 (17.6%)	1 (5.0%)	16 (15.2%)
Total	85 (81.0%)	20 (19.0%)	100 (100.0%)

Taula A.VI.46. Distribució dels individus segons sexe i circumstància d'alta.

	SEXE		Total
	Home	Dona	
Destinació al propi domicili	75 (88.2%)	19 (95.0%)	94 (89.5%)
Trasllat a un hospital general	2 (2.4%)	0 (0.0%)	2 (1.9%)
Trasllat a una residència social	3 (3.5%)	0 (0.0%)	3 (2.9%)
Defunció	5 (5.9%)	1 (5.0%)	6 (5.7%)
Total	85 (81.0%)	20 (19.0%)	100 (100.0%)

El 80% dels individus ingressats provenien del propi domicili o d'una residència social, seguit per un 5.7% que provenien de la consulta externa del mateix hospital (taula A.VI.47). En els homes es segueix aquesta característica global amb un 81.2% dels casos que provenien del propi domicili i el 11.8% de la consulta externa; en canvi en les dones el 80% provenien del domicili i la resta, el 20% van ser traslladades d'un centre d'atenció primària.

Tot seguit es mostra una primera taula (taula A.VI.48) amb els 10 GRD més freqüents i una segona taula (taula A.VI.49) amb uns descriptius del pes relatiu del total dels GRD. Com ha passat en tots els casos estudiats el GRD més freqüent és la malaltia pulmonar obstructiva crònica que s'ha catalogat en un 60% dels ingressos.

Taula A.VI.47. Distribució dels individus segons sexe i procedència.

	SEXE		Total
	Home	Dona	
Domicili o residència social	69 (81.2%)	16 (80.0%)	85 (81.0%)
Trasllat des d'un hospital general d'aguts	1 (1.2%)	0 (0.0%)	1 (1.0%)
Trasllat des d'un recurs de la xarxa sociosanitària	3 (3.5%)	0 (0.0%)	3 (2.9%)
Trasllat des d'un centre d'atenció primària	2 (2.4%)	4 (20.0%)	6 (5.7%)
Trasllat des de consulta externa del mateix hospital	10 (11.8%)	0 (0.0%)	10 (9.5%)
Total	85 (81.0%)	20 (19.0%)	100 (100.0%)

Taula A.VI.48. Distribució dels individus segons sexe i grup relacionat amb el diagnòstic.

	SEXE		Total
	Home	Dona	
088 MALALTIA PULMONAR OBSTRUCTIVA CRONICA	54 (63.5%)	9 (45.0%)	63 (60.0%)
127 INSUFICIENCIA CARDIACA i XOC	4 (4.7%)	9 (45.0%)	13 (12.4%)
089 PNEUMONIA SIMPLE/PLEURITIS >17a a/cc	9 (10.6%)	0 (0.0%)	9 (8.6%)
098 BRONQUITIS/ASMA 0-17a	5 (5.9%)	0 (0.0%)	5 (4.8%)
087 EDEMA PULMONAR/ INSUFICIÈNCIA RESPIRATÒRIA	3 (3.5%)	1 (5.0%)	4 (3.8%)
129 ATURADA CARDIACA DE CAUSA DESCONEGUDA	2 (2.4%)	0 (0.0%)	2 (1.9%)
140 ANGOR PECTORIS	2 (2.4%)	0 (0.0%)	2 (1.9%)
083 TRAUMATISME TORACIC MAJOR a/cc	1 (1.2%)	0 (0.0%)	1 (1.0%)
096 BRONQUITIS/ASMA >17a a/cc	1 (1.2%)	0 (0.0%)	1 (1.0%)
099 SIGNES/SIMPTOMES RESPIRATORIS a/cc	1 (1.2%)	0 (0.0%)	1 (1.0%)

En aquest cas la mitjana del pes relatiu pren un valor inferior a la unitat tan globalment com per sexes; tot i que en el cas dels homes, és quan es pren el valor més allunyat de la unitat.

Taula A.VI.49. Descriptius del pes relatiu del GRD globals i segons sexe.

	SEXE		Descriptius Globals
	Home	Dona	
Mitjana	0.9599	0.9855	0.9648
Mediana	0.9317	0.9723	0.9317
Mitjana retallada el 5%	0.9504	0.9723	0.9546
Desviació típica	0.1381	0.1055	0.1324
Moda	0.9317	0.9317 ⁽¹⁾	0.9317
Mínim	0.5740	0.8291	0.5740
Màxim	1.6194	1.3781	1.6194
Quartils			
Q1	0.9317	0.9317	0.9317
Q3	0.9317	1.0130	1.0130

(1)Existeixen diversos modes; es mostra el més petit de tots.

Finalment s'observa com la durada de l'estada ha estat inferior en els homes que en les dones; i globalment es veu com en el cas dels individus amb cinc ingressos la variable estada pren els seus valors més elevats.

Taula A.VI.50. Descriptius de l'estada globals i per sexe.

	SEXE		Descriptius Globals
	Home	Dona	
Mitjana	7.75	9.05	8.00
Mediana	7.00	7.00	7.00
Mitjana retallada el 5%	7.30	8.72	7.55
Desviació típica	5.40	6.61	5.64
Moda	6.00	6.00	6.00
Mínim	0.00	0.00	0.00
Màxim	30.00	24.00	30.00
Quartils			
Q1	4.00	4.25	4.00
Q3	10.00	13.25	10.00

ANNEX VII: RESULTATS OBTINGUTS EN L'APLICACIÓ DEL MÈTODE D'ANDERSEN

A.VII.1. MODEL SENSE EFECTES ESPACIALS

Random effects:

Formula: ~1 | id

(Intercept) Residual

StdDev: 0.2504197 0.1948502

Variance function:

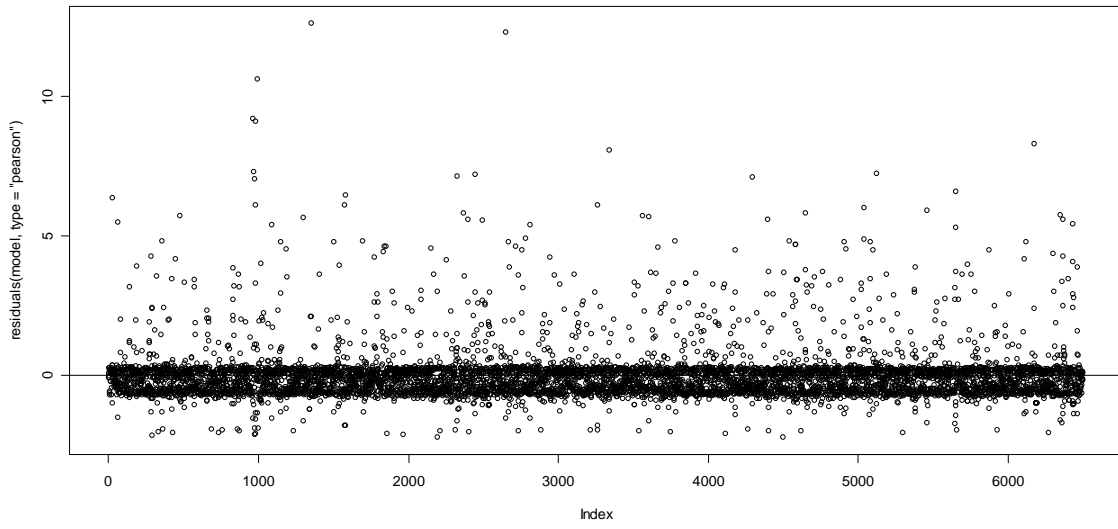
Structure: fixed weights

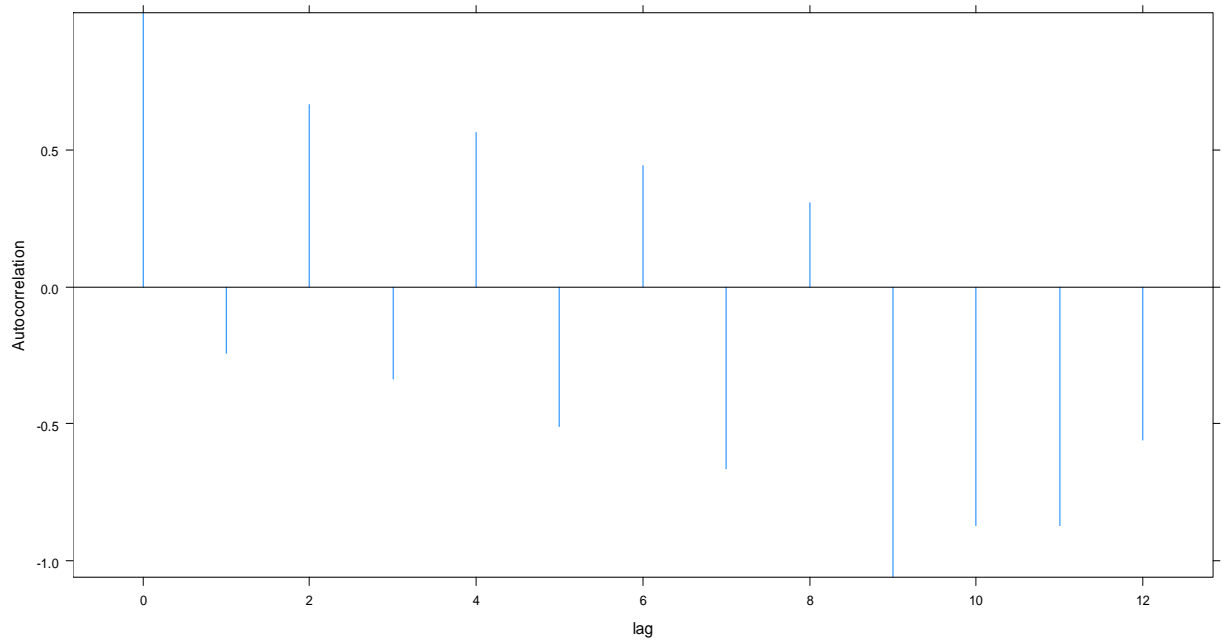
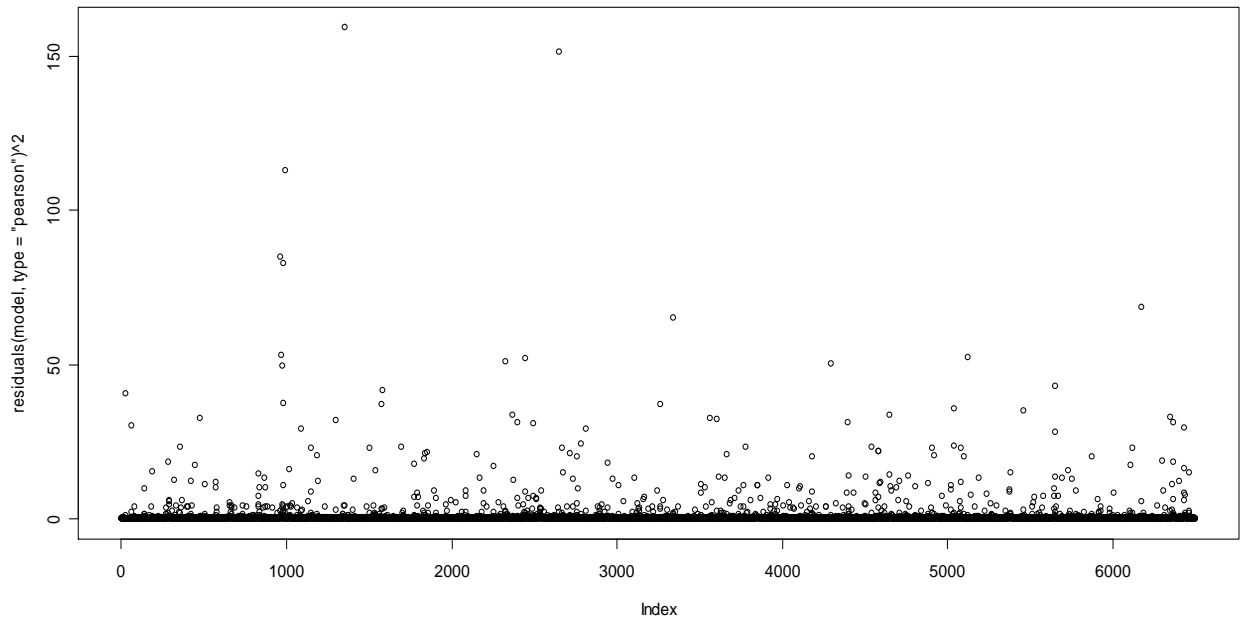
Formula: ~invwt

Fixed effects: pseudo ~ Sexe + factor(enum) + factor(edatc) + Valcohol + Vmpoc + Insufcar + Hta + Dislip + Dm2 + Tab + Extab + Nvisites

	Value	Std.Error	DF	t-value	p-value
(Intercept)	-1.6591315	0.02041477	4199	-81.27113	0.0000
Sexe	-0.0178611	0.01945938	2283	-0.91786	0.3588
factor(enum)2	-1.8072845	0.03366153	4199	-53.68991	0.0000
factor(enum)3	-2.3777500	0.07193913	4199	-33.05225	0.0000
factor(enum)4	-2.6605934	0.12585184	4199	-21.14068	0.0000
factor(enum)5	-1.8470911	0.12867885	4199	-14.35427	0.0000
factor(enum)6	-2.4377409	0.20091622	4199	-12.13312	0.0000
factor(edatc)1	-0.0274629	0.02695197	2283	-1.01896	0.3083
factor(edatc)2	-0.0297704	0.02677215	2283	-1.11199	0.2663
factor(edatc)3	0.0138328	0.02835240	2283	0.48789	0.6257
factor(edatc)4	0.0173410	0.02691343	2283	0.64432	0.5194
Valcohol	-0.0074664	0.06330646	2283	-0.11794	0.9061
Vmpoc	0.0024612	0.02754045	2283	0.08937	0.9288
Insufcar	-0.0153832	0.06480935	2283	-0.23736	0.8124
Hta	0.0216022	0.03215114	2283	0.67190	0.5017
Dislip	-0.0407920	0.03801328	2283	-1.07310	0.2833
Dm2	-0.0032552	0.03971651	2283	-0.08196	0.9347
Tab	-0.0437124	0.04003297	2283	-1.09191	0.2750
Extab	0.0482060	0.06912849	2283	0.69734	0.4857
Nvisites	-0.0164972	0.00601318	2283	-2.74351	0.0061

Anàlisi residual:

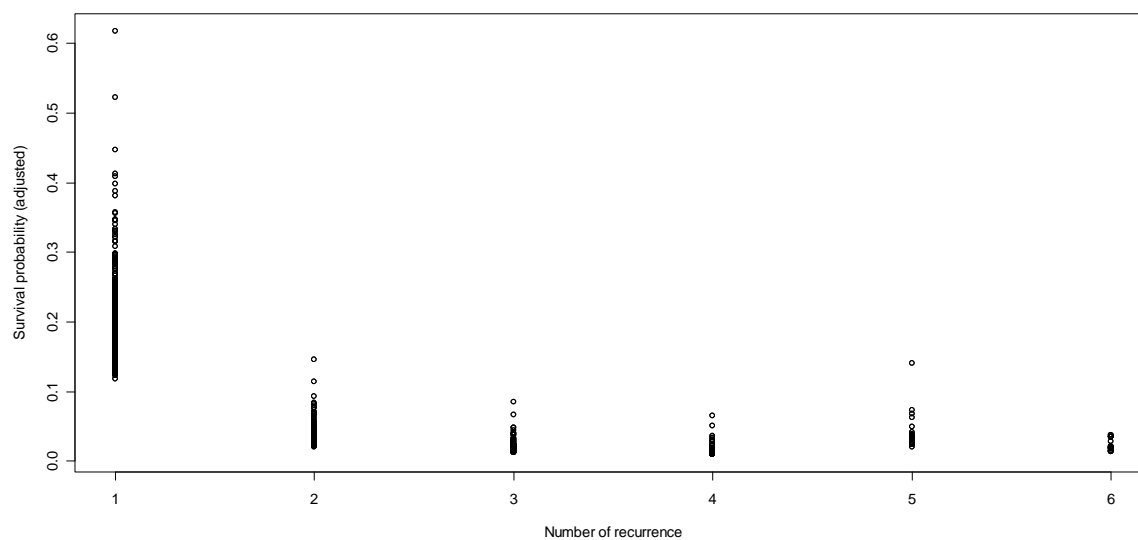


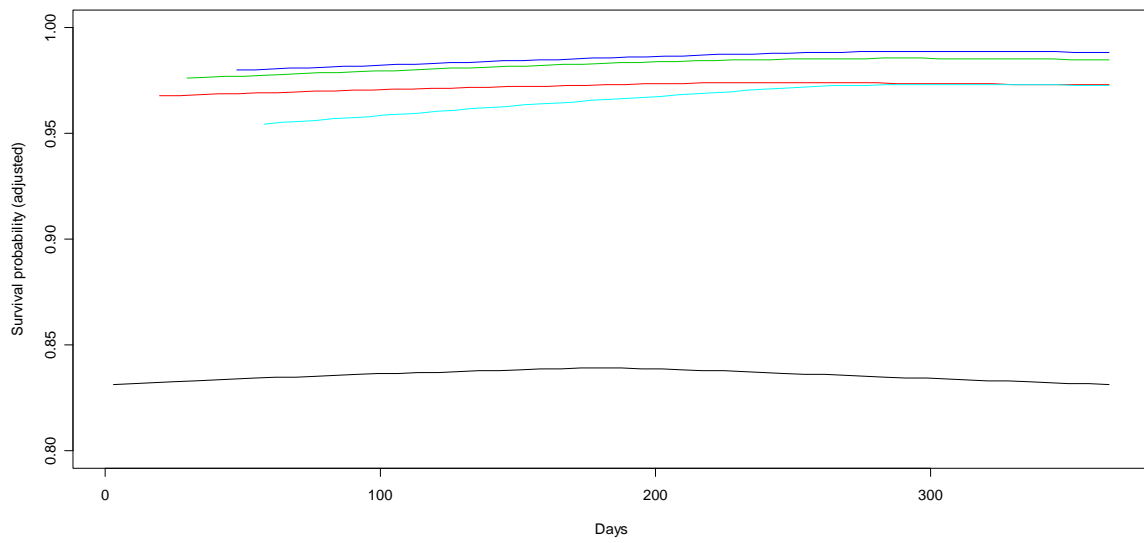


Càlcul de l'ODDS RATIO:

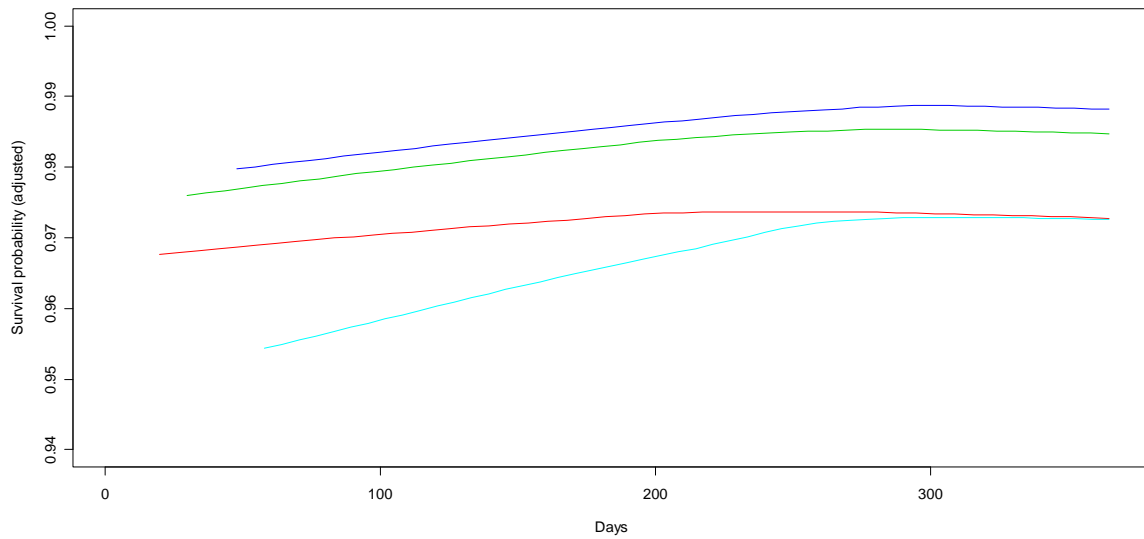
```
[ ,1]
(Intercept)    0.19030418
Sexe           0.98229750
factor(enum)2  0.16409915
factor(enum)3  0.09275905
factor(enum)4  0.06990673
factor(enum)5  0.15769522
factor(enum)6  0.08735798
factor(edatc)1 0.97291075
factor(edatc)2 0.97066839
factor(edatc)3 1.01392896
factor(edatc)4 1.01749219
Valcohol      0.99256142
Vmpoc        1.00246419
Insufcar     0.98473448
Hta          1.02183722
Dislip       0.96002877
Dm2         0.99675012
Tab          0.95722923
Extab       1.04938683
Nvisites    0.98363814
```

Gràfics de la probabilitat de supervivència segons el nombre d'ingressos:



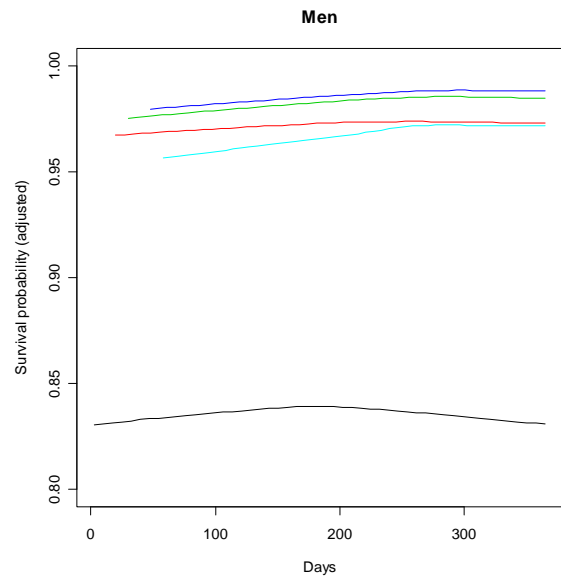
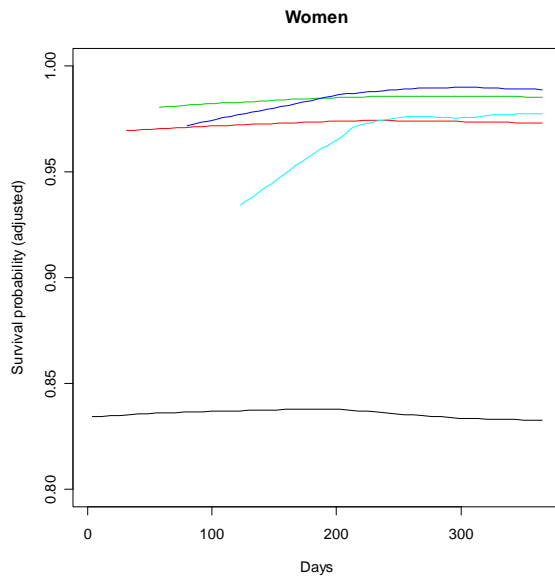


Negre: primer ingrés
 Vermell: segon ingrés
 Verd: tercer ingrés
 Blau: quart ingrés
 Turquesa: cinquè ingrés

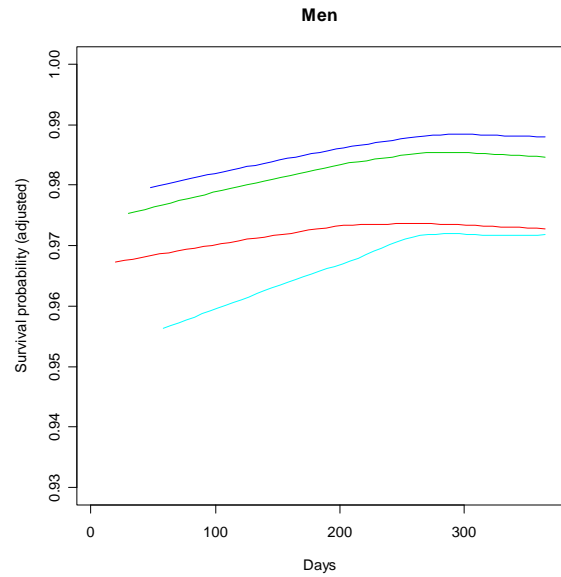
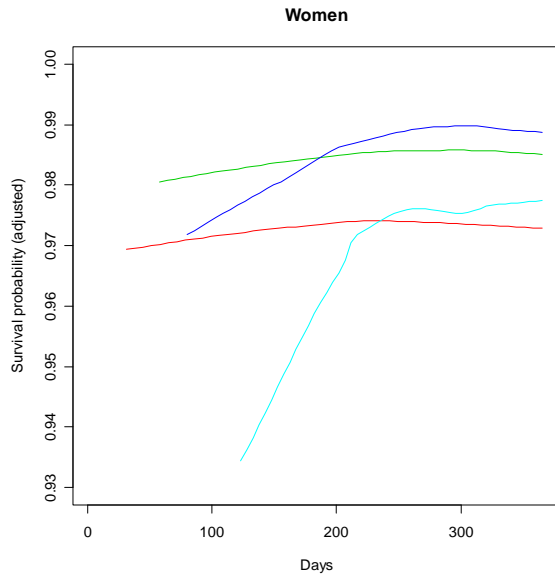


Vermell: segon ingrés
 Verd: tercer ingrés
 Blau: quart ingrés
 Turquesa: cinquè ingrés

Gràfics de la probabilitat de supervivència segons el nombre d'ingressos per sexes:

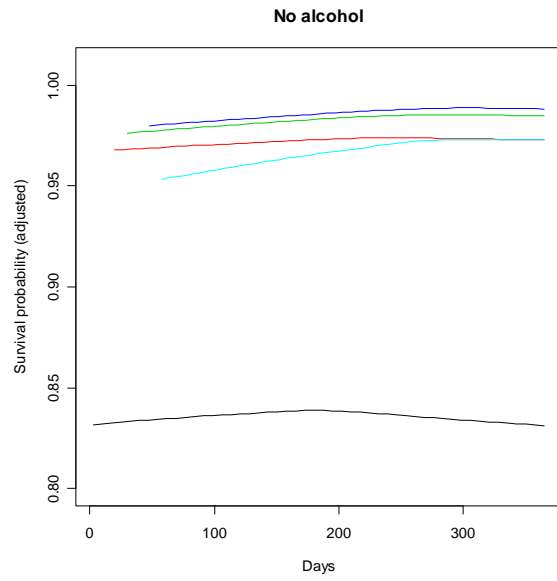
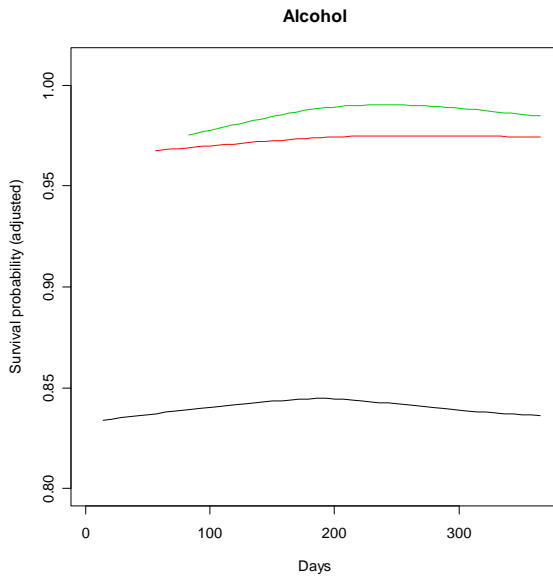


Negre: primer ingrés
 Vermell: segon ingrés
 Verd: tercer ingrés
 Blau: quart ingrés
 Turquesa: cinquè ingrés

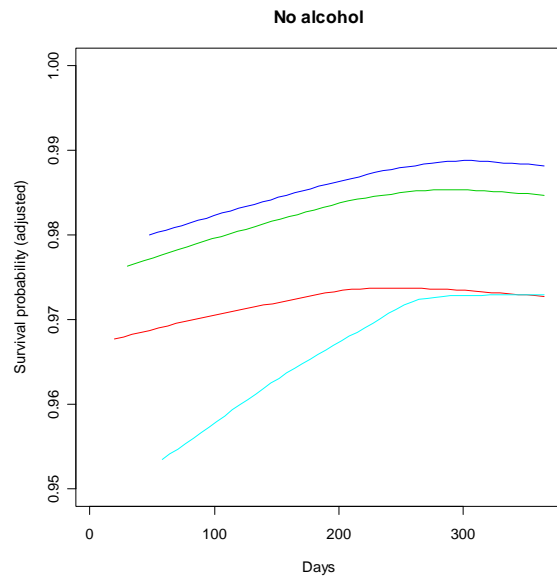
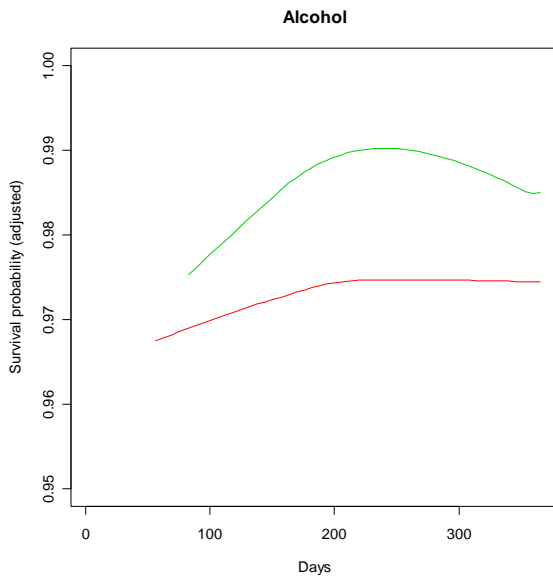


Vermell: segon ingrés
 Verd: tercer ingrés
 Blau: quart ingrés
 Turquesa: cinquè ingrés

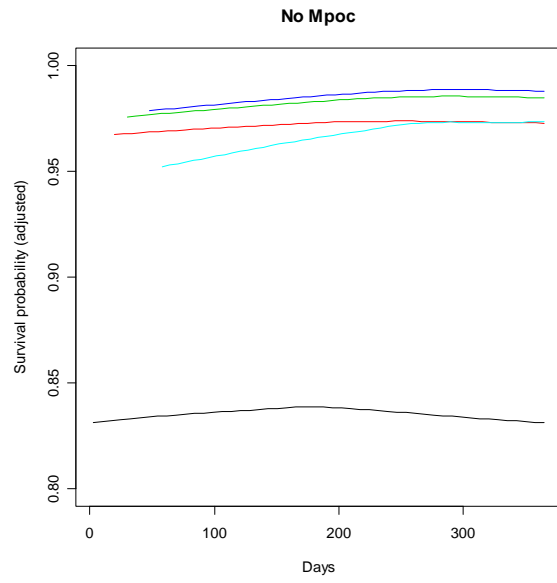
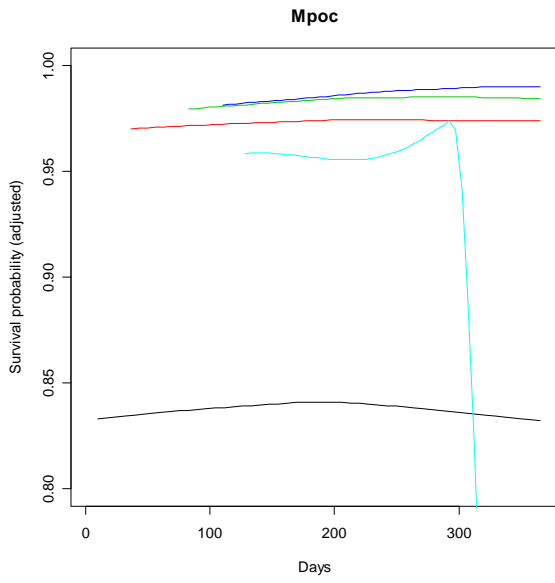
Gràfics de la probabilitat de supervivència segons el nombre d'ingressos per problemes crònics::



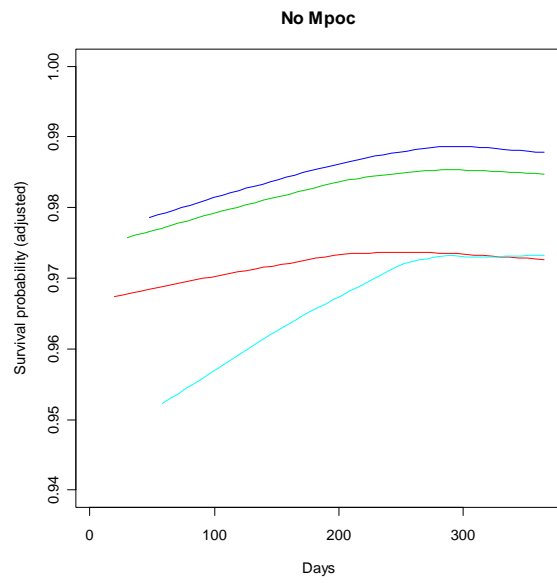
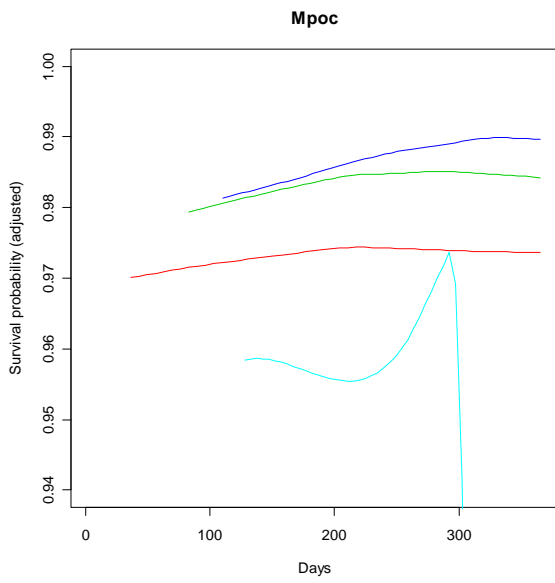
Negre: primer ingrés
 Vermell: segon ingrés
 Verd: tercer ingrés
 Blau: quart ingrés
 Turquesa: cinquè ingrés



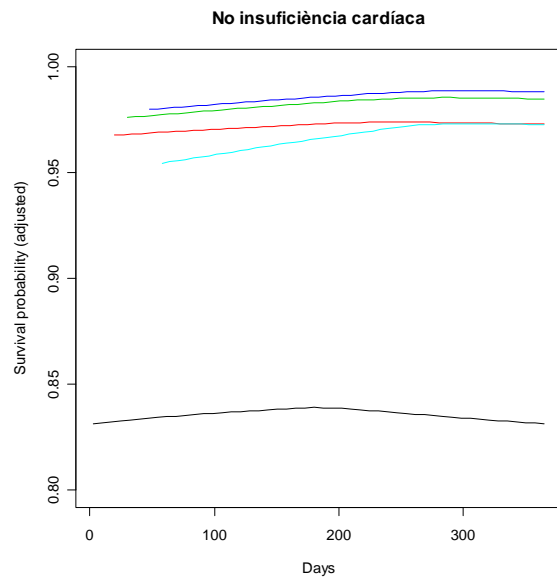
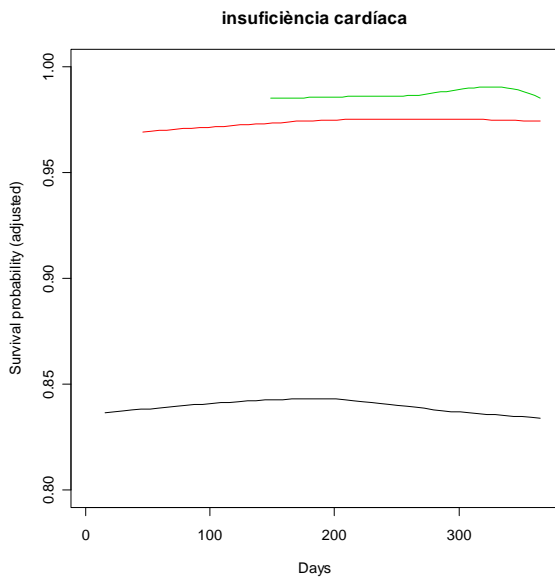
Vermell: segon ingrés
 Verd: tercer ingrés
 Blau: quart ingrés
 Turquesa: cinquè ingrés



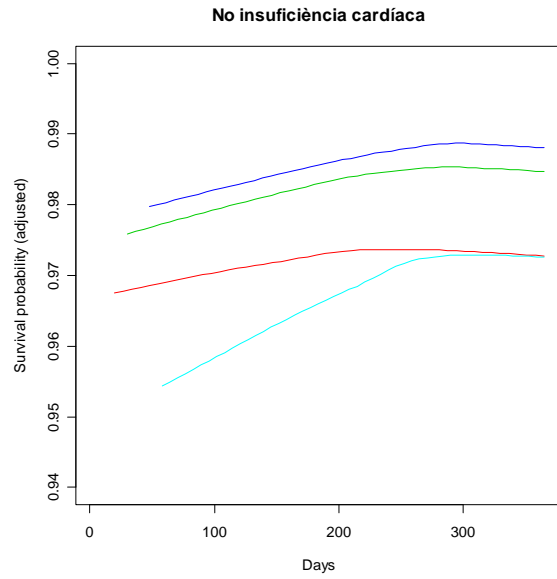
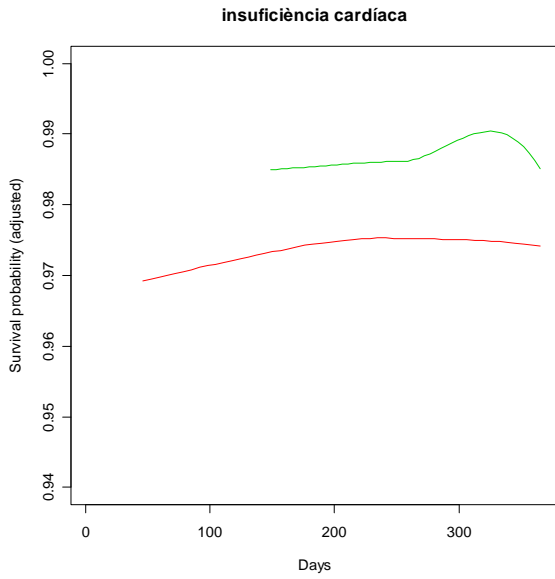
Negre: primer ingrès
 Vermell: segon ingrès
 Verd: tercer ingrès
 Blau: quart ingrès
 Turquesa: cinquè ingrès



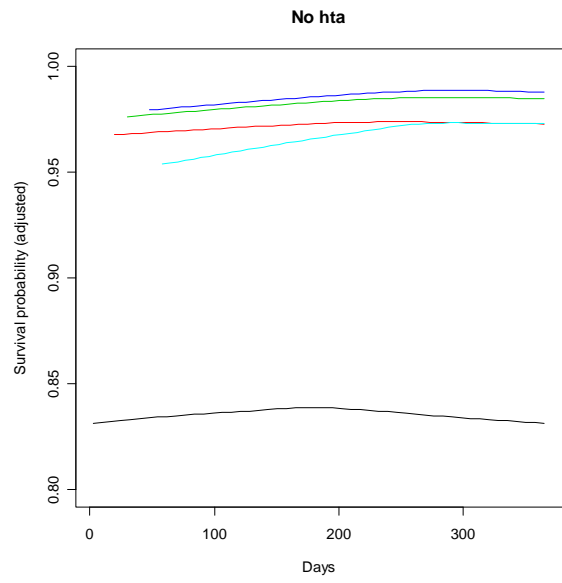
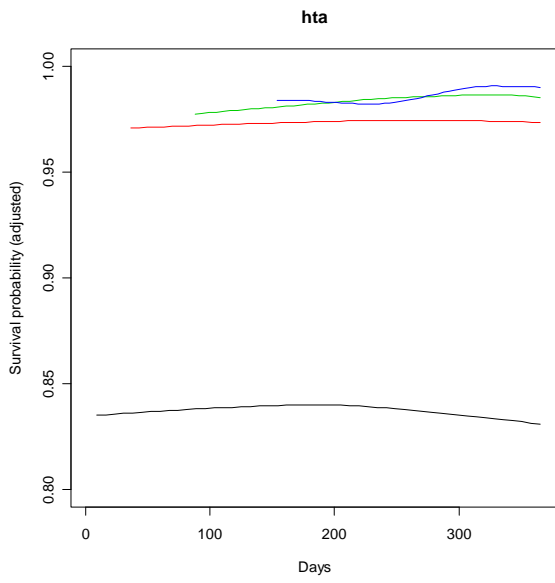
Vermell: segon ingrès
 Verd: tercer ingrès
 Blau: quart ingrès
 Turquesa: cinquè ingrès



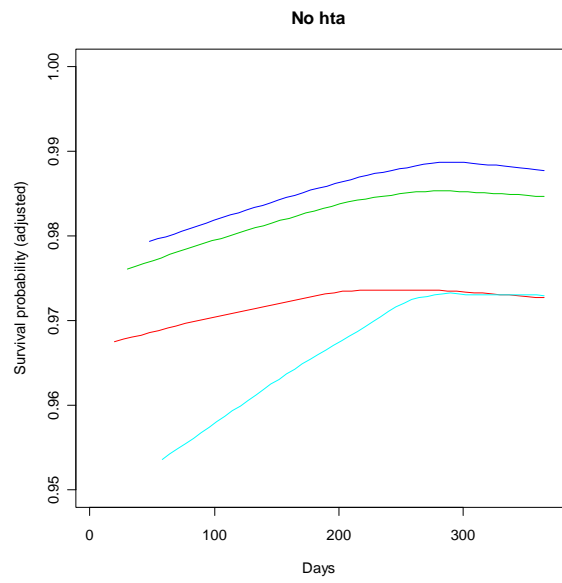
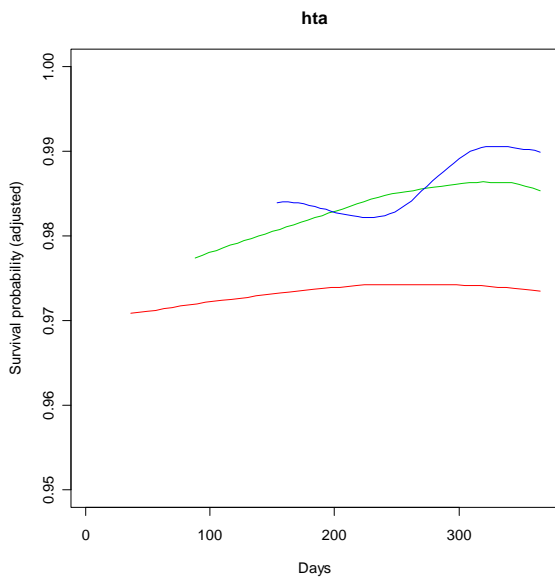
Negre: primer ingrés
 Vermell: segon ingrés
 Verd: tercer ingrés
 Blau: quart ingrés
 Turquesa: cinquè ingrés



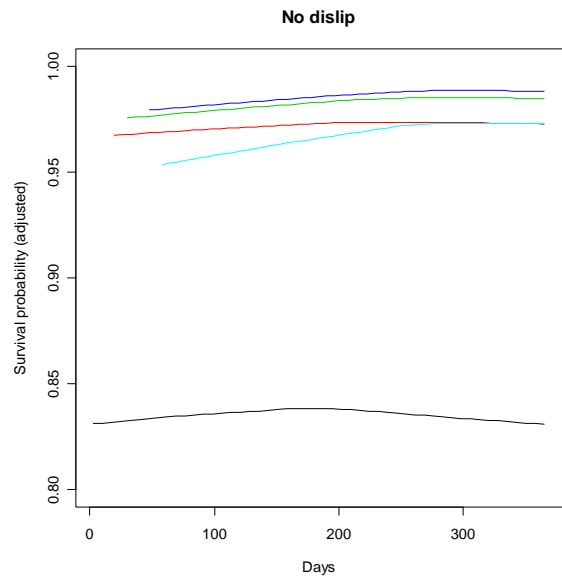
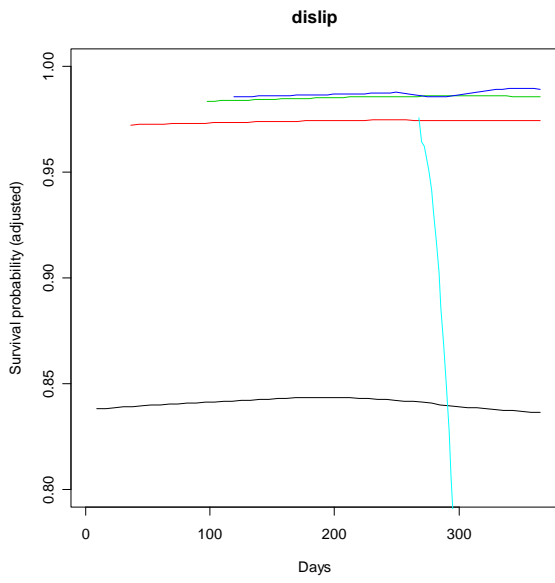
Vermell: segon ingrés
 Verd: tercer ingrés
 Blau: quart ingrés
 Turquesa: cinquè ingrés



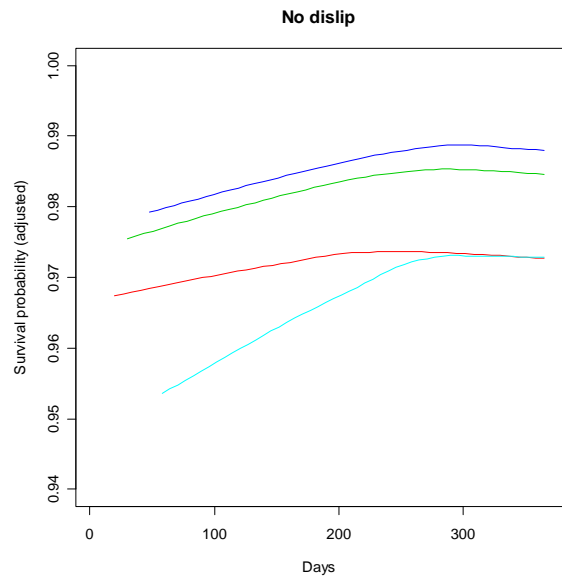
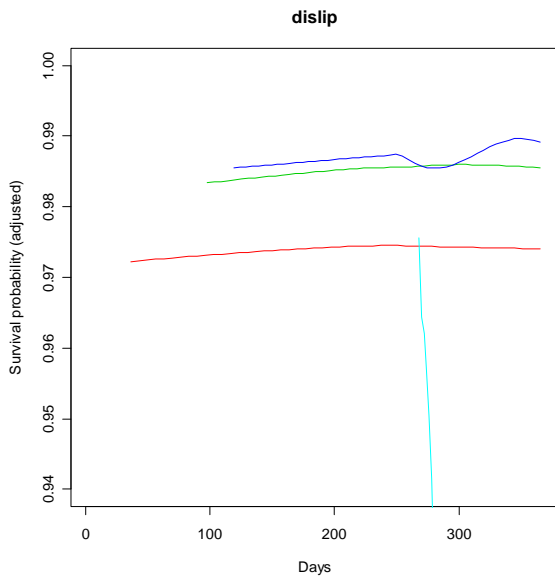
Negre: primer ingrès
 Vermell: segon ingrès
 Verd: tercer ingrès
 Blau: quart ingrès
 Turquesa: cinquè ingrès



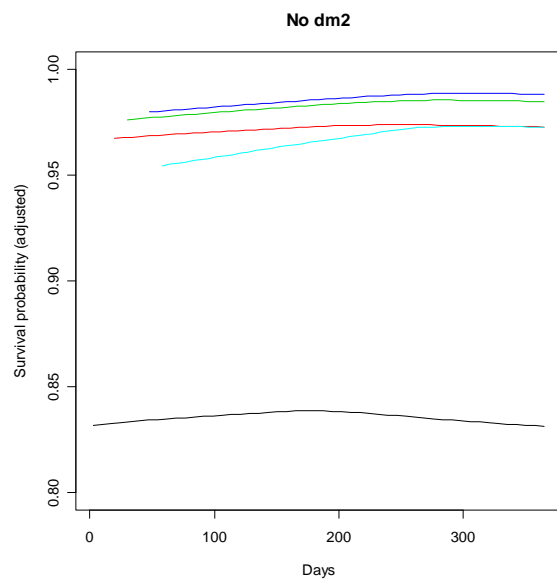
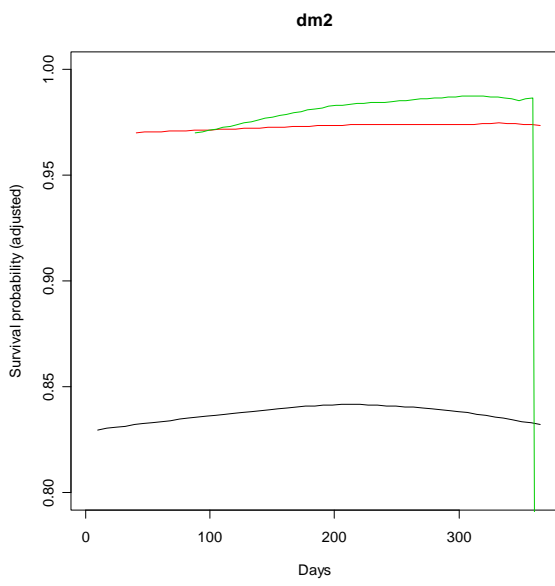
Vermell: segon ingrès
 Verd: tercer ingrès
 Blau: quart ingrès
 Turquesa: cinquè ingrès



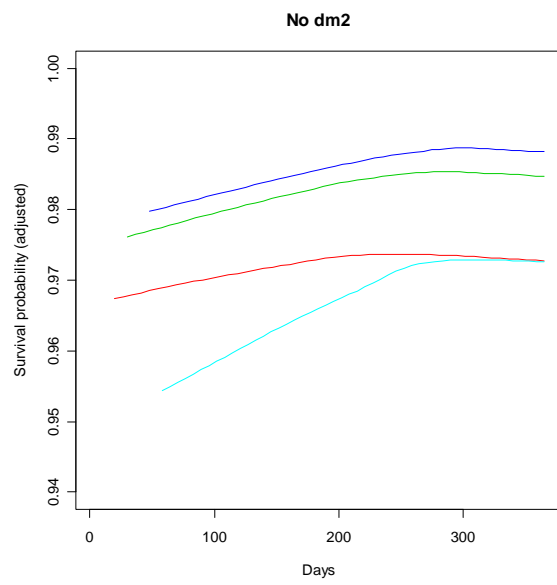
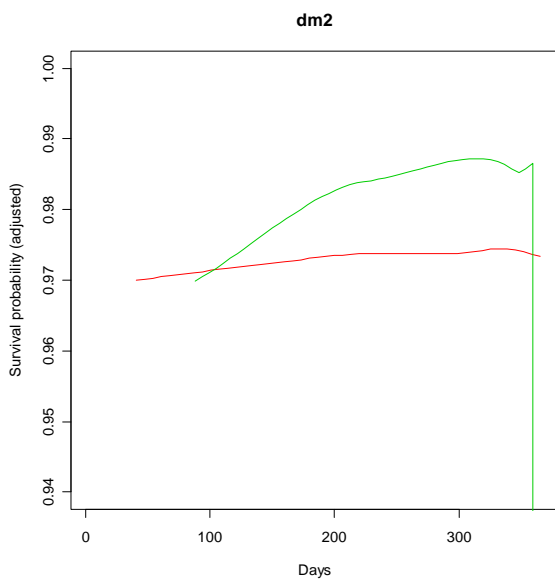
Negre: primer ingrés
 Vermell: segon ingrés
 Verd: tercer ingrés
 Blau: quart ingrés
 Turquesa: cinquè ingrés



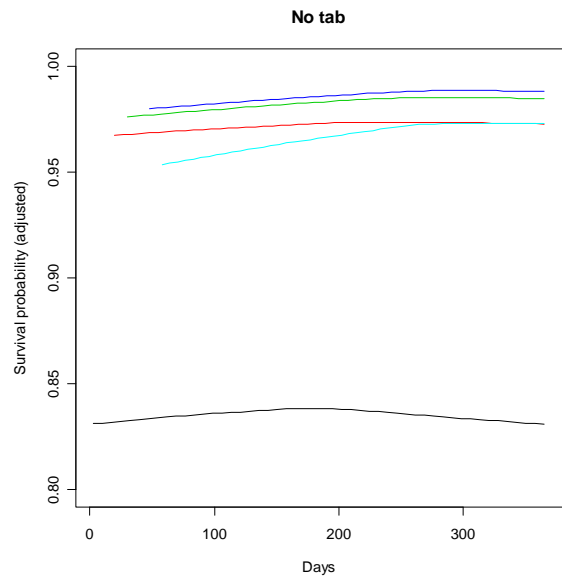
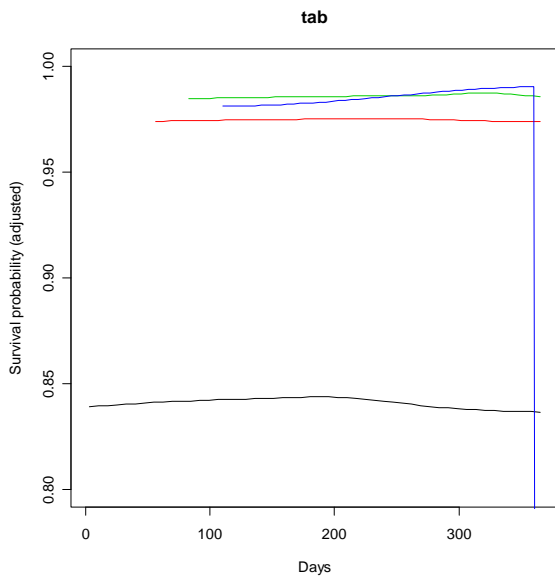
Vermell: segon ingrés
 Verd: tercer ingrés
 Blau: quart ingrés
 Turquesa: cinquè ingrés



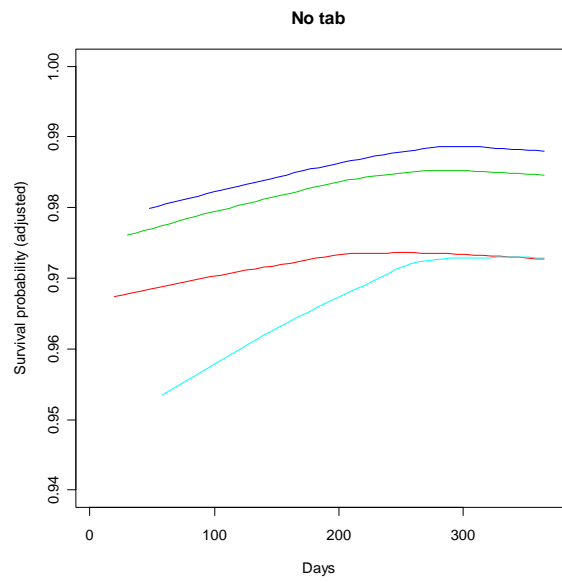
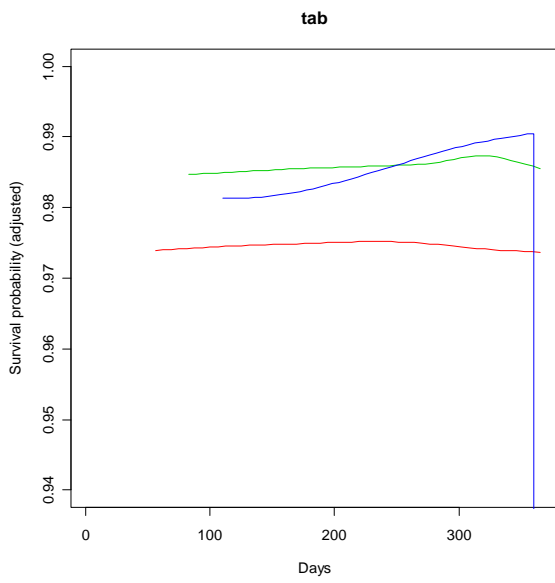
Negre: primer ingrés
 Vermell: segon ingrés
 Verd: tercer ingrés
 Blau: quart ingrés
 Turquesa: cinquè ingrés



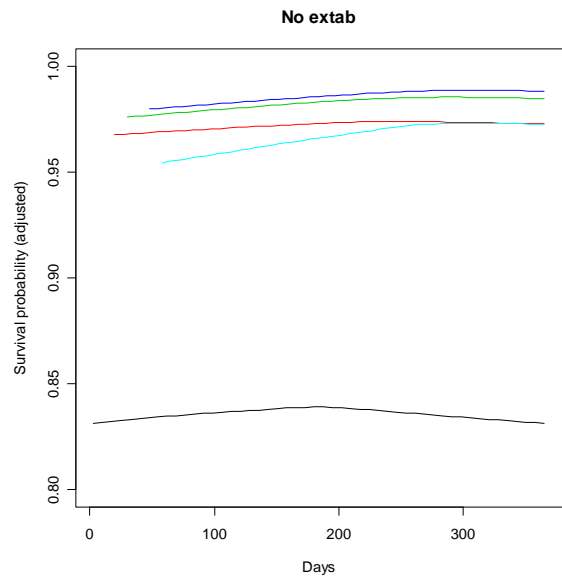
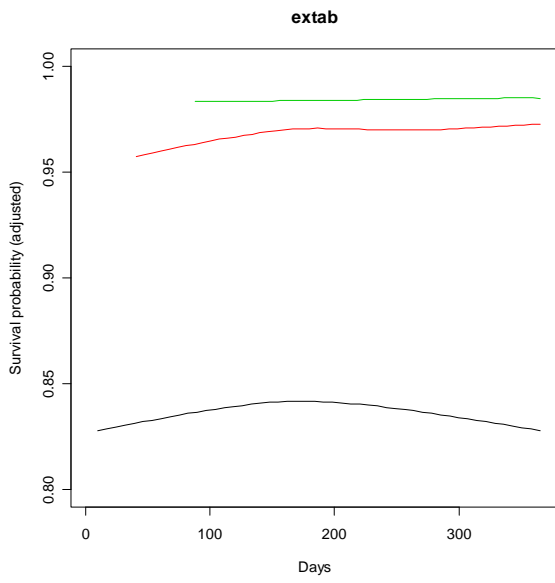
Vermell: segon ingrés
 Verd: tercer ingrés
 Blau: quart ingrés
 Turquesa: cinquè ingrés



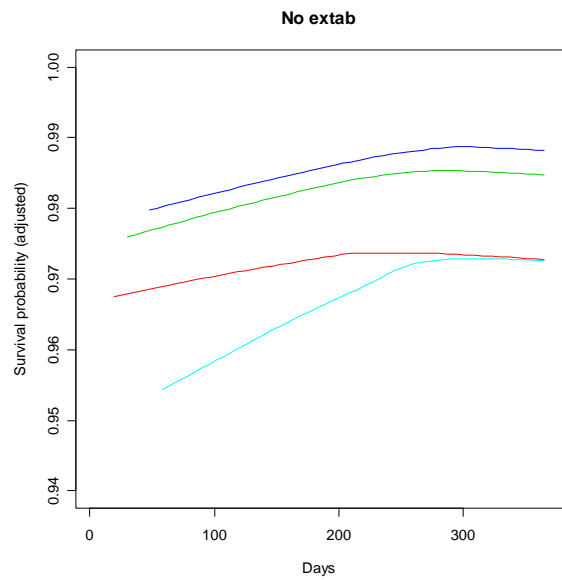
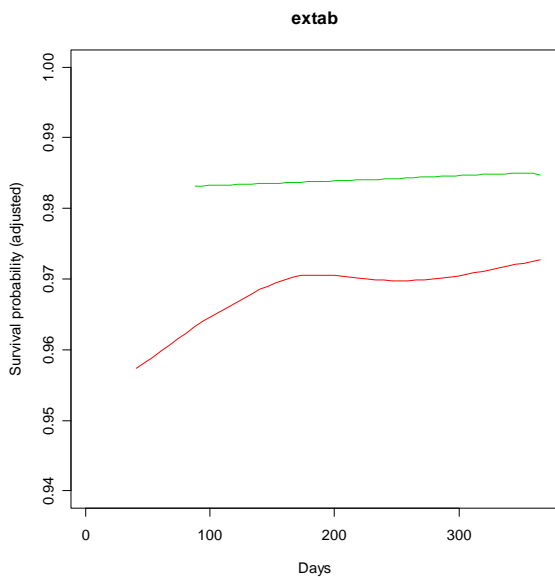
Negre: primer ingrés
 Vermell: segon ingrés
 Verd: tercer ingrés
 Blau: quart ingrés
 Turquesa: cinquè ingrés



Vermell: segon ingrés
 Verd: tercer ingrés
 Blau: quart ingrés
 Turquesa: cinquè ingrés

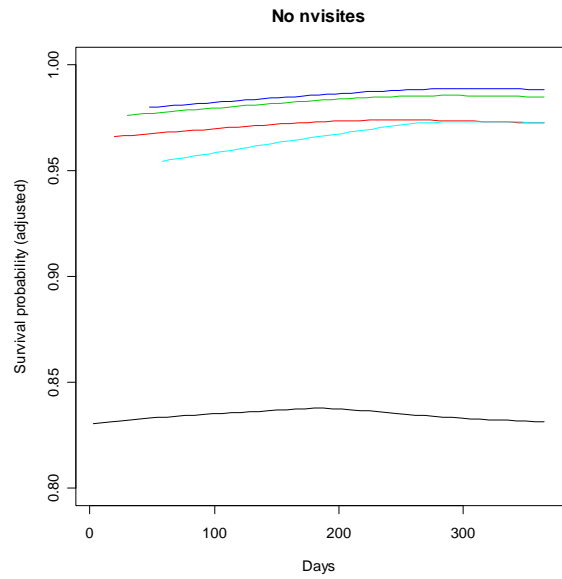
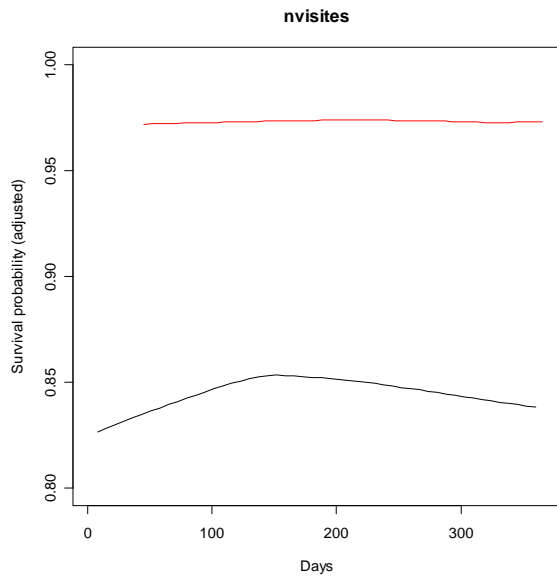


Negre: primer ingrés
 Vermell: segon ingrés
 Verd: tercer ingrés
 Blau: quart ingrés
 Turquesa: cinquè ingrés

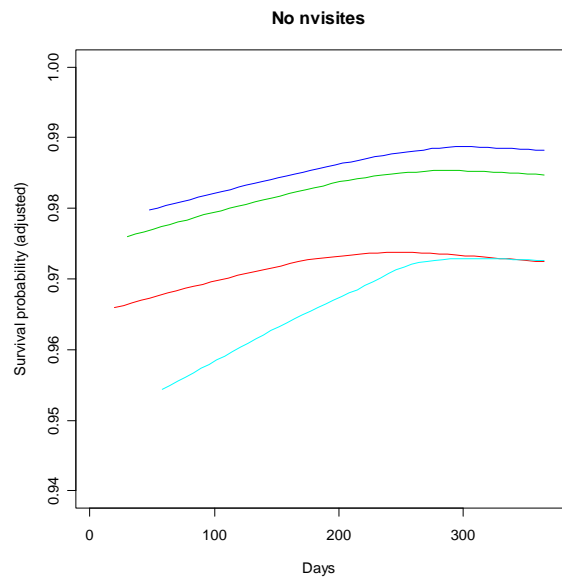
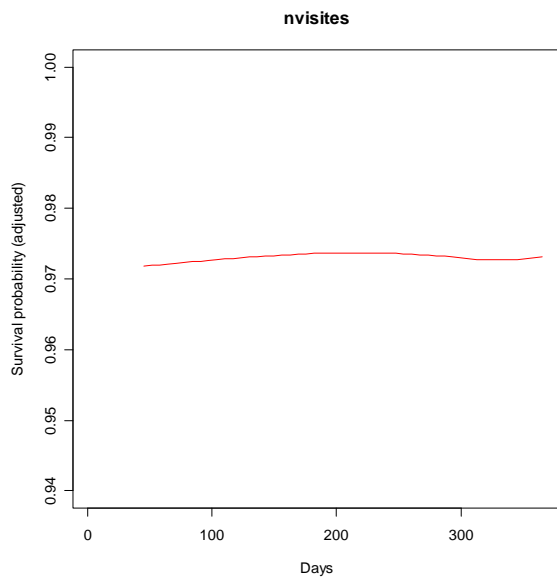


Vermell: segon ingrés
 Verd: tercer ingrés
 Blau: quart ingrés
 Turquesa: cinquè ingrés

Gràfics de la probabilitat de supervivència segons el nombre d'ingressos per nombre de visites realitzades a primària:



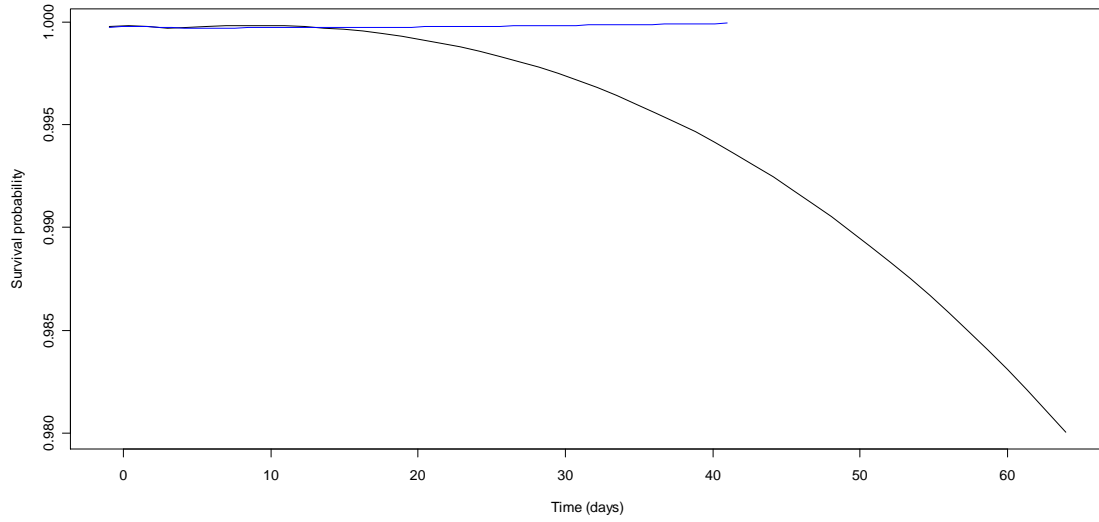
Negre: primer ingrès
 Vermell: segon ingrès
 Verd: tercer ingrès
 Blau: quart ingrès
 Turquesa: cinquè ingrès



Vermell: segon ingrès
 Verd: tercer ingrès
 Blau: quart ingrès
 Turquesa: cinquè ingrès

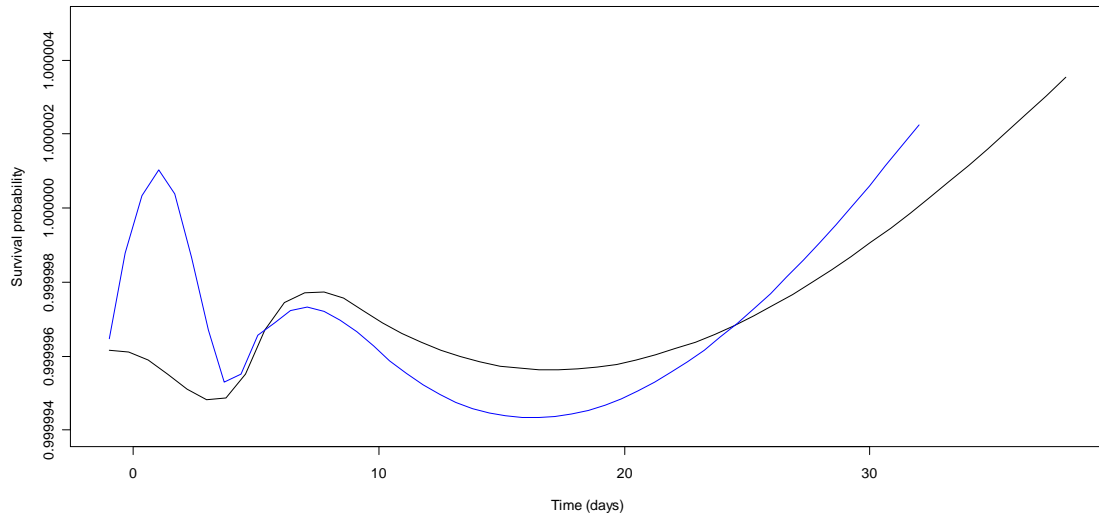
A.VII.2. MODEL AMB EFECTES ESPACIALS

Gràfic de la probabilitat de supervivència per el primer ingrès segons el sexe:



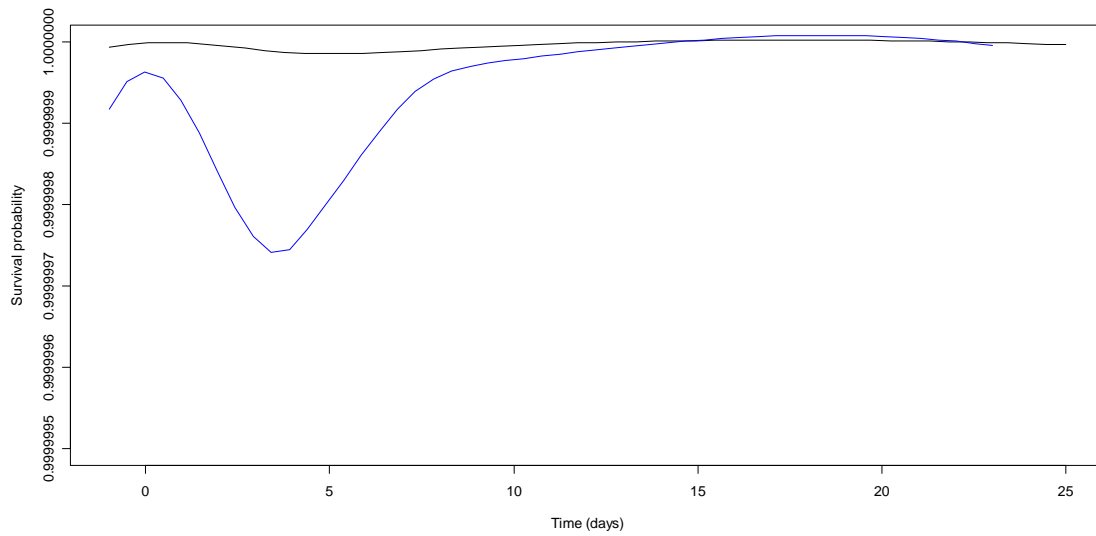
Negre: homes
Blau: dones

Gràfic de la probabilitat de supervivència per el segon ingrès segons el sexe:



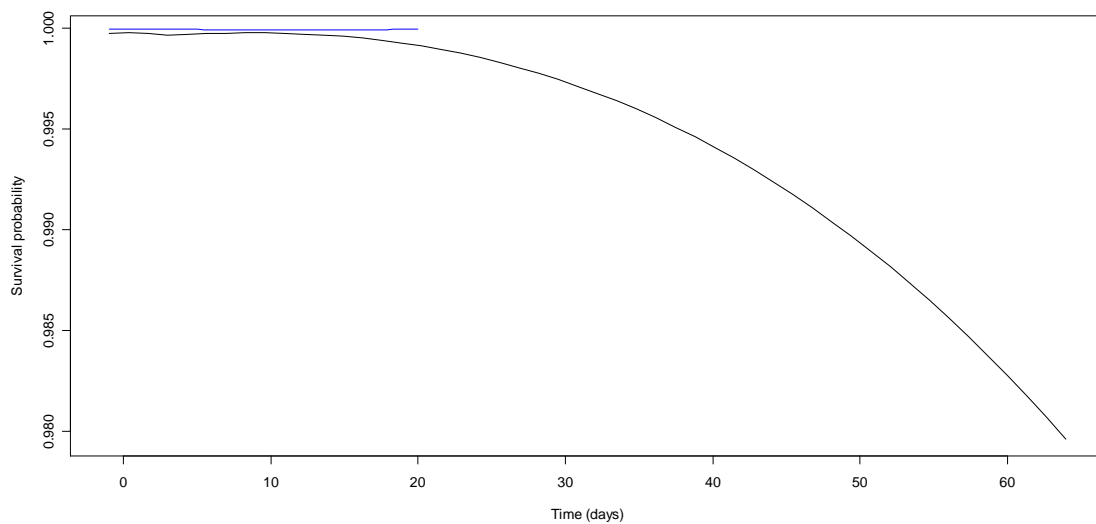
Negre: homes
Blau: dones

Gràfic de la probabilitat de supervivència per el tercer ingrès segons el sexe:



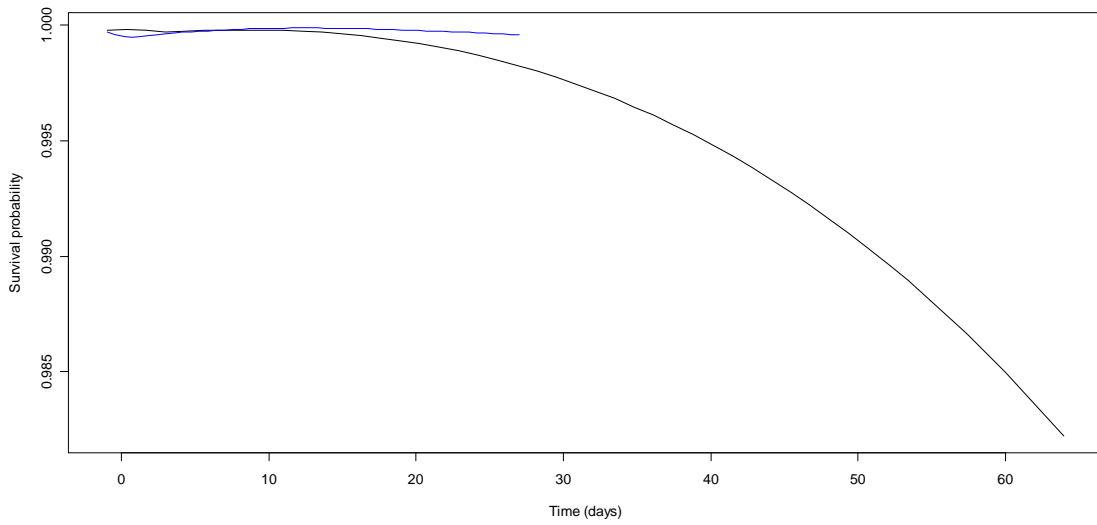
Negre: homes
Blau: dones

Gràfic de la probabilitat de supervivència per el primer ingrès segons presenti o no el problema d'alcohol:



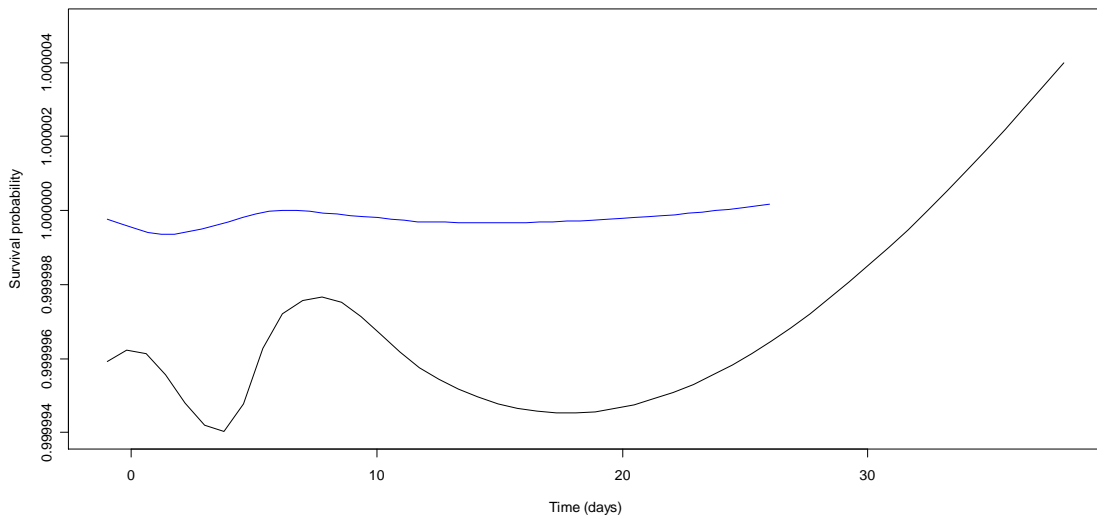
Negre: no es pateix el problema d'alcohol
Blau: si es pateix el problema d'alcohol

Gràfic de la probabilitat de supervivència per el primer ingrès segons presentí o no el problema de mpoc:



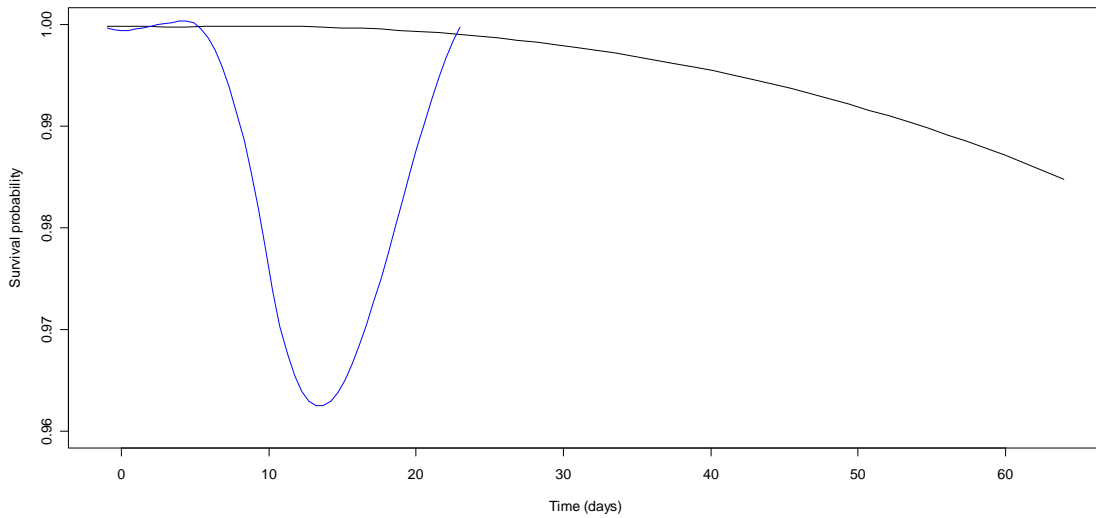
Negre: no es pateix el problema de mpoc
Blau: si es pateix el problema de mpoc

Gràfic de la probabilitat de supervivència per el segon ingrès segons presentí o no el problema de mpoc:



Negre: no es pateix el problema de mpoc
Blau: si es pateix el problema de mpoc

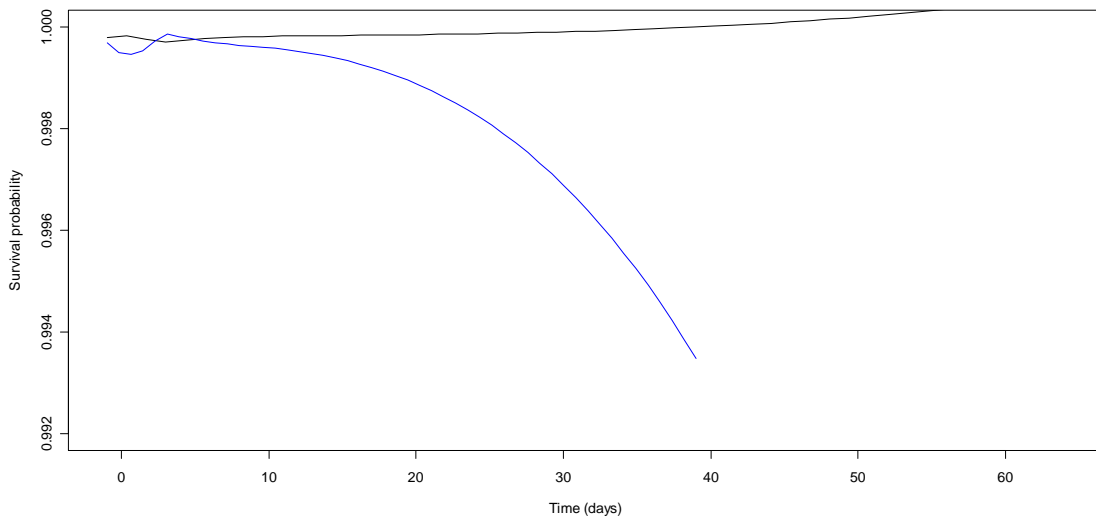
Gràfic de la probabilitat de supervivència per el primer ingrès segons presenti o no el problema de insuficiència cardíaca:



Negre: no es pateix el problema de insuficiència cardíaca

Blau: si es pateix el problema de insuficiència cardíaca

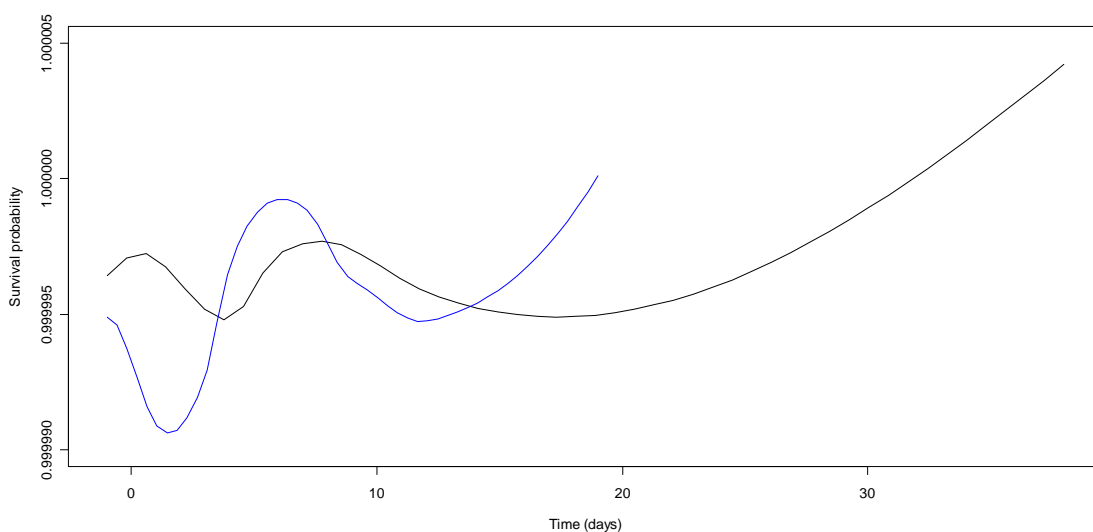
Gràfic de la probabilitat de supervivència per el primer ingrès segons presenti o no el problema de hipertensió arterial:



Negre: no es pateix el problema de hipertensió arterial

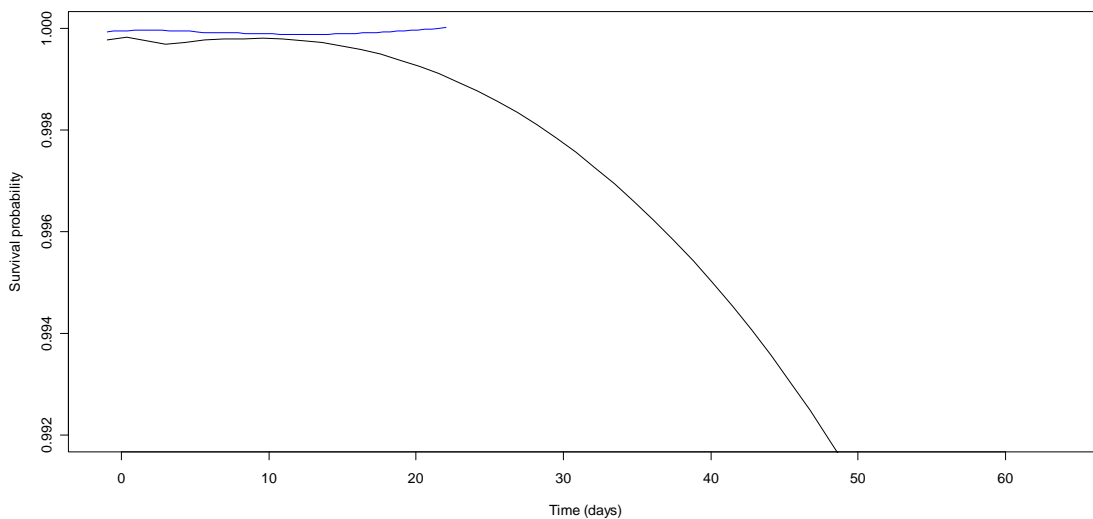
Blau: si es pateix el problema de hipertensió arterial

Gràfic de la probabilitat de supervivència per el segon ingrès segons presenti o no el problema de hipertensió arterial:



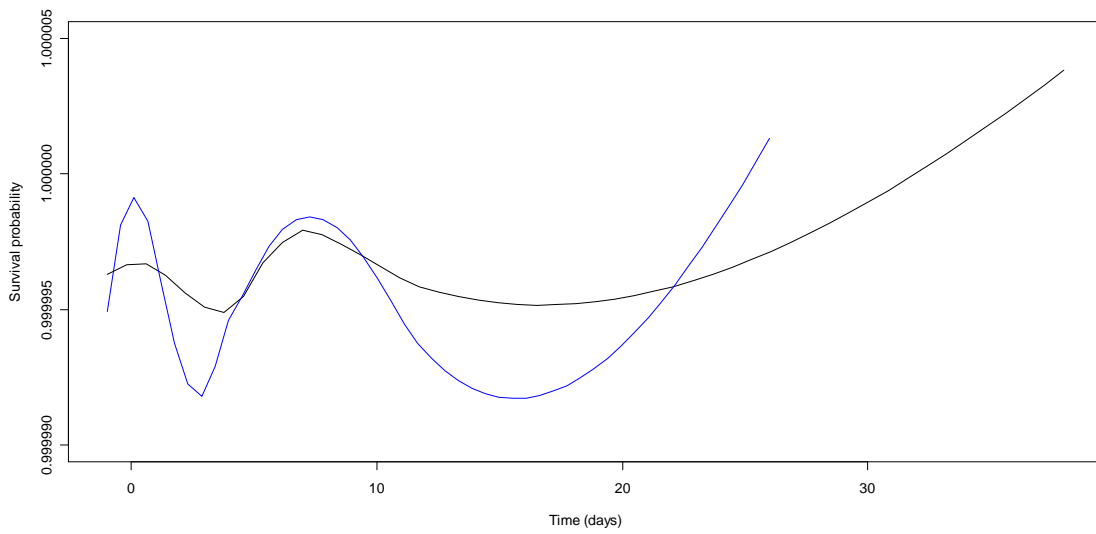
Negre: no es pateix el problema de hipertensió arterial
Blau: si es pateix el problema de hipertensió arterial

Gràfic de la probabilitat de supervivència per el primer ingrès segons presenti o no el problema de dislipèmia:



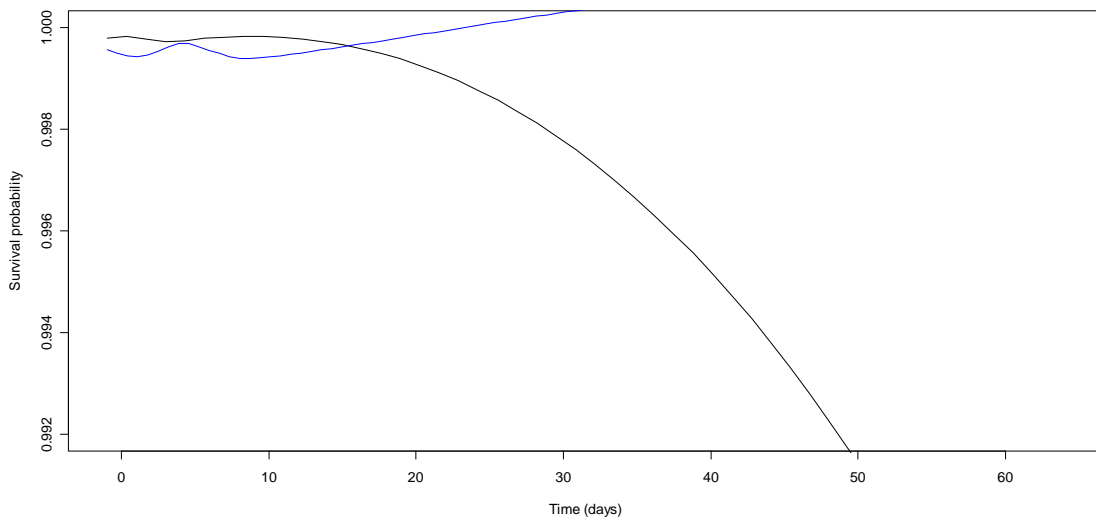
Negre: no es pateix el problema de dislipèmia
Blau: si es pateix el problema de dislipèmia

Gràfic de la probabilitat de supervivència per el segon ingrès segons presenti o no el problema de dislipèmia:



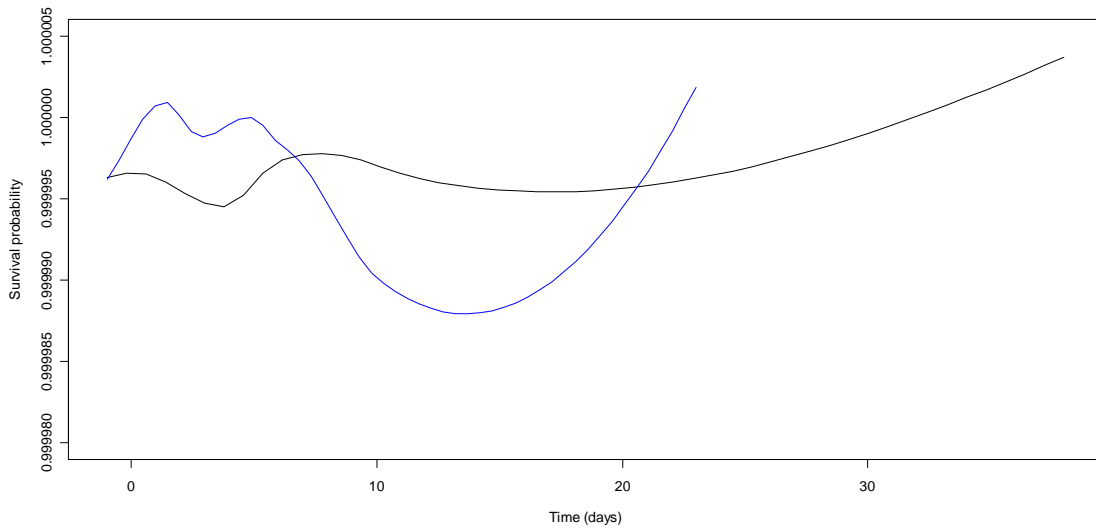
Negre: no es pateix el problema de dislipèmia
 Blau: si es pateix el problema de dislipèmia

Gràfic de la probabilitat de supervivència per el primer ingrès segons presenti o no el problema de diabetis mellitus 2:



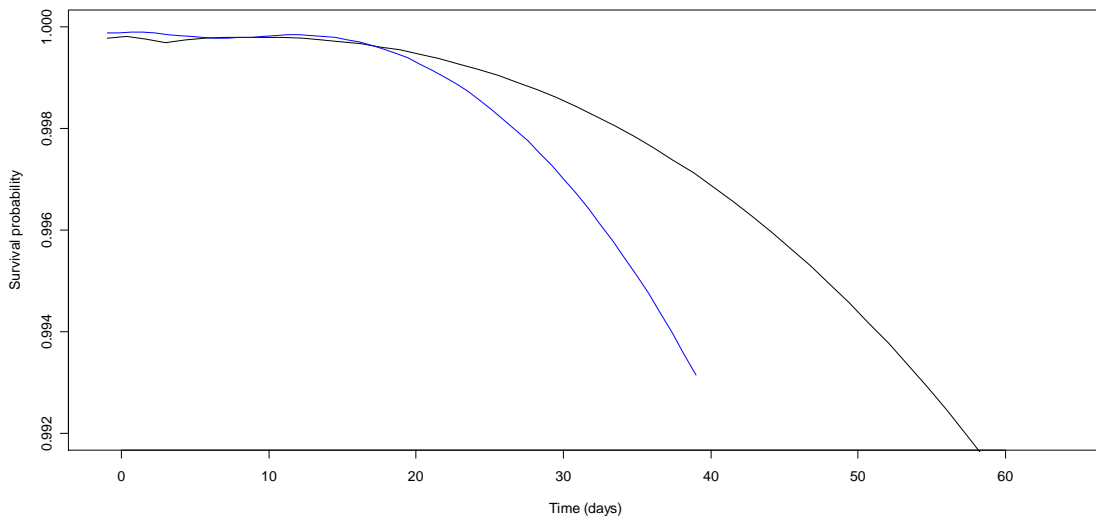
Negre: no es pateix el problema de dm2
 Blau: si es pateix el problema de dm2

Gràfic de la probabilitat de supervivència per el segon ingrès segons presenti o no el problema de diabetis mellitus 2:



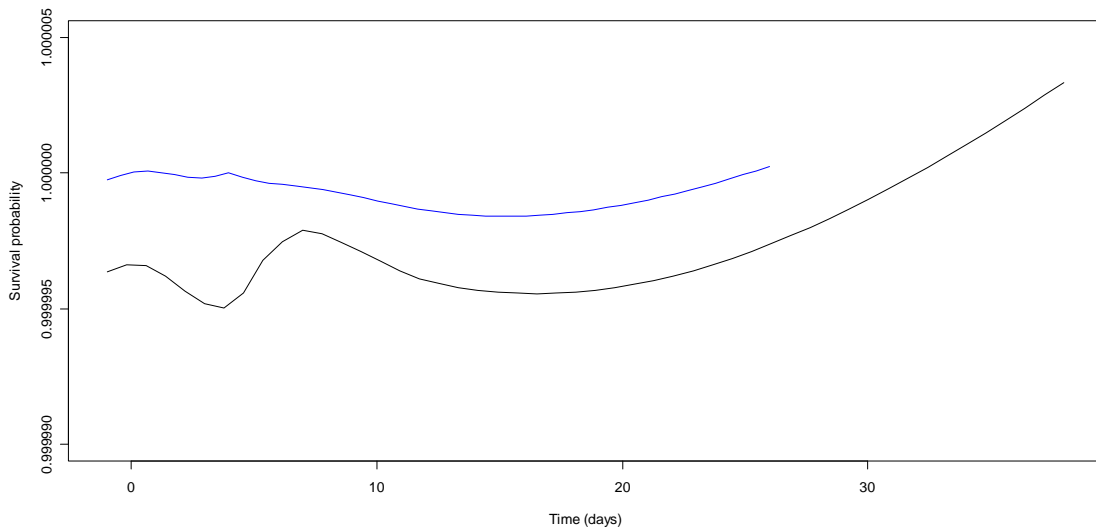
Negre: no es pateix el problema de dm2
Blau: si es pateix el problema de dm2

Gràfic de la probabilitat de supervivència per el primer ingrès segons presenti o no el problema de tabac:



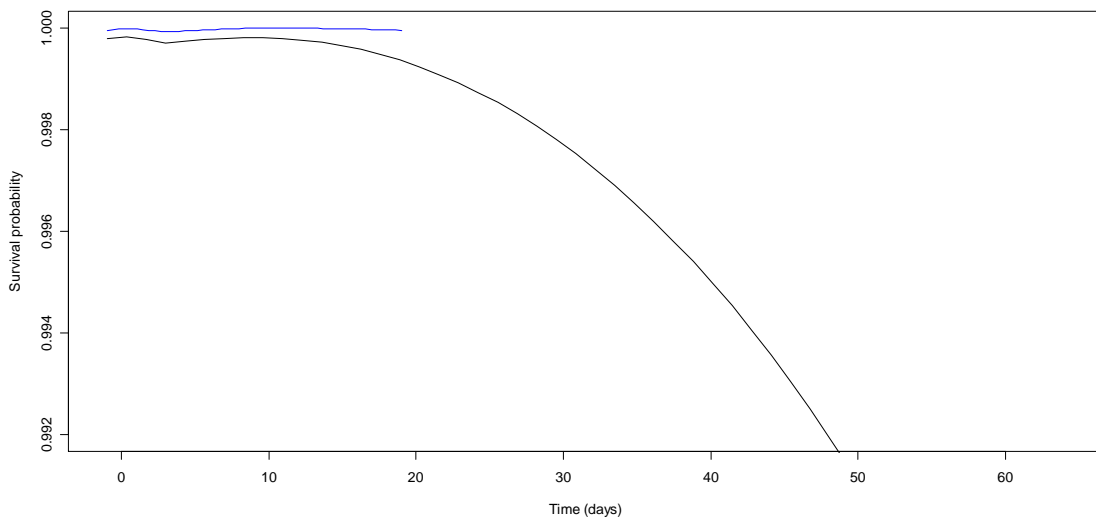
Negre: no es pateix el problema de tabac
Blau: si es pateix el problema de tabac

Gràfic de la probabilitat de supervivència per el segon ingrès segons presentí o no el problema de tabac:



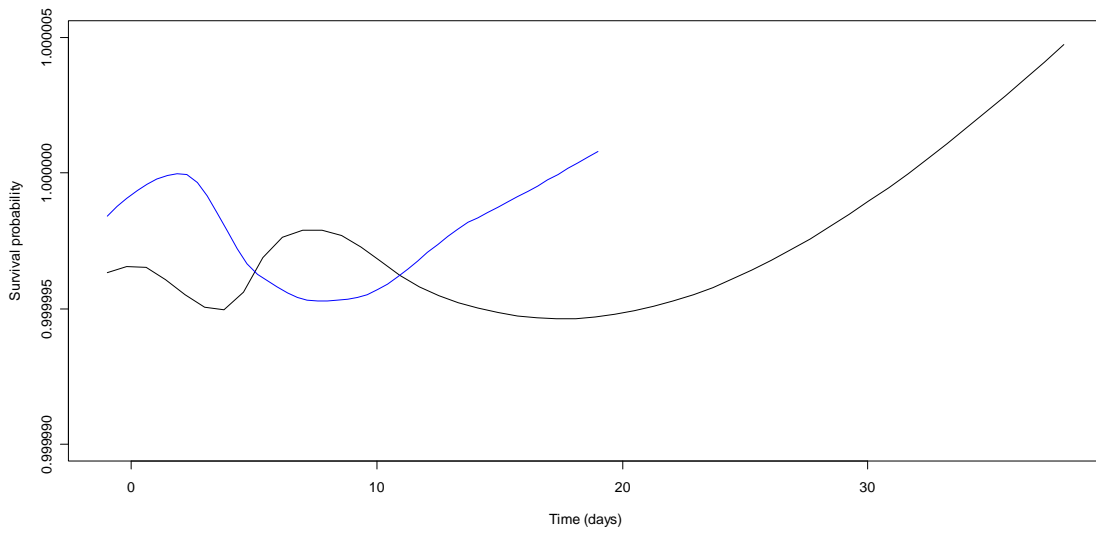
Negre: no es pateix el problema de tabac
Blau: si es pateix el problema de tabac

Gràfic de la probabilitat de supervivència per el primer ingrès segons presentí o no el problema de extabac:



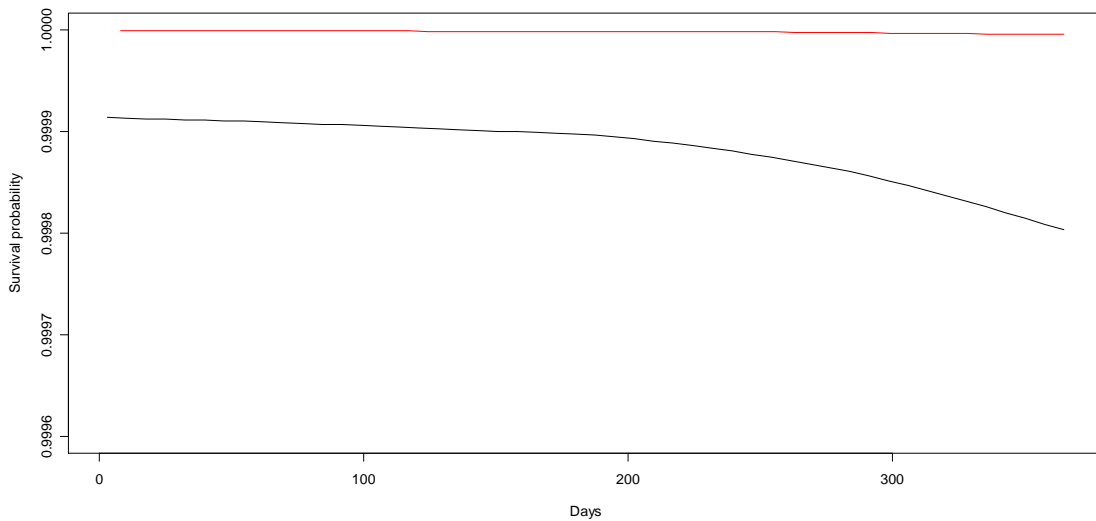
Negre: no es pateix el problema de extabac
Blau: si es pateix el problema de extabac

Gràfic de la probabilitat de supervivència per el segon ingrès segons presenti o no el problema de extabac:



Negre: no es pateix el problema de extabac
Blau: si es pateix el problema de extabac

Gràfic de la probabilitat de supervivència segons si ha realitzat una visita a primària o no:



Negre: no ha realitzat cap visita
Vermell: ha realitzat una visita

ANNEX VIII: RESULTATS OBTINGUTS EN L'APLICACIÓ DEL MÈTODE DE GOLDSTEIN

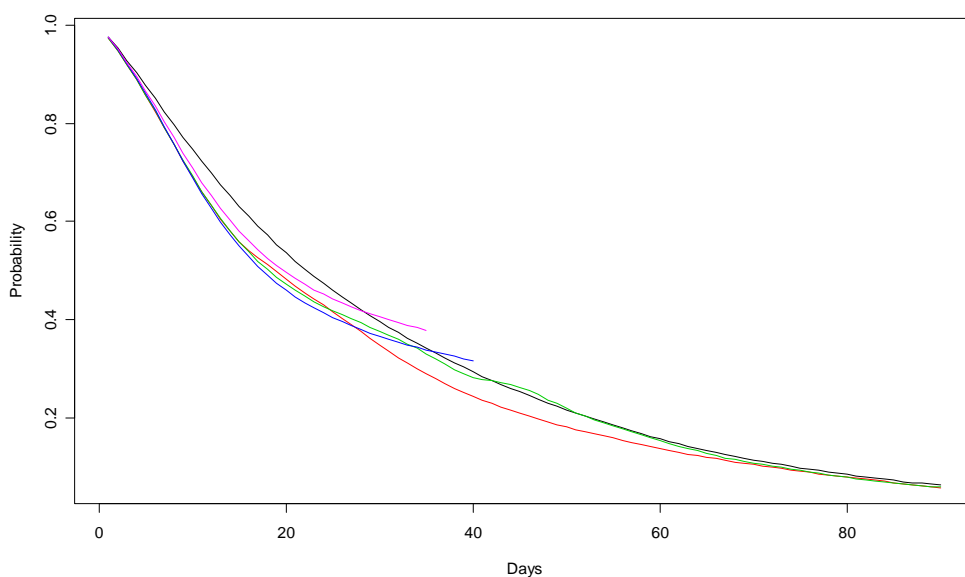
A.VII.1. MODEL SENSE EFECTES ESPACIALS

Inference for Bugs model at "Modelo.txt", fit using winbugs,
2 chains, each with 10000 iterations (first 1000 discarded), n.thin = 30
n.sims = 600 iterations saved

	mean	sd	2.5%	25%	50%	75%	97.5%	Rhat	n.eff
b0	-3.63346	0.17424	-3.84600	-3.81000	-3.66350	-3.49700	-3.20073	2.00702	4
b1	-0.10895	0.06147	-0.21623	-0.15170	-0.11270	-0.07117	0.01521	1.08372	25
b21	-0.05005	0.09597	-0.24813	-0.11285	-0.04037	0.01869	0.11223	1.42646	7
b22	-0.09628	0.09773	-0.30428	-0.16030	-0.08791	-0.02996	0.08002	1.35854	8
b23	-0.12143	0.09516	-0.31447	-0.18020	-0.11655	-0.05356	0.04657	1.42850	7
b24	-0.08803	0.09711	-0.28936	-0.15552	-0.08113	-0.01837	0.07979	1.38346	7
c11	0.69442	2.04265	-2.36920	-1.11525	0.56805	2.58575	3.71325	7.17245	2
c12	-1.82834	0.19297	-2.22005	-1.95125	-1.82750	-1.69100	-1.47578	1.40717	7
c13	1.34805	0.32526	0.64900	1.12750	1.39500	1.59250	1.88217	1.67729	5
c14	1.71592	0.16737	1.36295	1.61075	1.72800	1.82325	2.04200	1.04625	37
b3	-0.02520	0.18567	-0.40211	-0.15337	-0.01785	0.10828	0.31026	1.04674	42
b4	-0.00085	0.08152	-0.17651	-0.05869	0.00366	0.05464	0.15350	1.01886	100
b5	0.06561	0.19960	-0.33263	-0.07182	0.06782	0.20223	0.46091	1.01011	210
b6	-0.01070	0.10191	-0.19750	-0.08739	-0.00701	0.05611	0.19631	1.00572	450
b7	0.01138	0.12070	-0.23095	-0.07036	0.01950	0.09502	0.22882	1.00046	600
b8	-0.03250	0.11532	-0.25940	-0.11292	-0.03299	0.04625	0.18772	1.00795	190
b9	-0.17717	0.12363	-0.41880	-0.26922	-0.17495	-0.09329	0.07085	1.01337	110
b10	0.03201	0.19504	-0.32279	-0.10493	0.04344	0.17357	0.39353	1.00090	600
b11	-0.71021	2.04155	-3.70635	-2.60100	-0.59643	1.08825	2.35907	7.17045	2
sigma.beta0	0.02645	0.01492	0.00345	0.01480	0.02489	0.03758	0.05704	1.04588	600

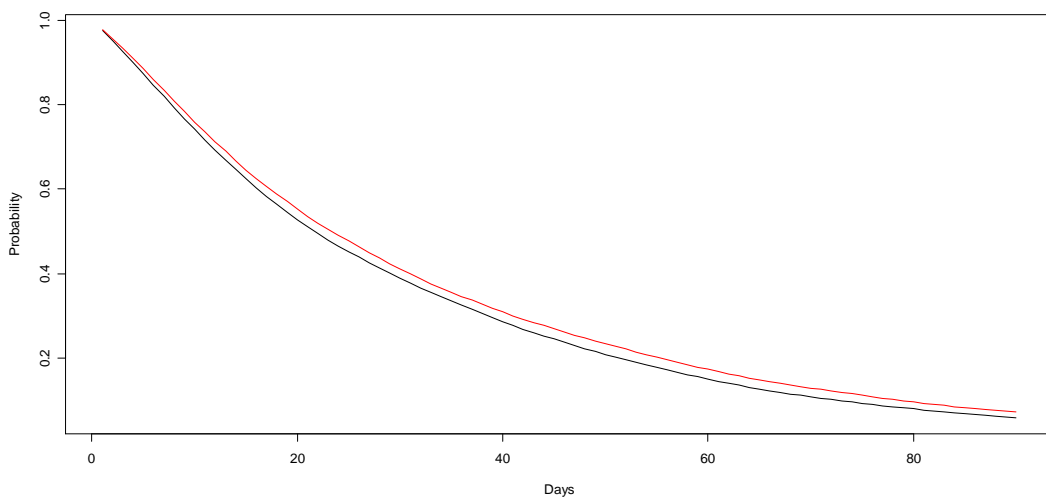
For each parameter, n.eff is a crude measure of effective sample size, and Rhat is the potential scale reduction factor (at convergence, Rhat=1).

Gràfic de la probabilitat de supervivència per cada ingrés.



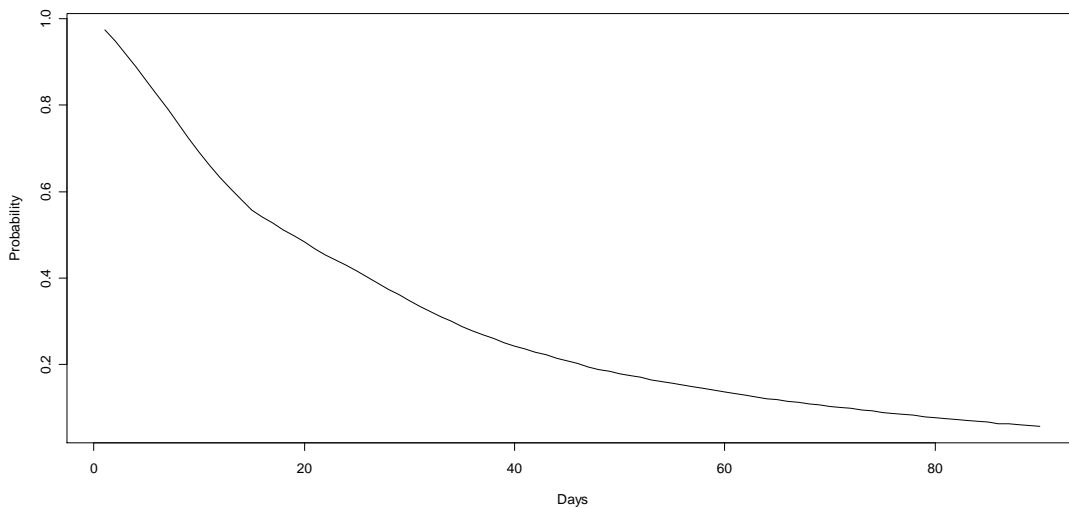
Negre: primer ingrés Vermell: segon ingrés Verd: tercer ingrés Blau: quart ingrés Rosa: cinquè ingrés

Gràfic de la probabilitat de supervivència per el primer ingrès segons el sexe:



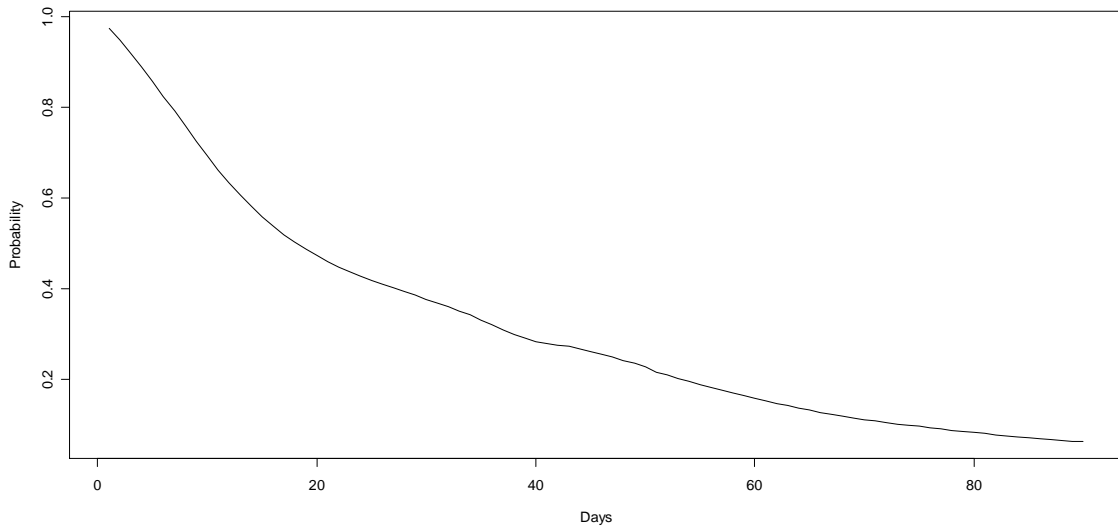
Negre: home
Vermell: dona

Gràfic de la probabilitat de supervivència per el segon ingrès segons el sexe:



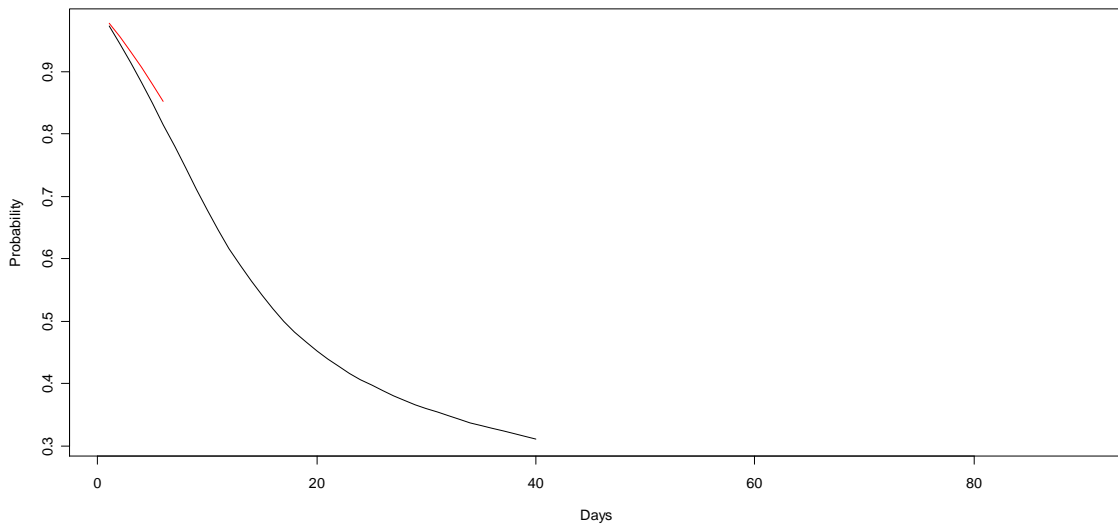
Negre: home
Vermell: dona

Gràfic de la probabilitat de supervivència per el tercer ingrès segons el sexe:



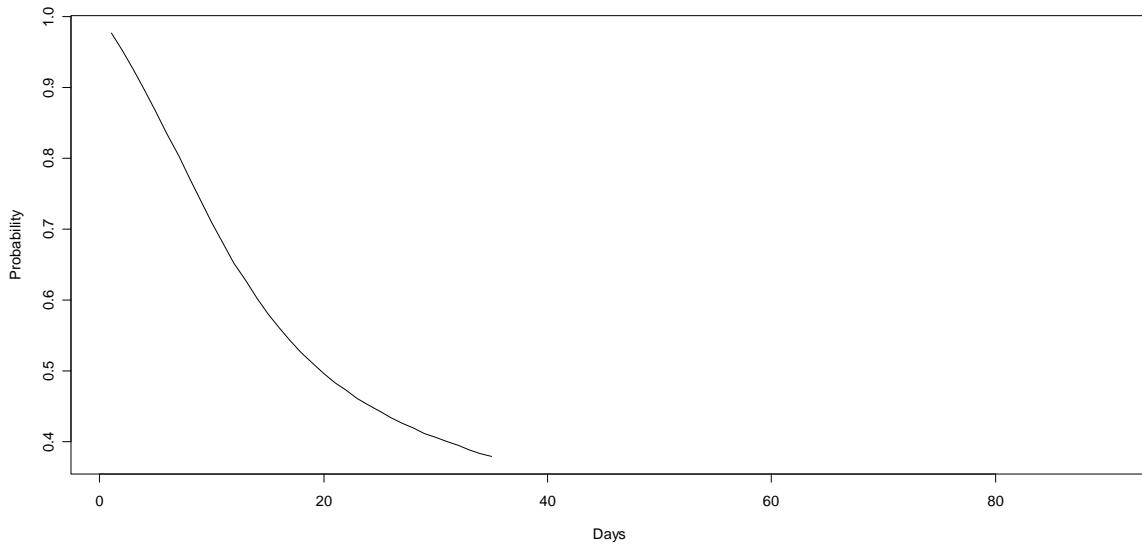
Negre: home
Vermell: dona

Gràfic de la probabilitat de supervivència per el quart ingrès segons el sexe:



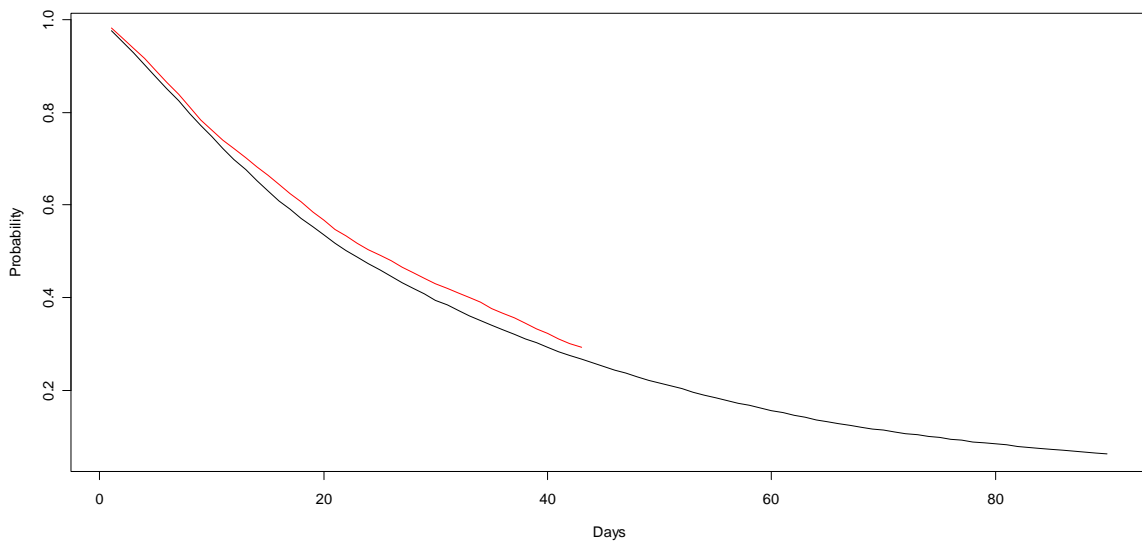
Negre: home
Vermell: dona

Gràfic de la probabilitat de supervivència per el cinquè ingrès segons el sexe:



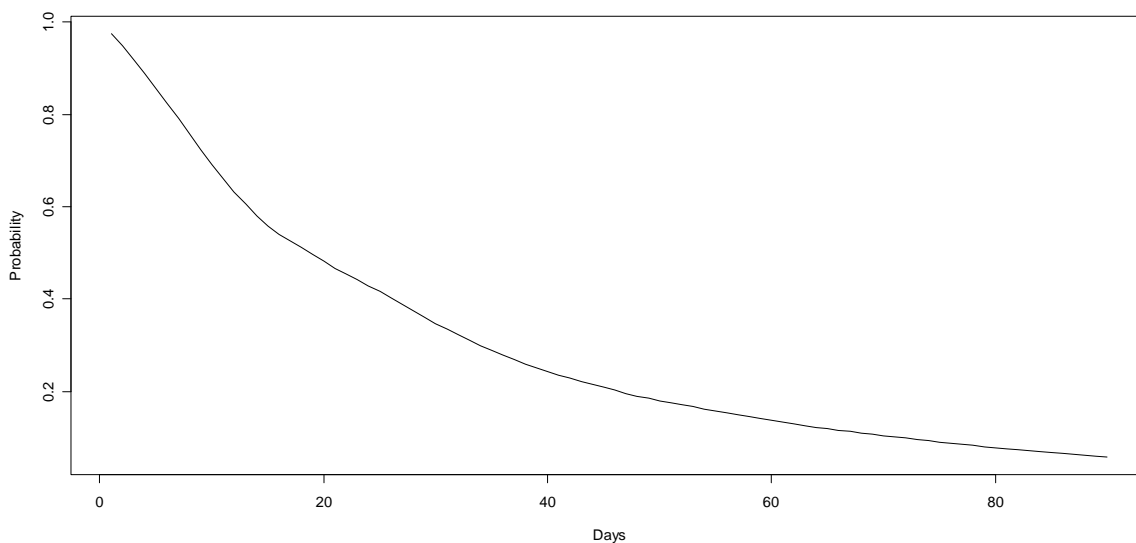
Negre: home
Vermell: dona

Gràfic de la probabilitat de supervivència per el primer ingrès segons presenti o no el problema d'alcohol:



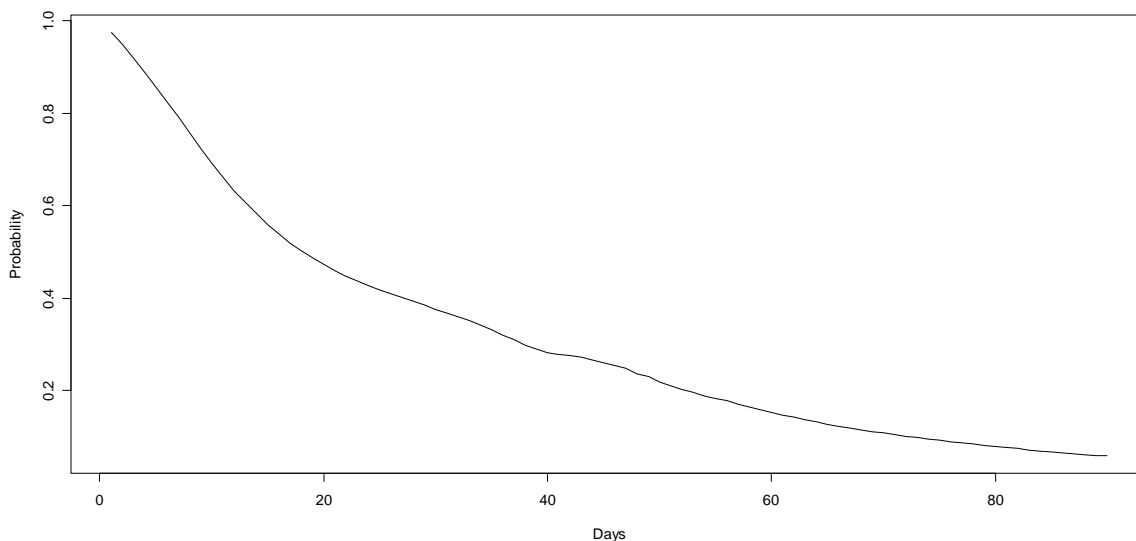
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el segon ingrès segons presenti o no el problema d'alcohol:



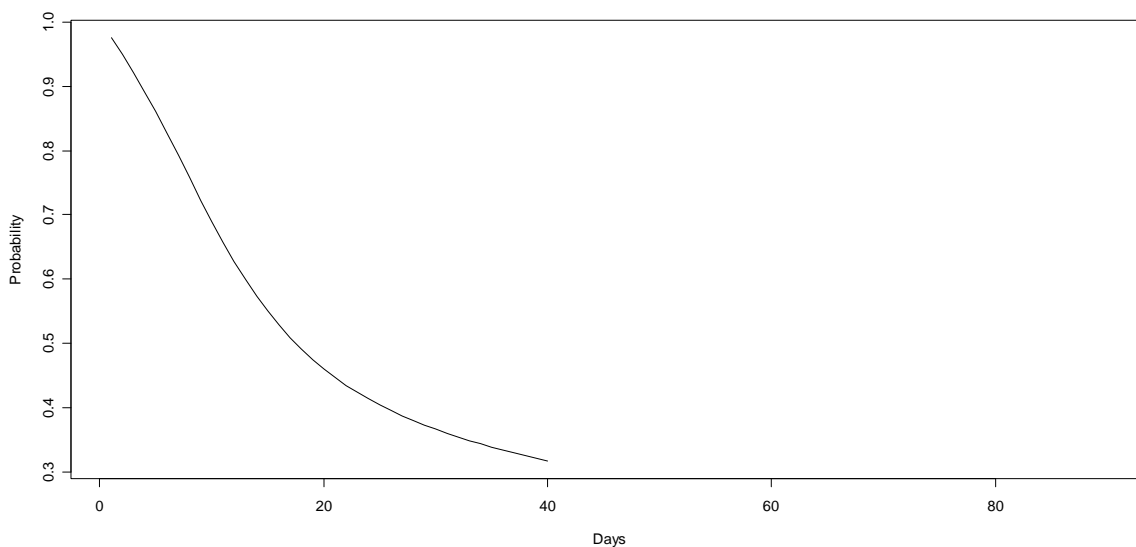
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el tercer ingrès segons presenti o no el problema d'alcohol:



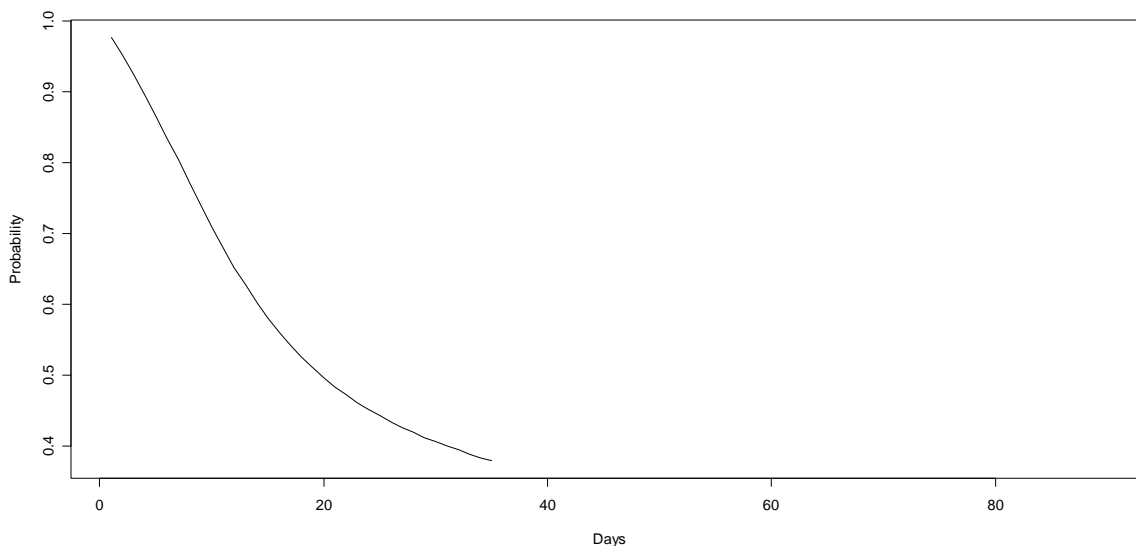
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el quart ingrès segons presenti o no el problema d'alcohol:



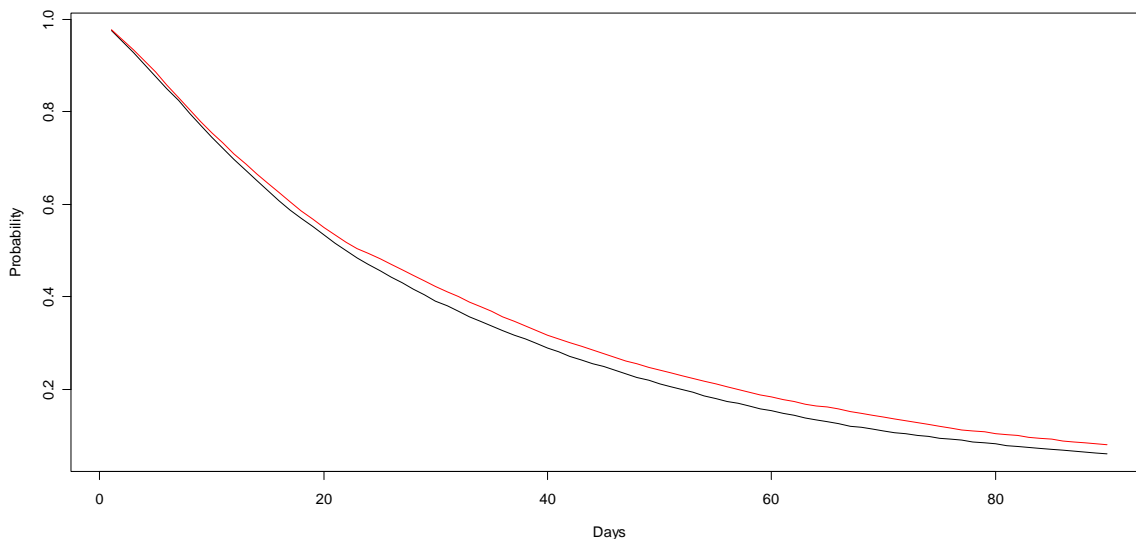
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el cinquè ingrès segons presenti o no el problema d'alcohol:



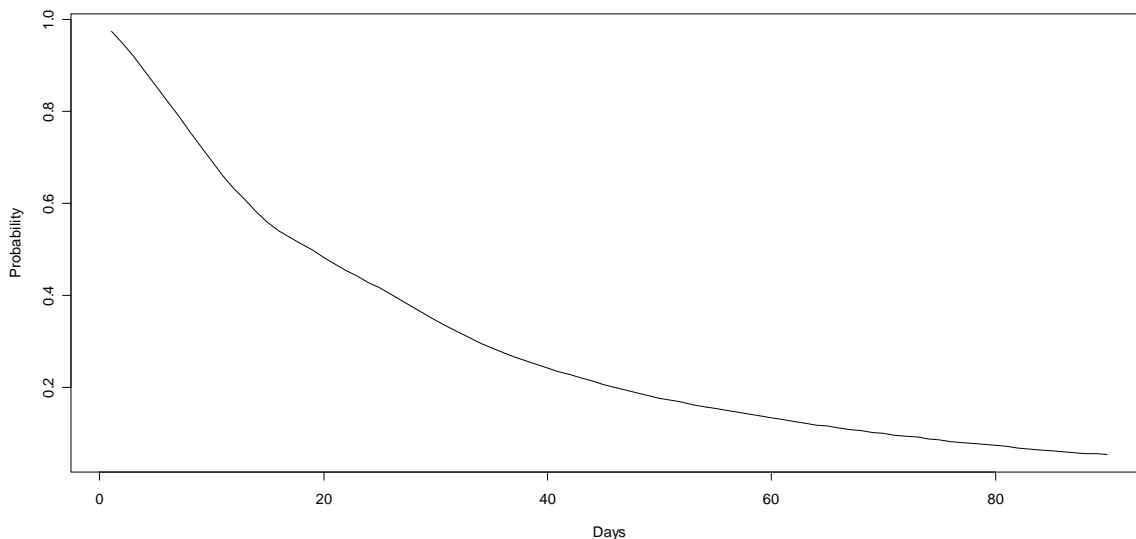
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el primer ingrès segons presenti o no el problema de MPOC:



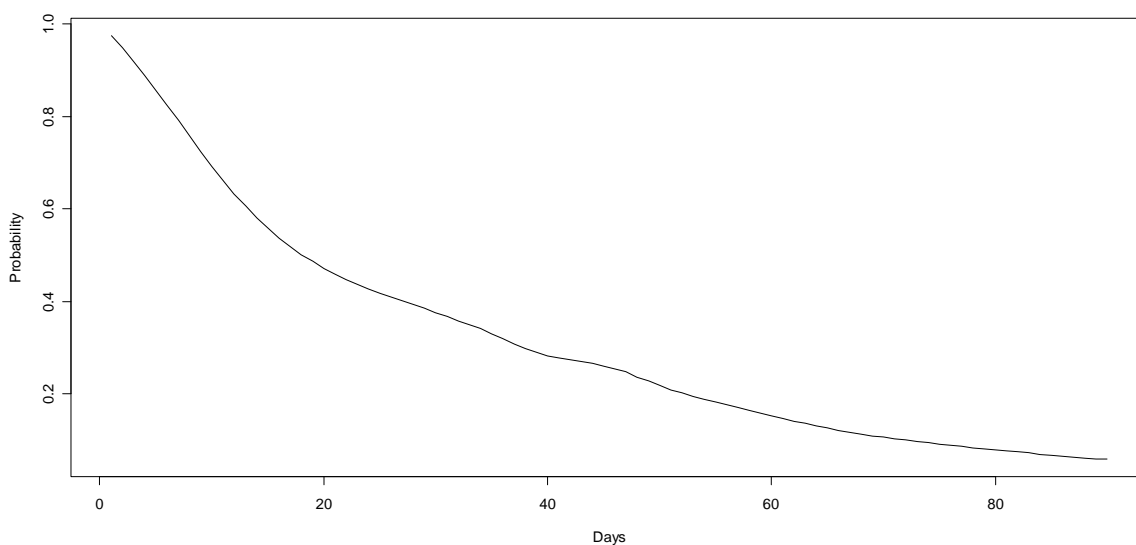
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el segon ingrès segons presenti o no el problema de MPOC:



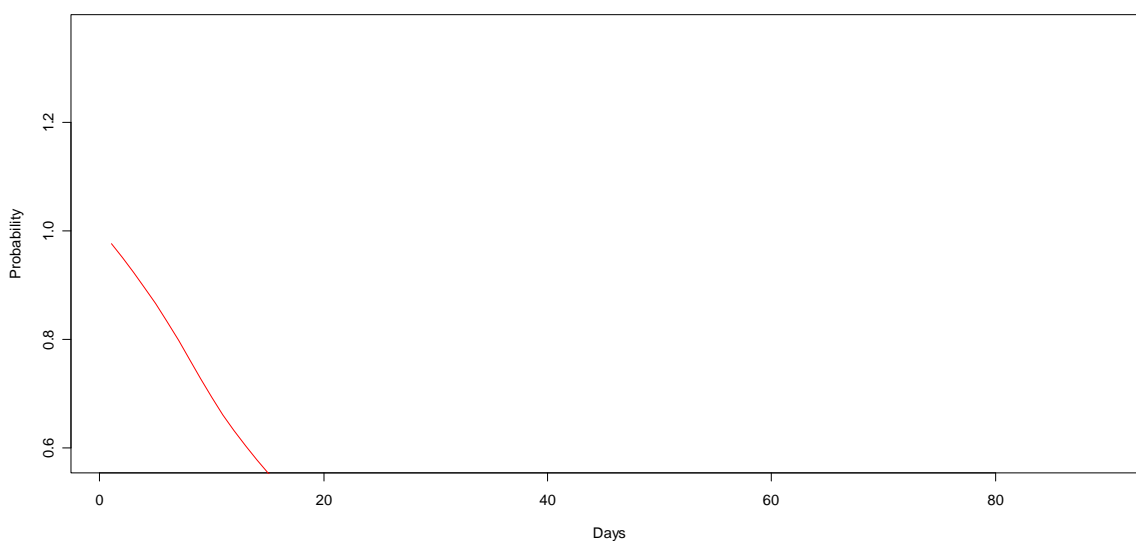
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el tercer ingrès segons presenti o no el problema de MPOC:



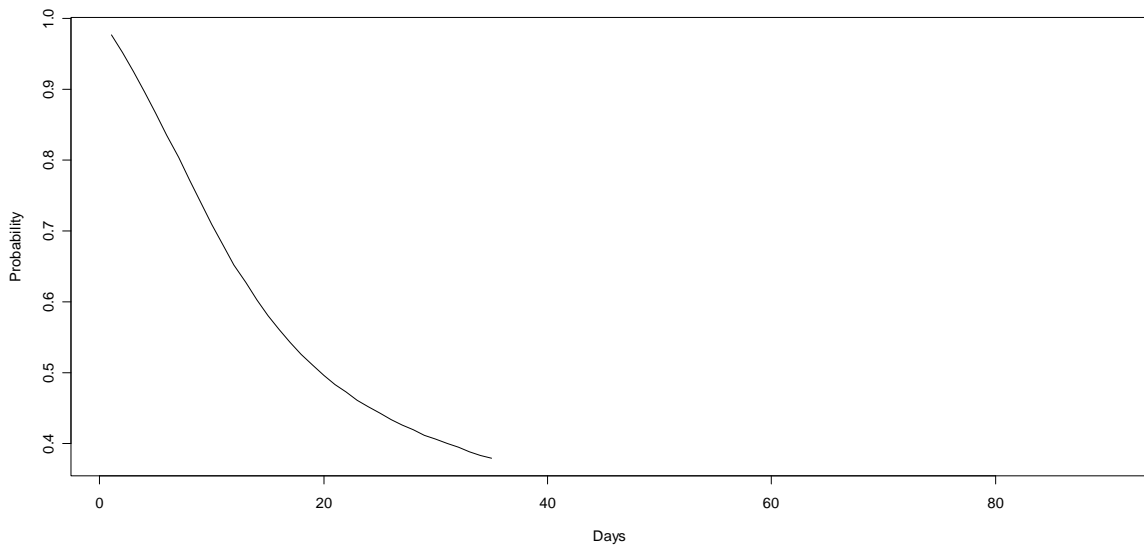
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el quart ingrès segons presenti o no el problema de MPOC:



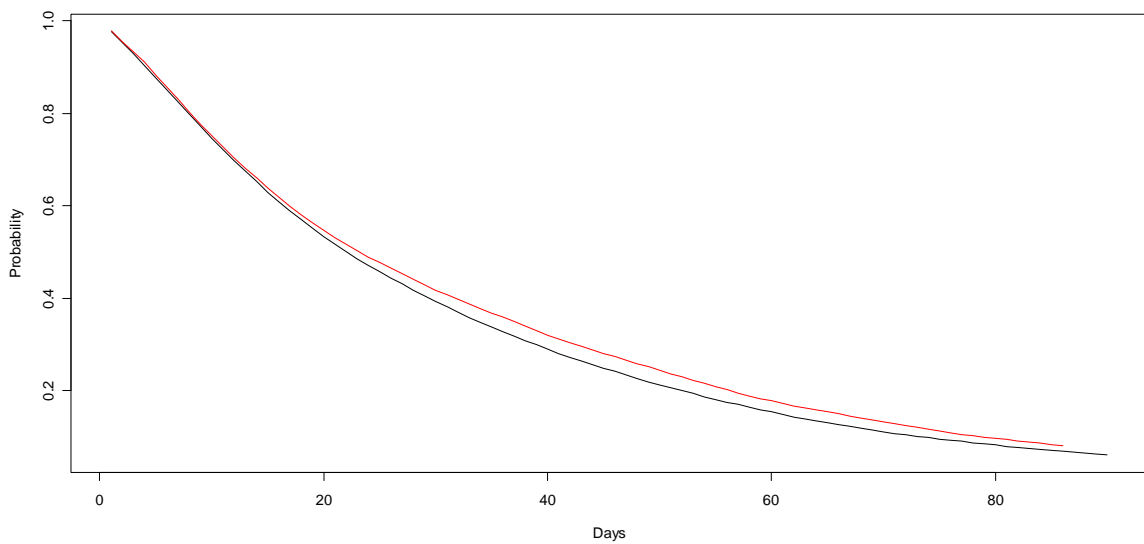
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el cinquè ingrès segons presenti o no el problema de MPOC:



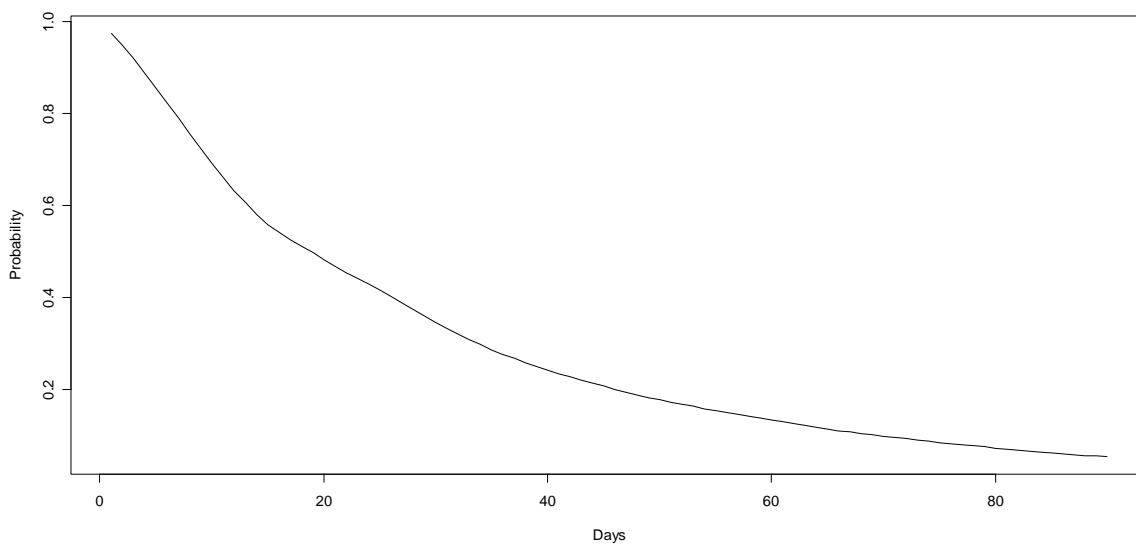
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el primer ingrès segons presenti o no el problema de hipertensió arterial:



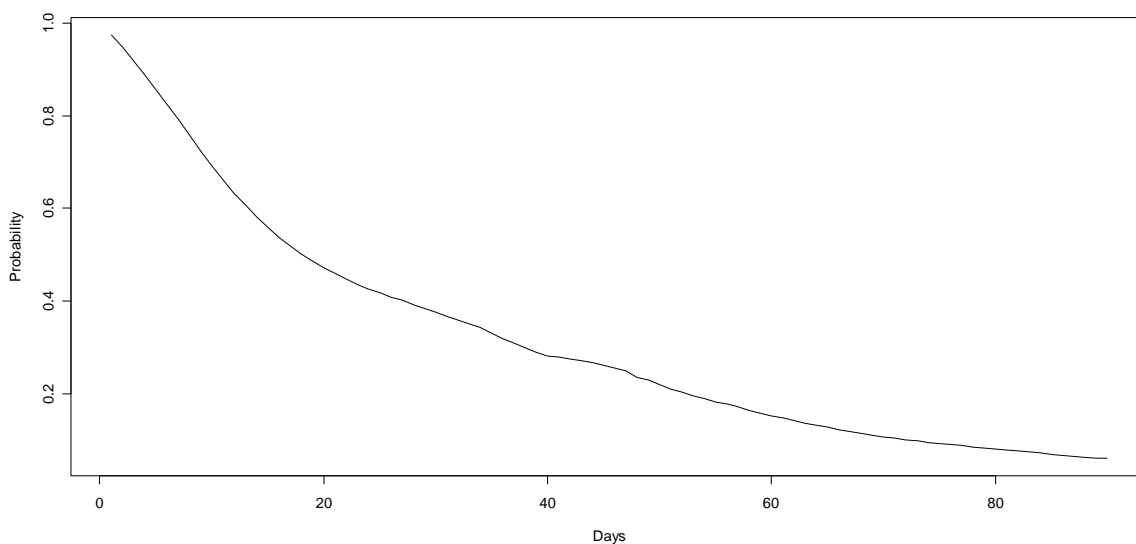
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el segon ingrès segons presenti o no el problema de hipertensió arterial:



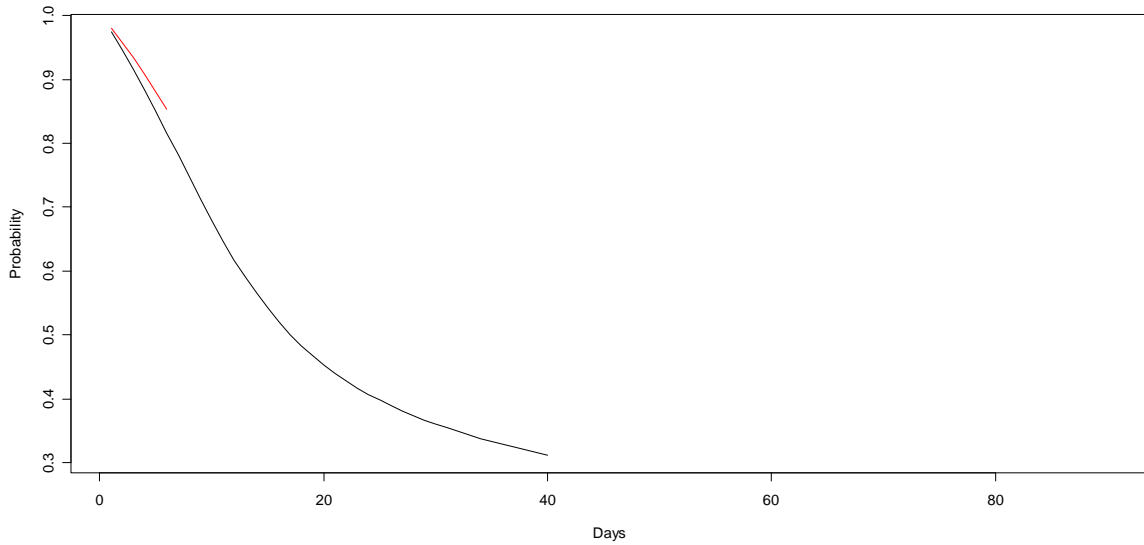
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el tercer ingrès segons presenti o no el problema de hipertensió arterial:



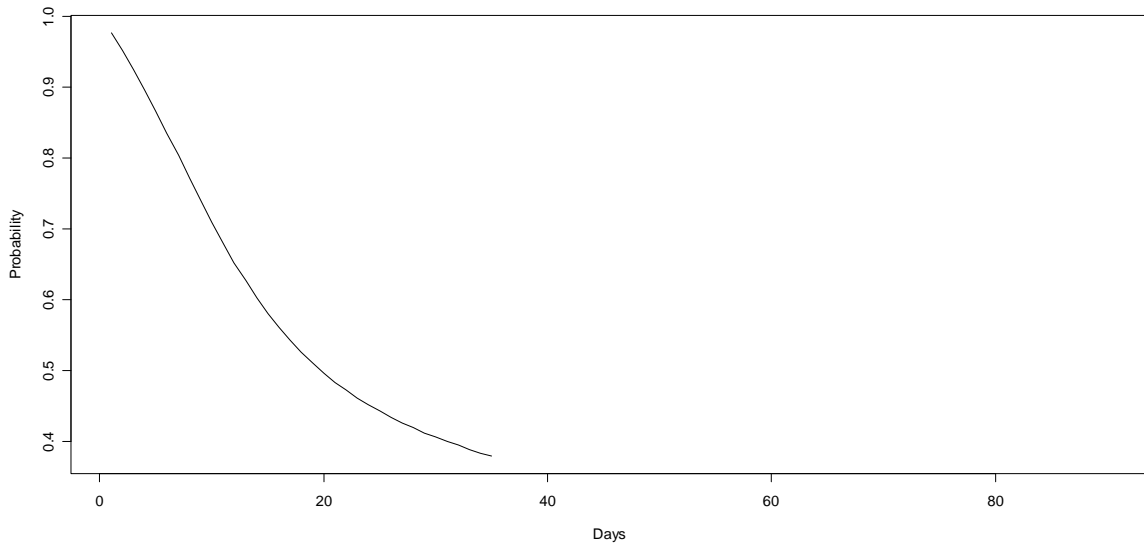
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el quart ingrès segons presenti o no el problema de hipertensió arterial:



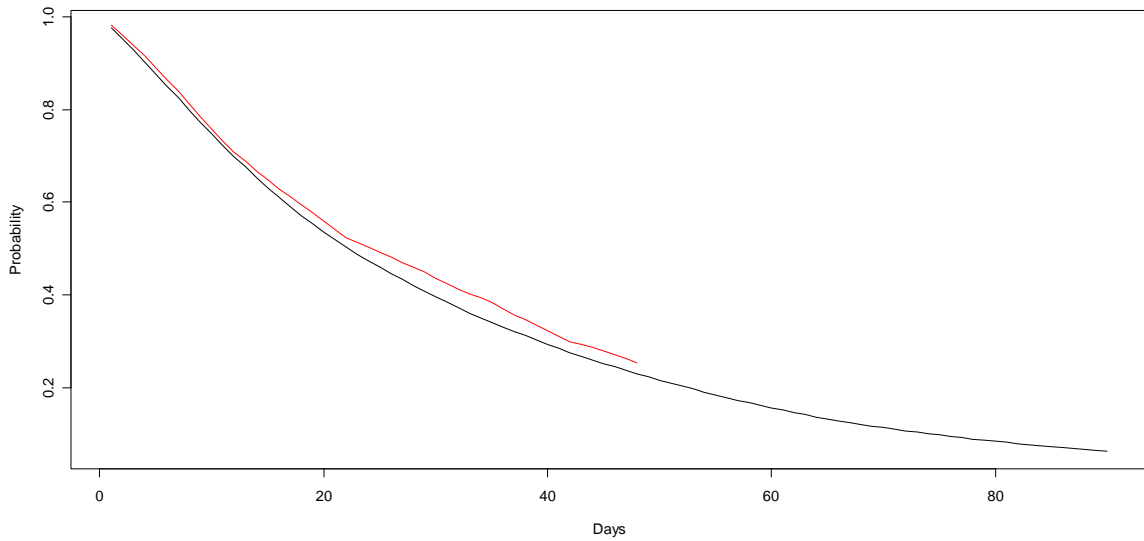
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el cinquè ingrès segons presenti o no el problema de hipertensió arterial:



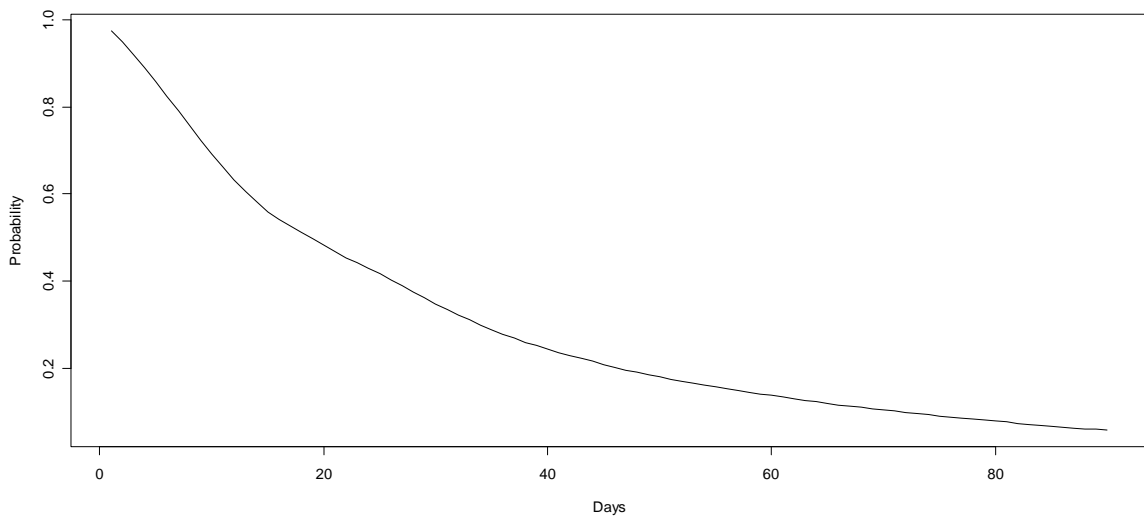
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el primer ingrès segons presenti o no el problema de insuficiència cardíaca:



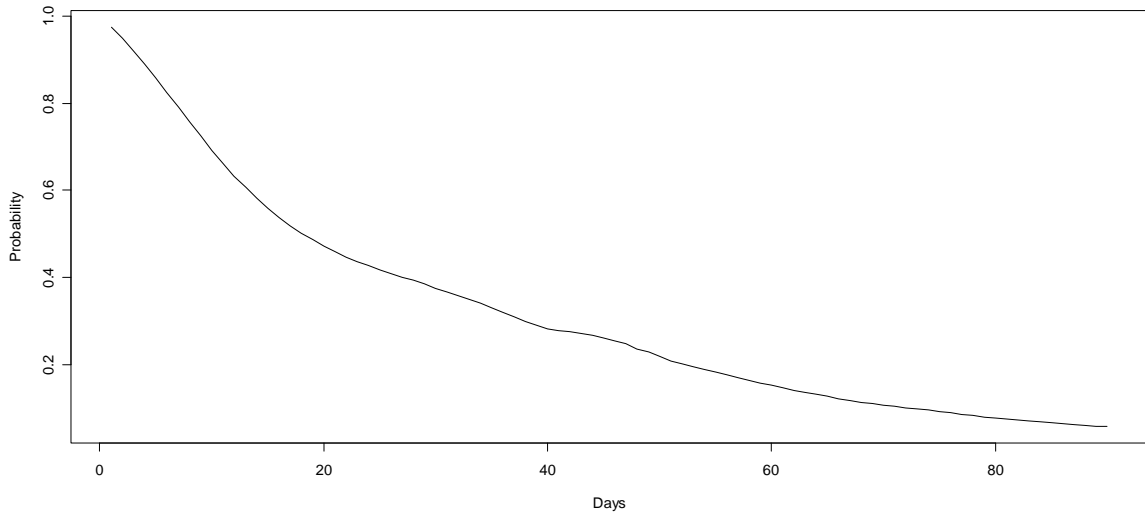
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el segon ingrès segons presenti o no el problema de insuficiència cardíaca:



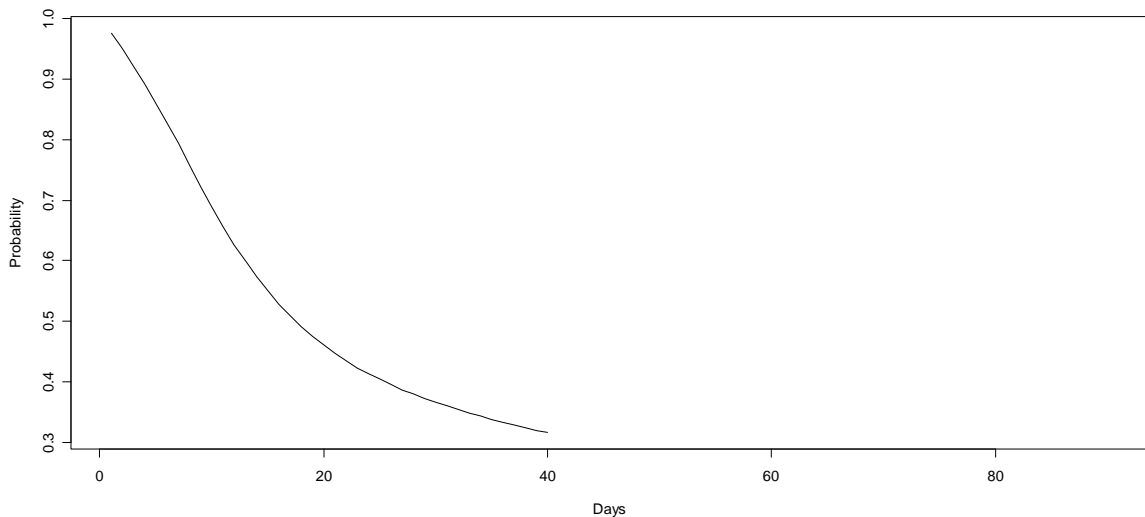
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el tercer ingrès segons presenti o no el problema de insuficiència cardíaca:



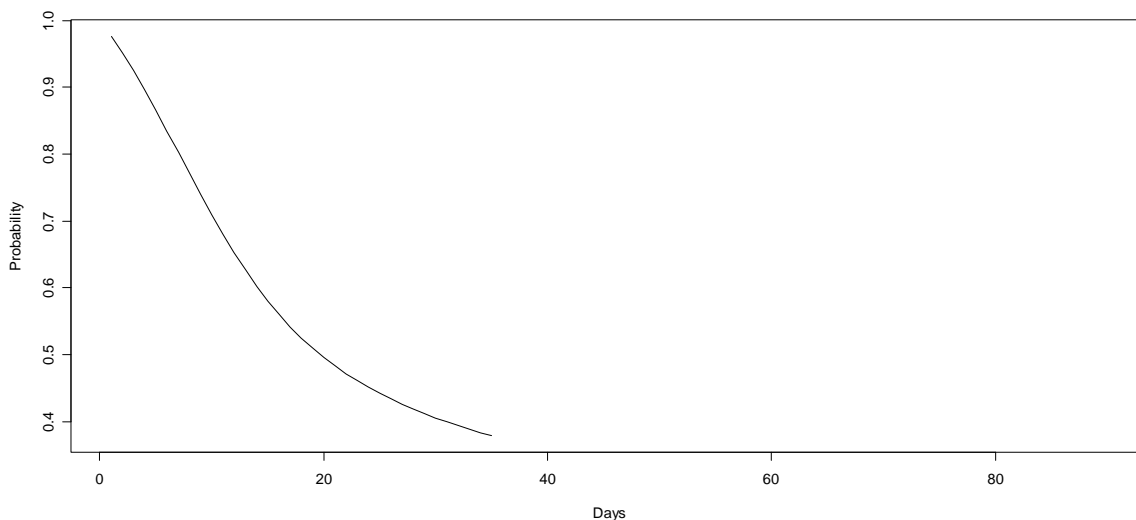
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el quart ingrès segons presenti o no el problema de insuficiència cardíaca:



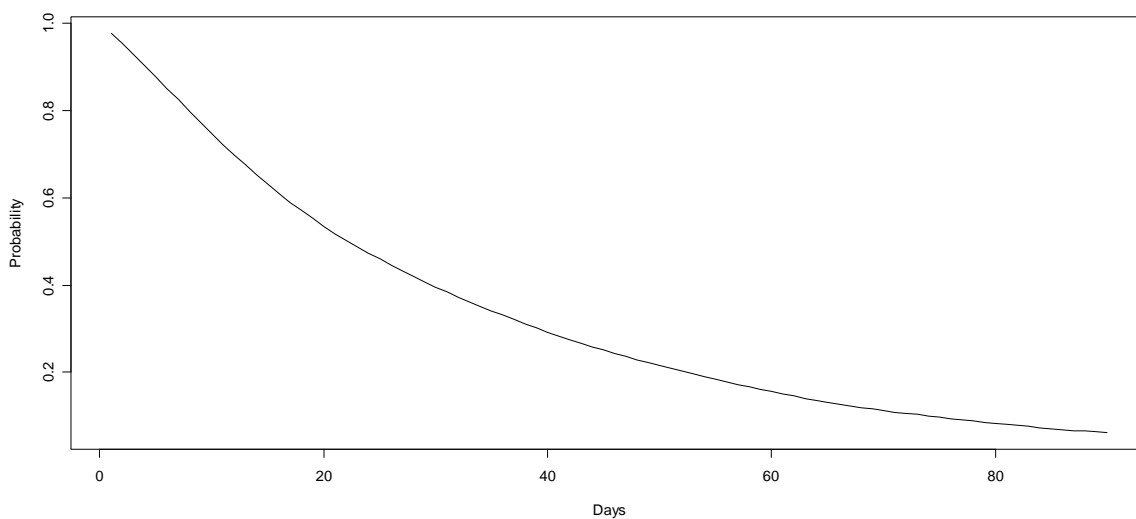
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el cinquè ingrès segons presenti o no el problema de insuficiència cardíaca:



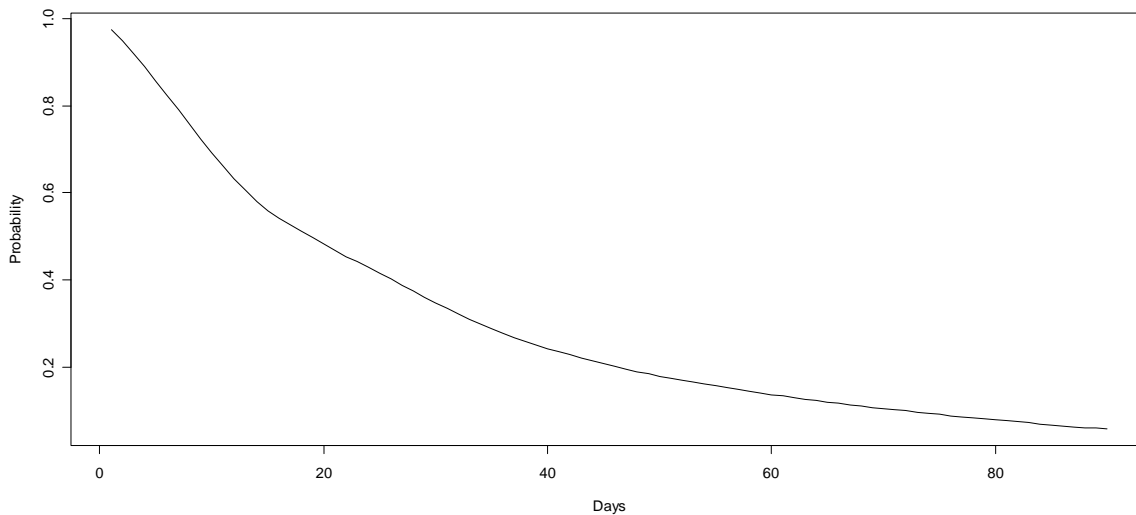
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el primer ingrès segons presenti o no el problema de dislipèmia:



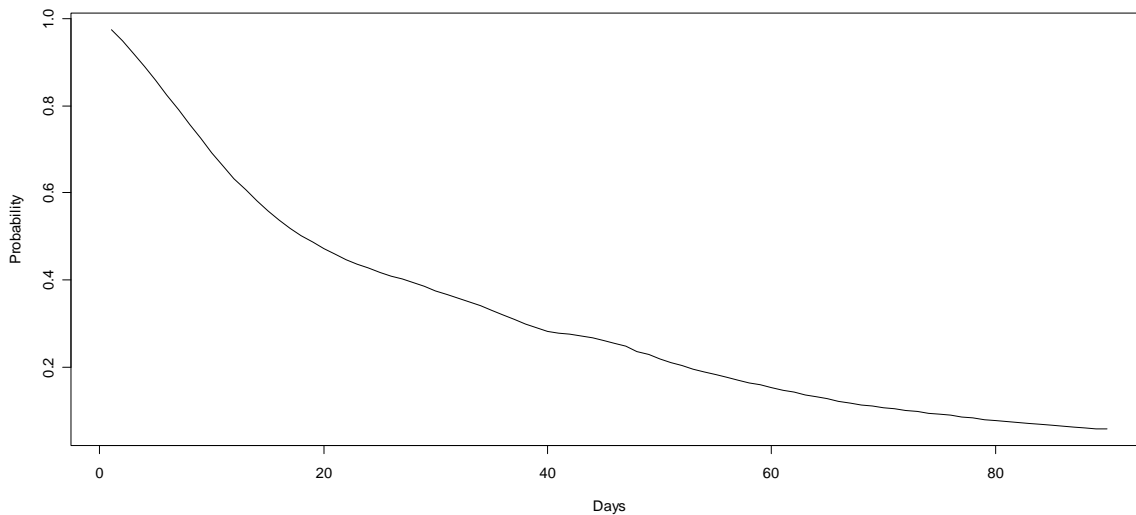
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el segon ingrès segons presenti o no el problema de dislipèmia:



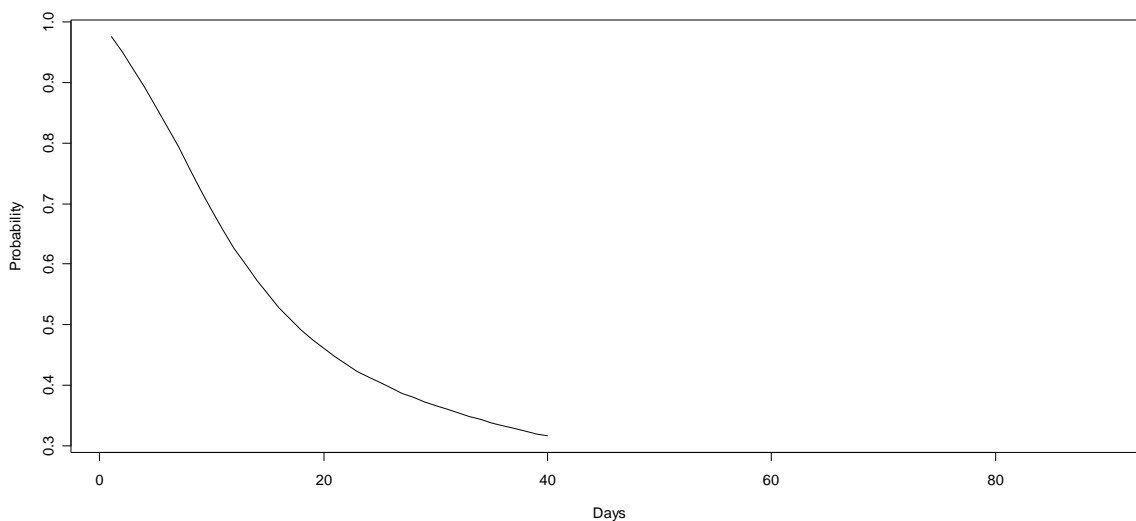
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el tercer ingrès segons presenti o no el problema de dislipèmia:



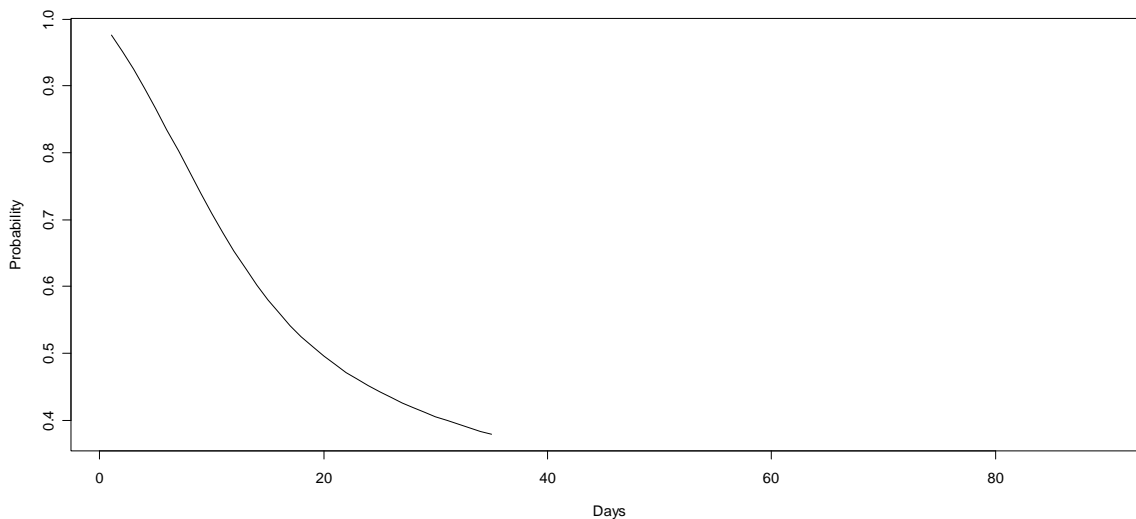
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el quart ingrès segons presenti o no el problema de dislipèmia:



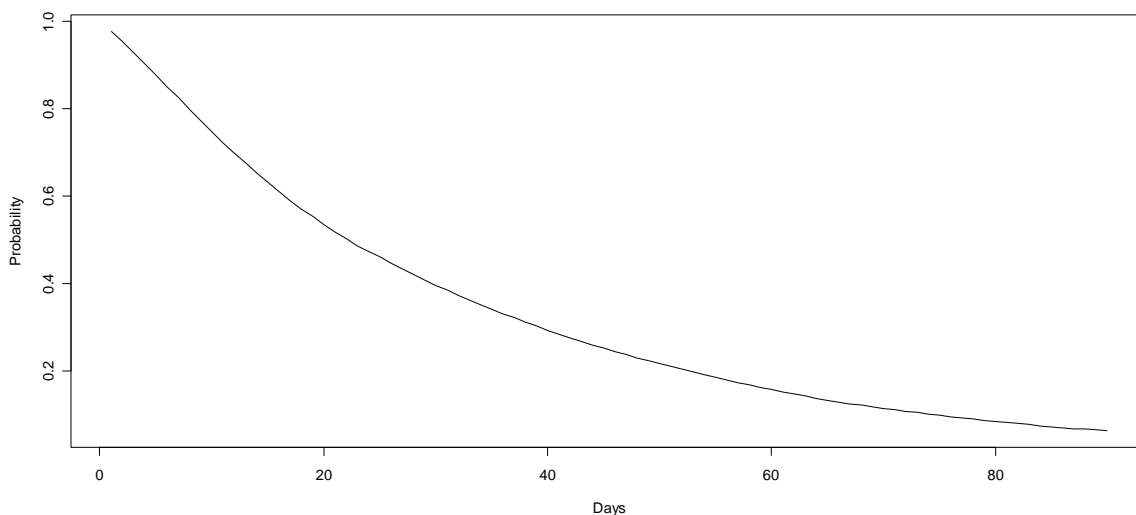
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el cinquè ingrès segons presenti o no el problema de dislipèmia:



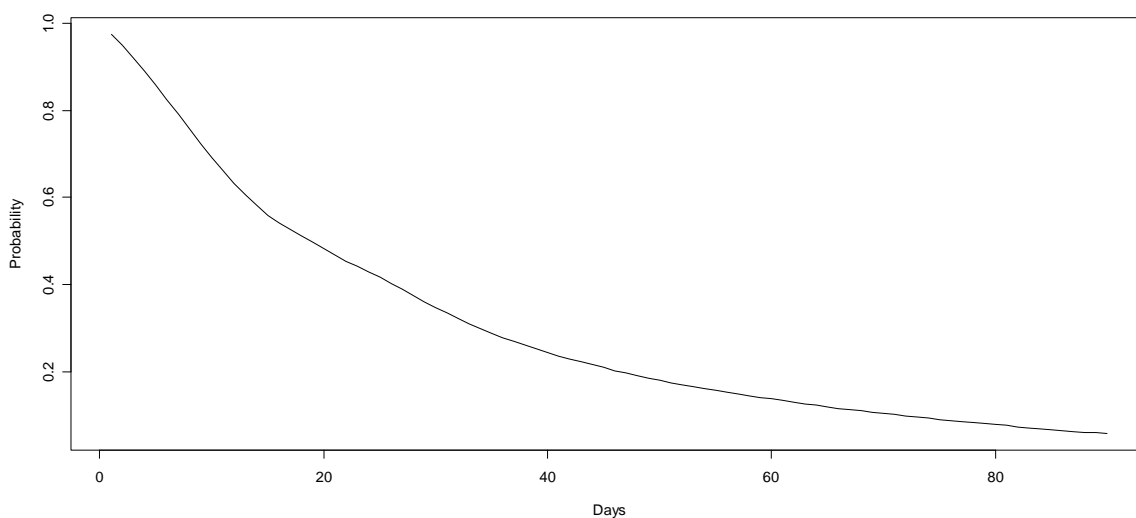
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el primer ingrès segons presenti o no el problema de diabetis mellitus 2:



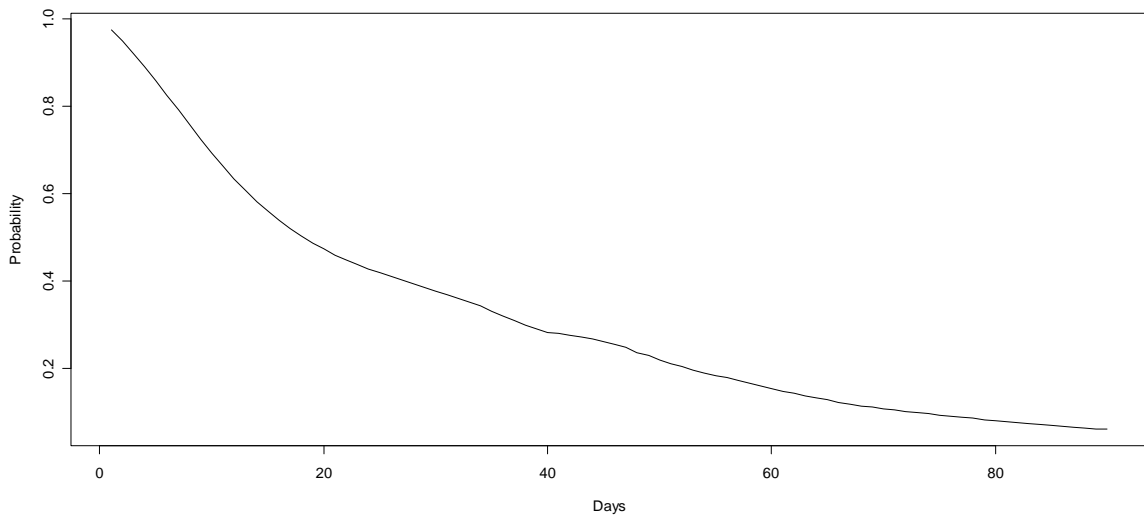
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el segon ingrès segons presenti o no el problema de diabetis mellitus 2:



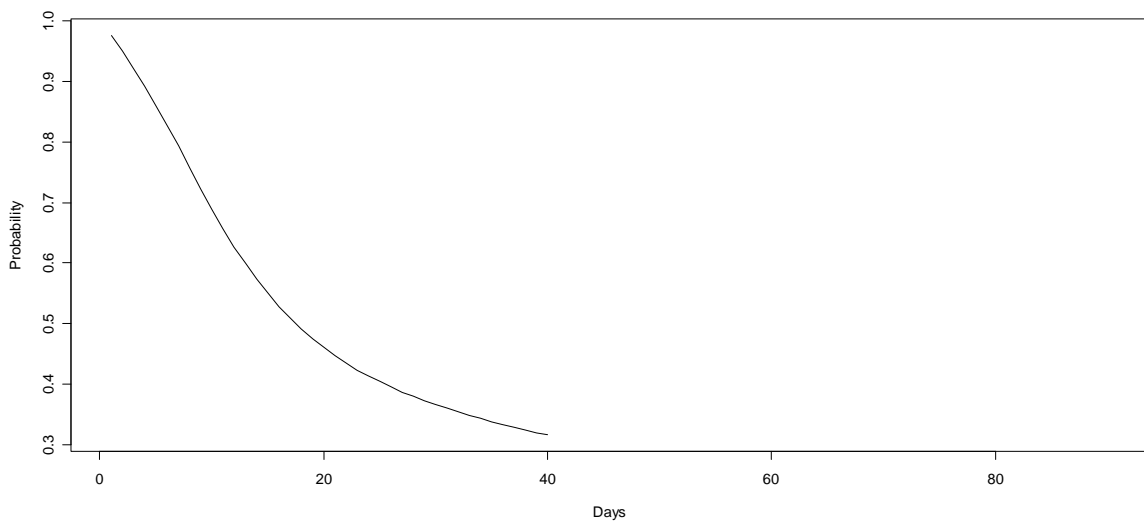
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el tercer ingrès segons presenti o no el problema de diabetis mellitus 2:



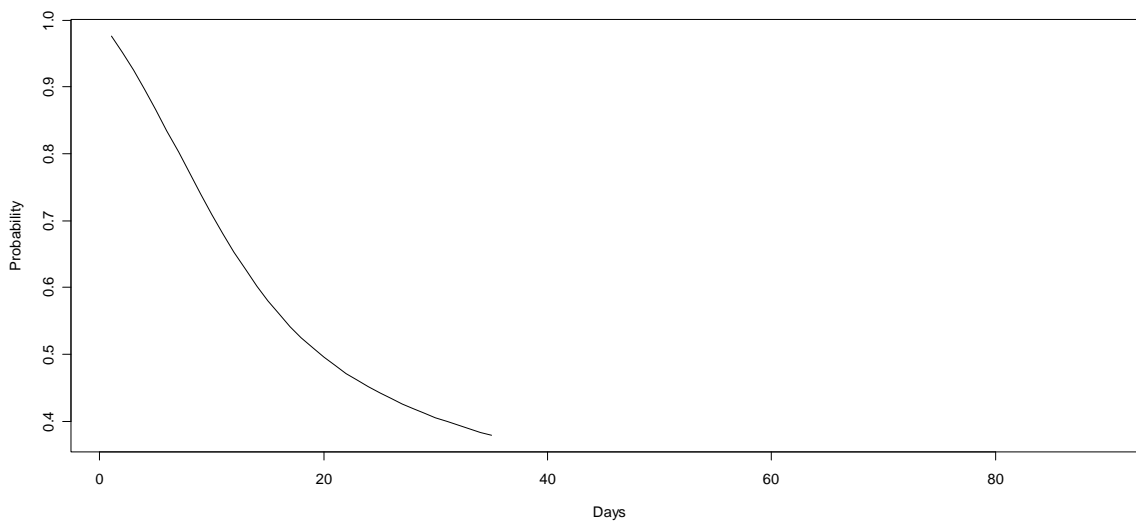
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el quart ingrès segons presenti o no el problema de diabetis mellitus 2:



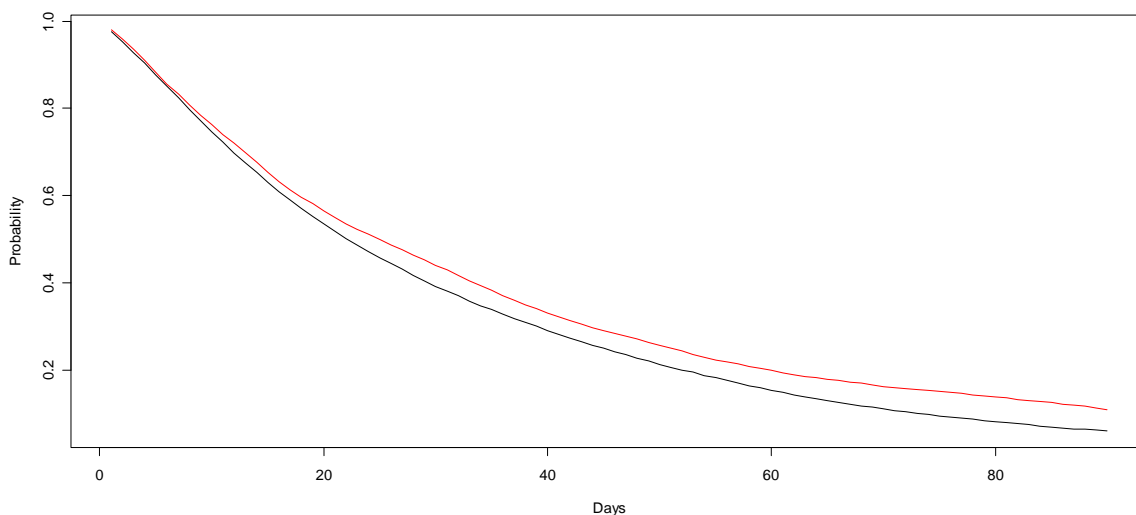
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el cinquè ingrès segons presenti o no el problema de diabetis mellitus 2:



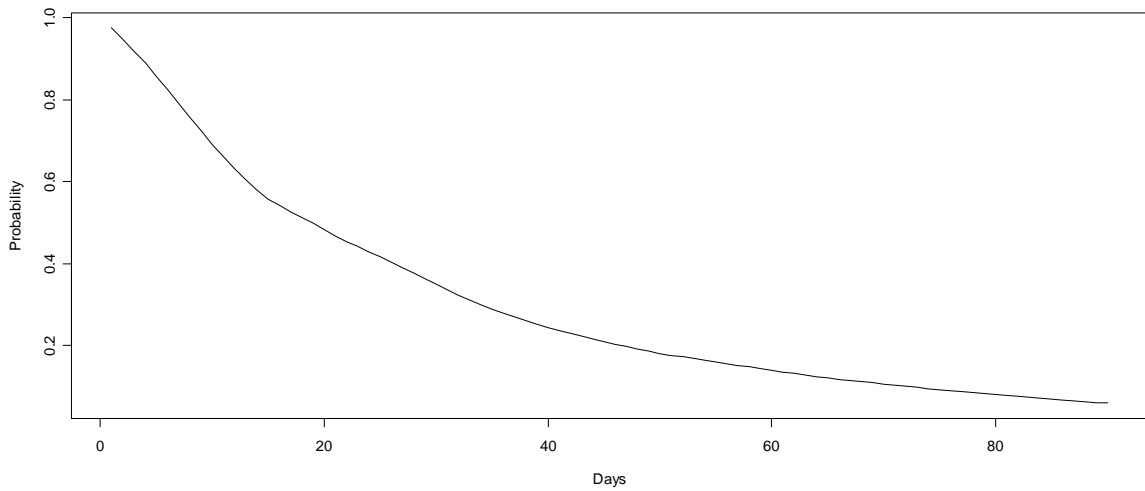
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el primer ingrès segons presenti o no el problema de tabaquisme:



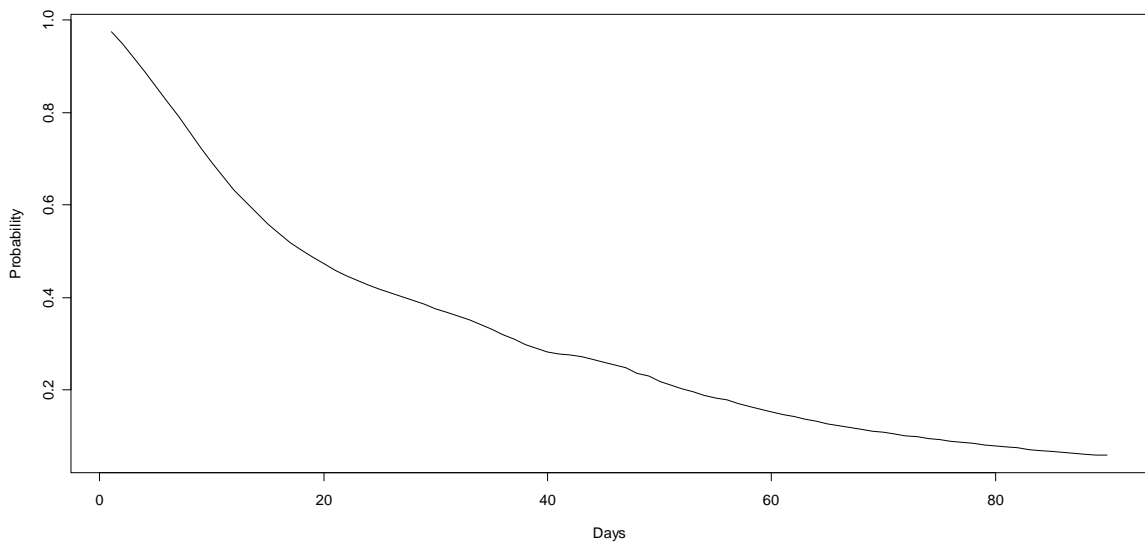
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el segon ingrès segons presenti o no el problema de tabaquisme:



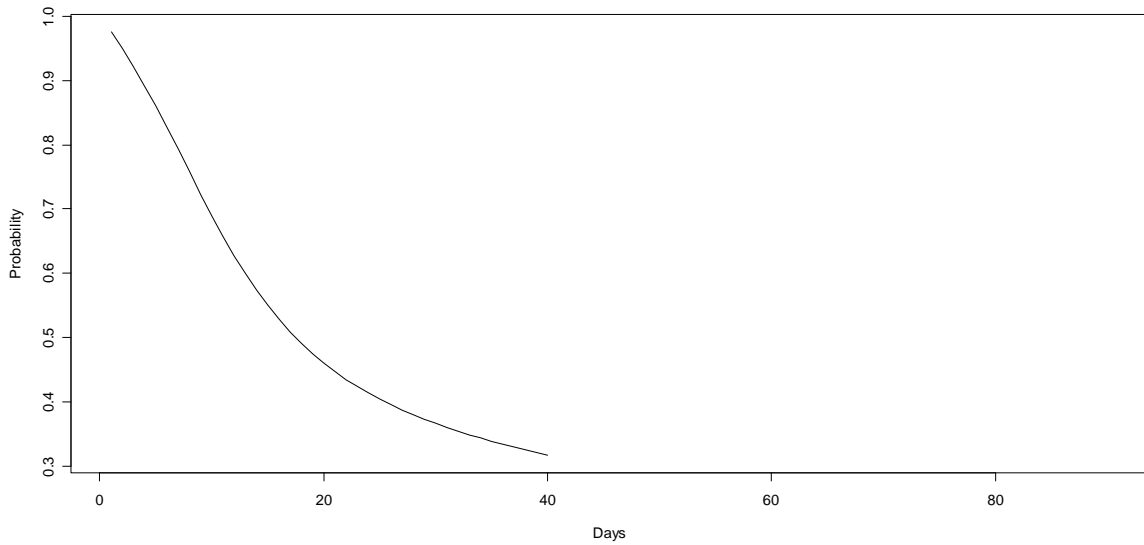
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el tercer ingrès segons presenti o no el problema de tabaquisme:



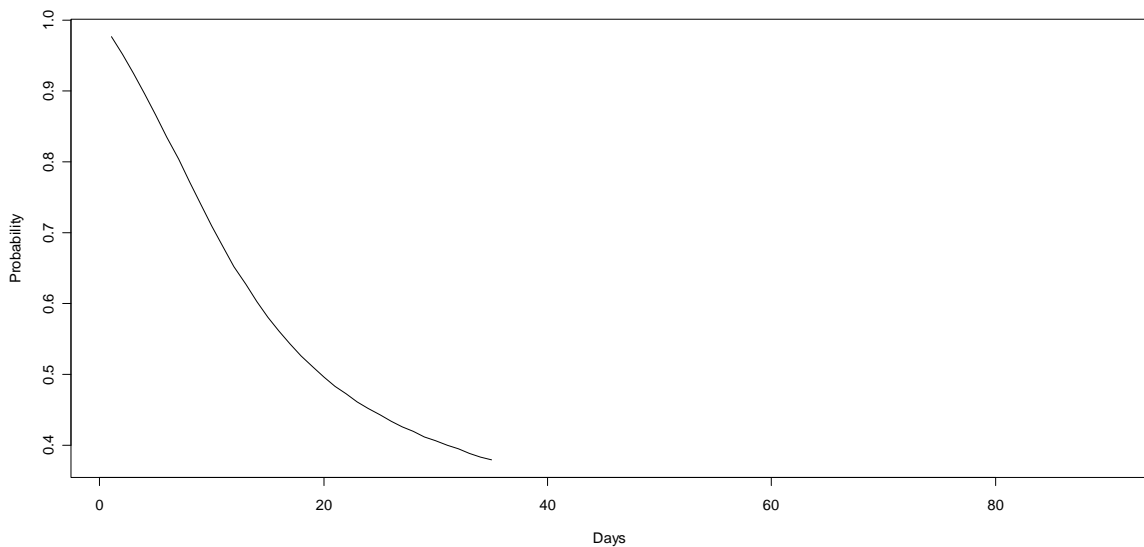
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el quart ingrès segons presenti o no el problema de tabaquisme:



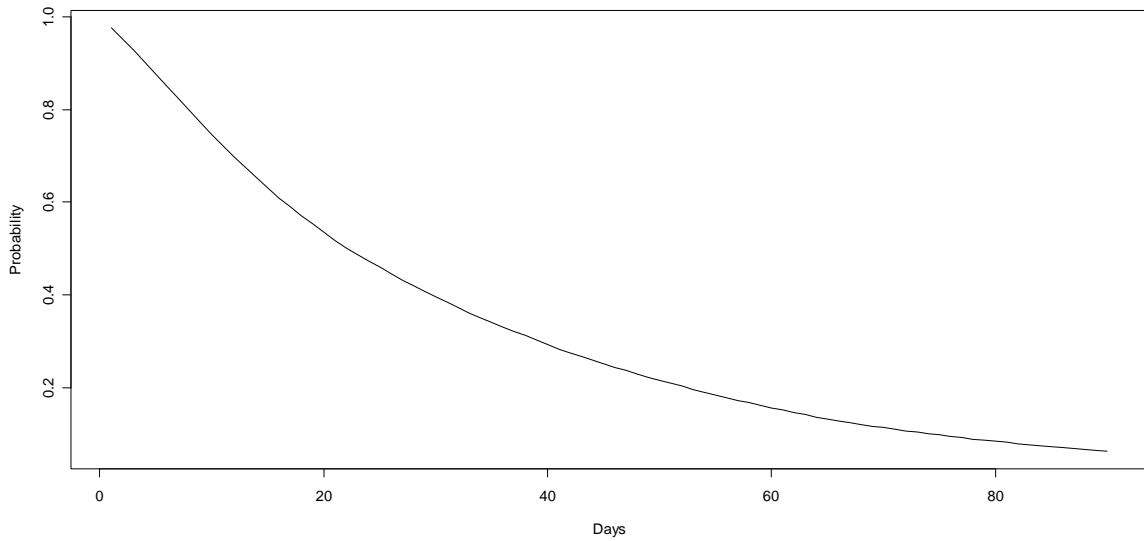
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el cinquè ingrès segons presenti o no el problema de tabaquisme:



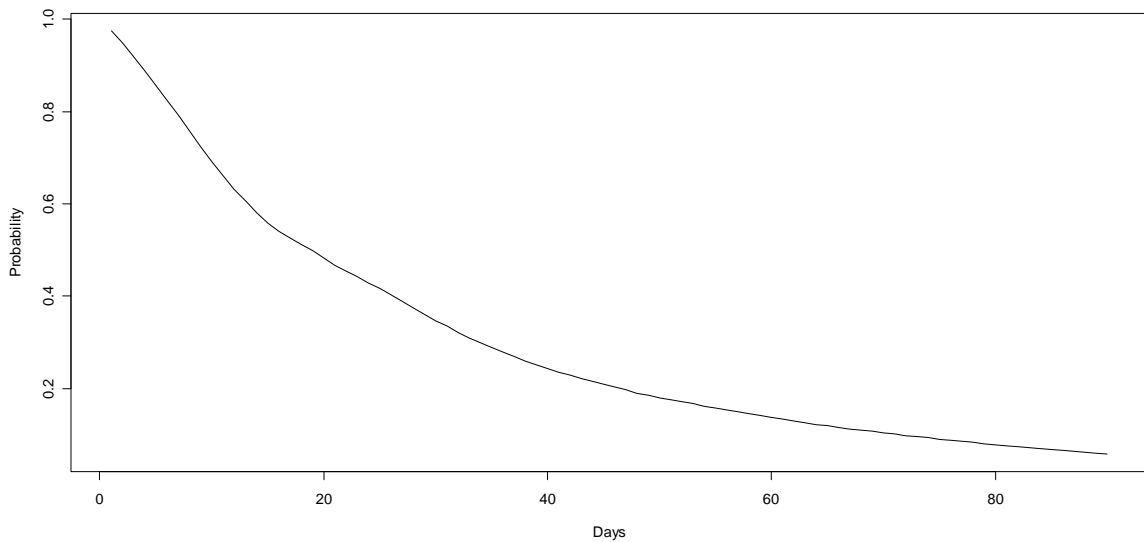
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el primer ingrès segons presentí o no el problema de extabaquisme:



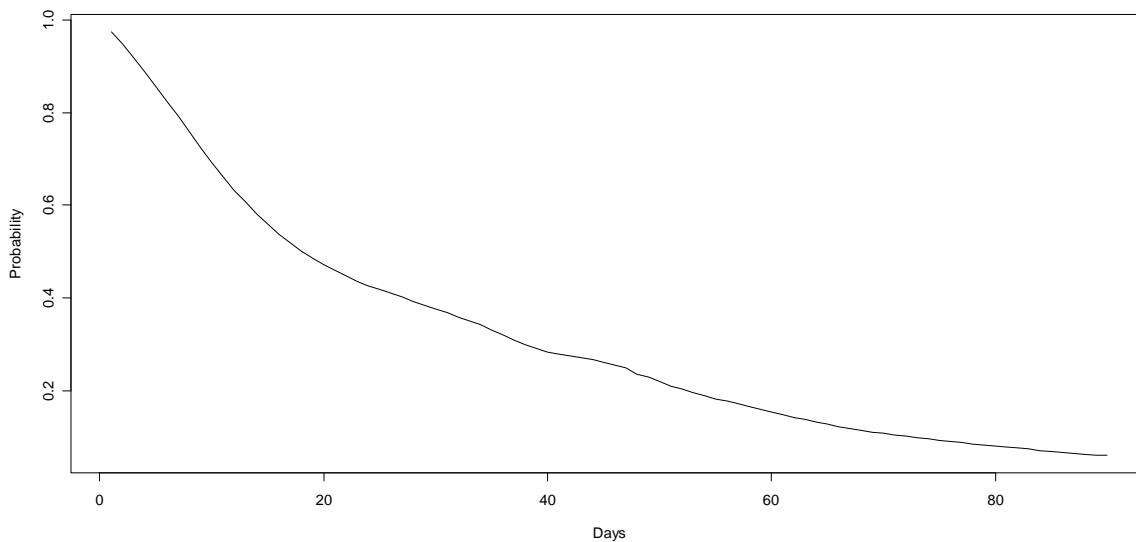
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el segon ingrès segons presentí o no el problema de extabaquisme:



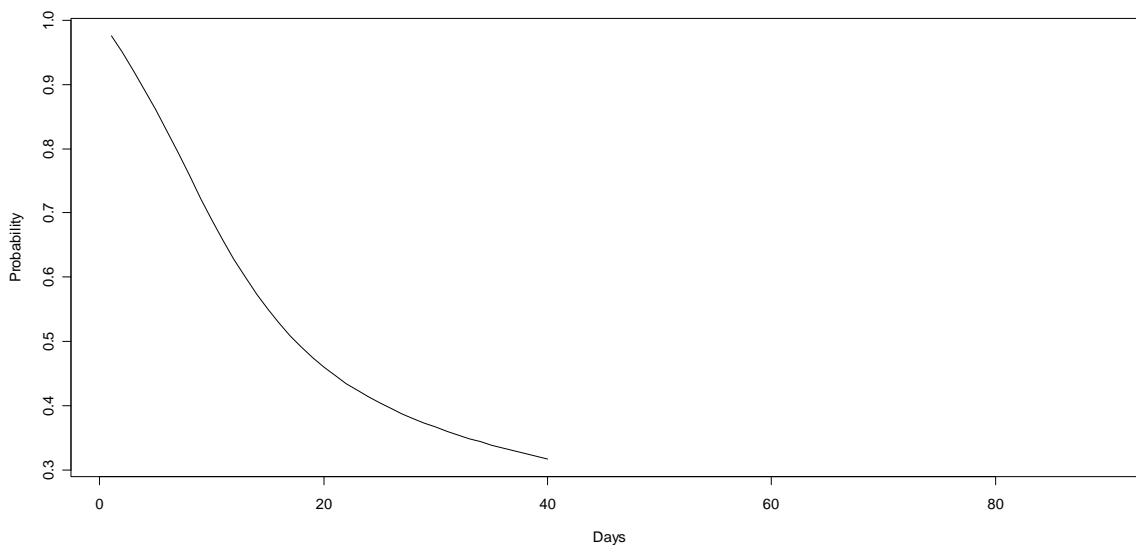
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el tercer ingrès segons presenti o no el problema de extabaquisme:



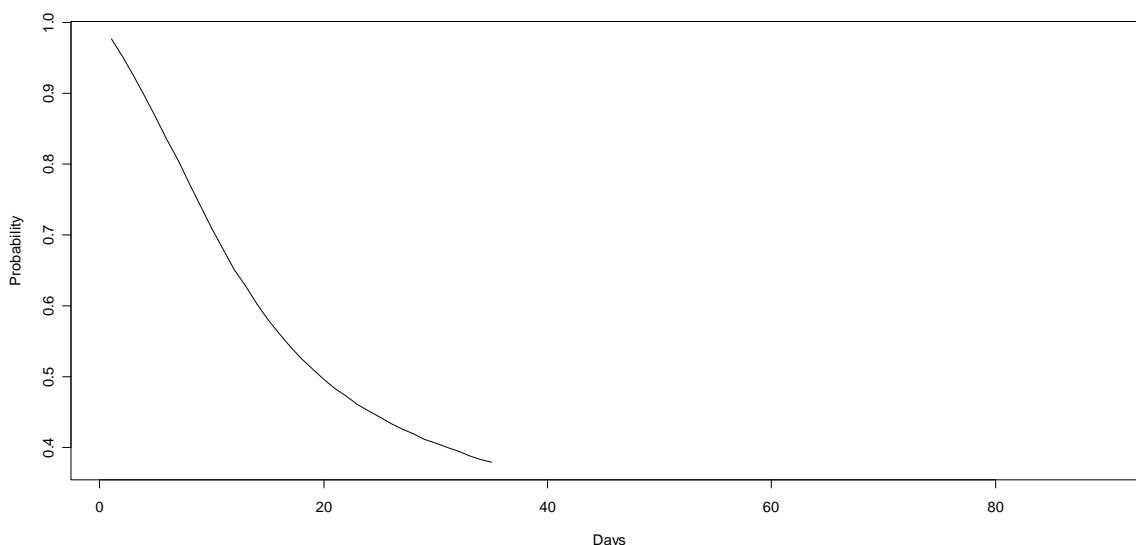
Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el quart ingrès segons presenti o no el problema de extabaquisme:



Negre: no presenta el problema
Vermell: si presenta el problema

Gràfic de la probabilitat de supervivència per el cinquè ingrès segons presenti o no el problema de extabaquisme:



Negre: no presenta el problema
 Vermell: si presenta el problema

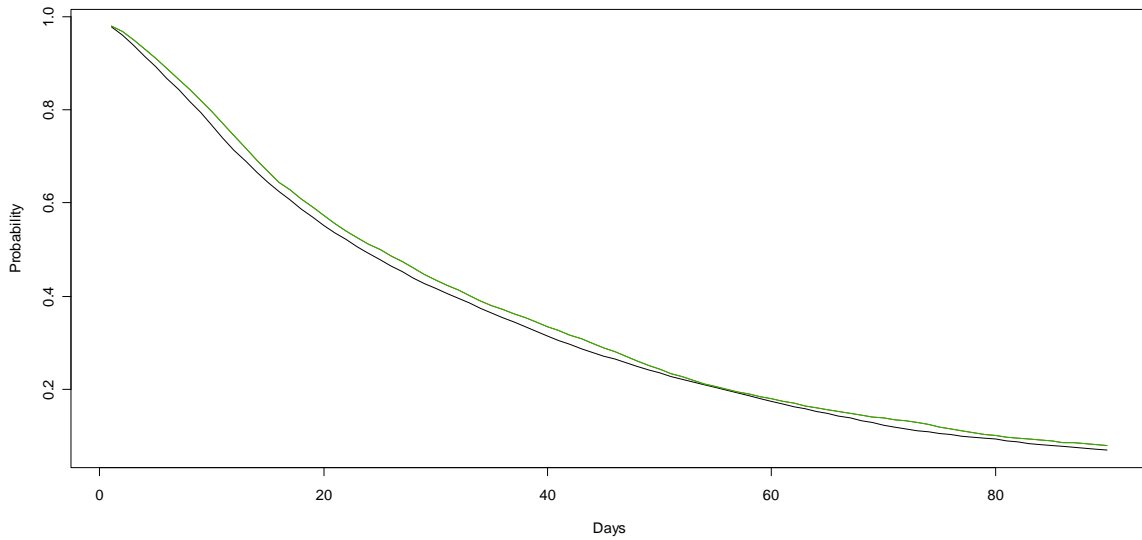
A.VII.2. MODEL AMB EFECTES ESPACIALS

Inference for Bugs model at "ModeloBYMvarexplSIN.txt", fit using winbugs,
 2 chains, each with 10000 iterations (first 1000 discarded), n.thin = 30
 n.sims = 600 iterations saved

	mean	sd	2.5%	25%	50%	75%	97.5%	Rhat	n.eff
b0	-4.08975	0.12093	-4.30717	-4.18000	-4.09250	-4.00475	-3.85293	1.00476	600
b1	-0.07142	0.06257	-0.19466	-0.11495	-0.06893	-0.02578	0.04661	1.00743	190
b21	0.04612	0.08501	-0.12440	-0.01117	0.04985	0.10367	0.21865	0.99940	600
b22	0.00039	0.08962	-0.17769	-0.05986	-0.00269	0.05797	0.17721	1.00085	600
b23	-0.01386	0.08663	-0.18262	-0.07146	-0.01697	0.04370	0.15034	0.99951	600
b24	0.01096	0.08394	-0.15052	-0.04541	0.01261	0.07001	0.16805	1.00231	500
c11	-0.74099	0.89972	-2.19730	-1.45225	-0.84230	-0.05121	1.07705	3.08894	3
c12	-1.64310	0.15124	-1.94500	-1.74425	-1.64350	-1.53675	-1.34200	0.99942	600
c13	1.68824	0.20758	1.26187	1.54900	1.68800	1.82500	2.09227	1.00013	600
c14	1.66167	0.15958	1.34295	1.55175	1.66050	1.76500	1.99202	1.00753	240
b3	-0.01969	0.20424	-0.42687	-0.14275	-0.01824	0.11730	0.38124	1.00429	300
b4	-0.10430	0.08845	-0.27095	-0.16380	-0.10665	-0.04965	0.06970	0.99996	600
b5	0.29688	0.21163	-0.10244	0.16305	0.30735	0.44002	0.71310	0.99986	600
b6	-0.06486	0.11291	-0.29610	-0.14313	-0.06303	0.00824	0.16002	0.99937	600
b7	-0.03372	0.13511	-0.30330	-0.12685	-0.03483	0.05408	0.24191	1.01162	120
b8	-0.04997	0.12864	-0.29356	-0.12885	-0.05282	0.03321	0.21150	1.00676	200
b9	-0.21608	0.14348	-0.50366	-0.30667	-0.21205	-0.12523	0.06194	0.99934	600
b10	0.25330	0.22034	-0.17206	0.09123	0.26595	0.40747	0.66535	0.99975	600
b11	0.71185	0.90024	-1.09305	0.01237	0.81670	1.43650	2.16905	3.10217	3
sigma.beta	0.30955	0.04308	0.22412	0.27957	0.31095	0.34182	0.38706	1.01530	160
sdS	0.22777	0.10730	0.02798	0.15745	0.22985	0.29897	0.43551	1.00539	250

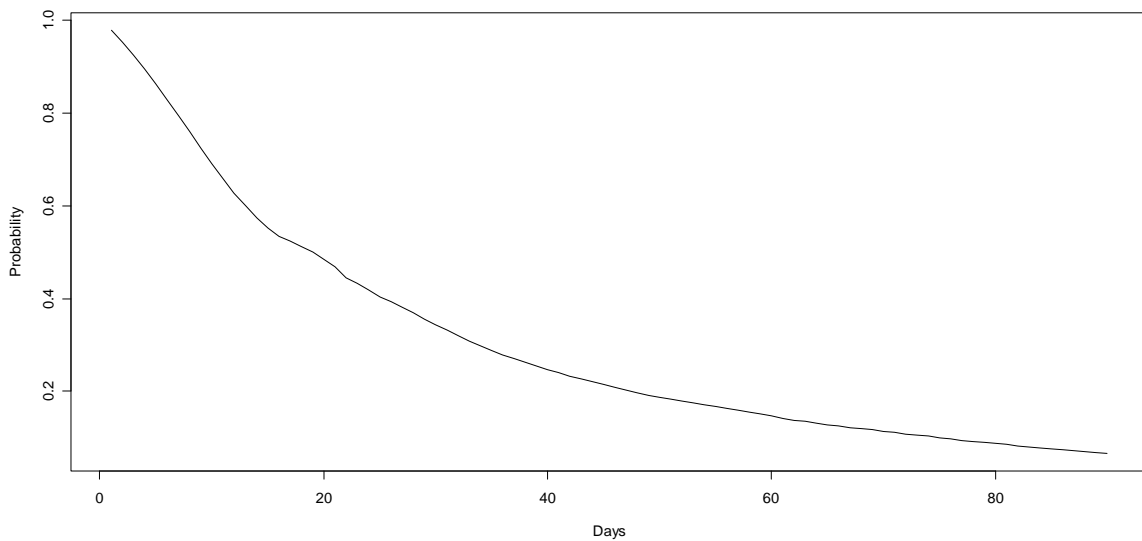
For each parameter, n.eff is a crude measure of effective sample size,
 and Rhat is the potential scale reduction factor (at convergence, Rhat=1).

Gràfic de la probabilitat de supervivència per el primer ingrès segons el sexe:



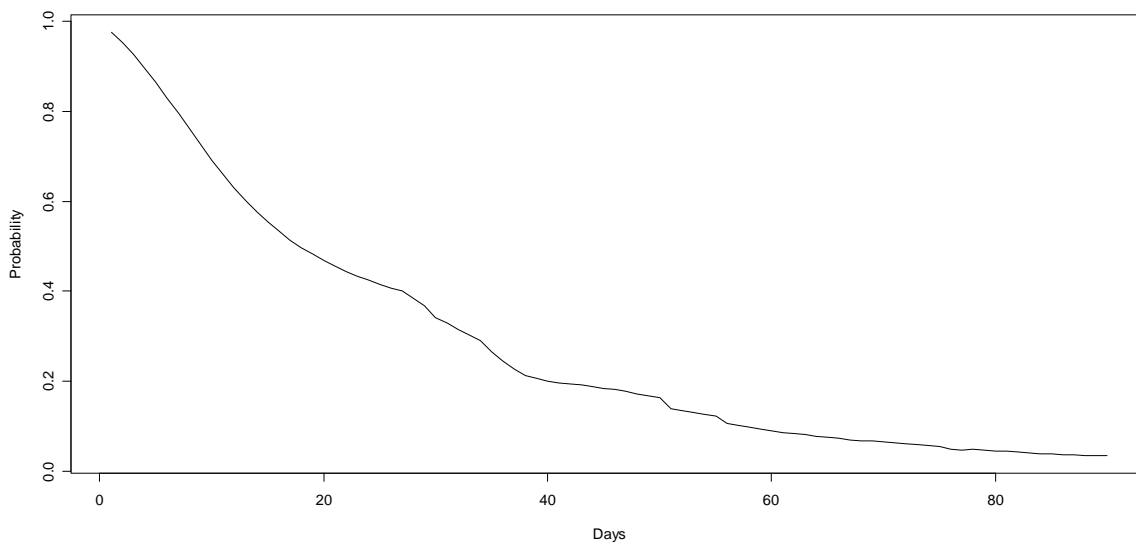
Negre: home
Verd: dona

Gràfic de la probabilitat de supervivència per el segon ingrès segons el sexe:



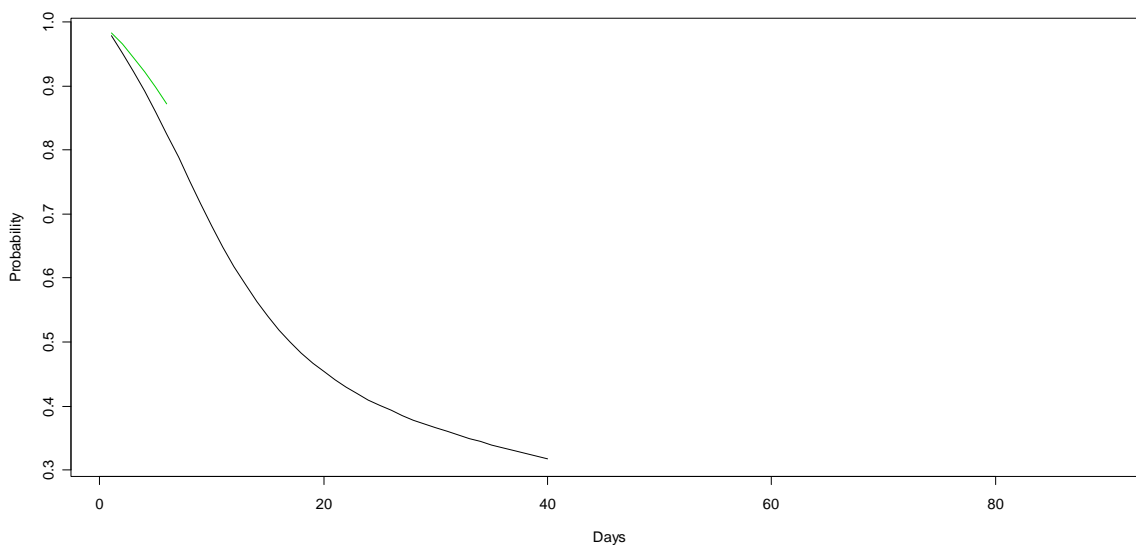
Negre: home
Verd: dona

Gràfic de la probabilitat de supervivència per el tercer ingrès segons el sexe:



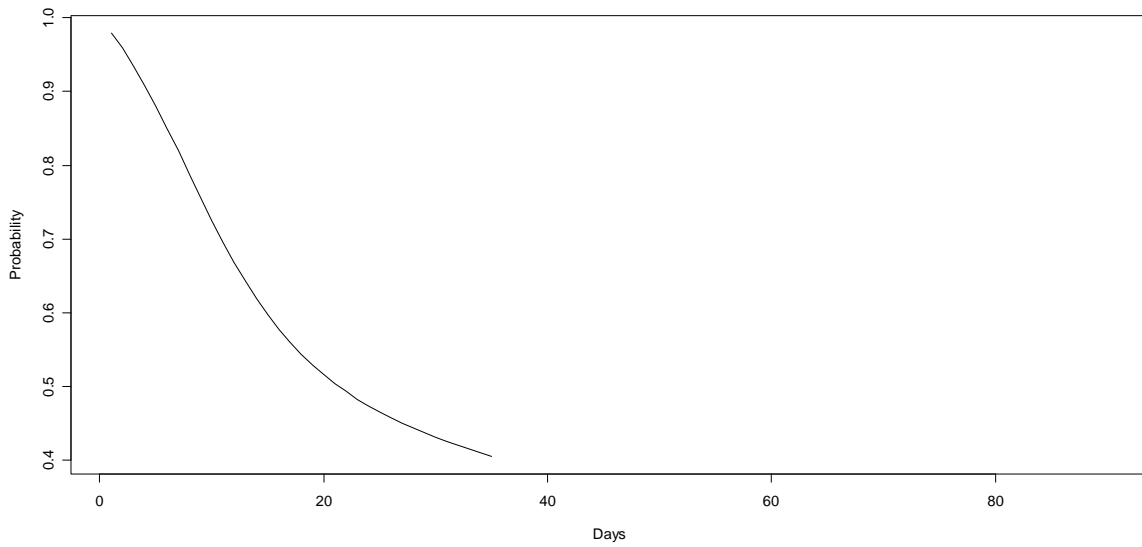
Negre: home
Verd: dona

Gràfic de la probabilitat de supervivència per el quart ingrès segons el sexe:



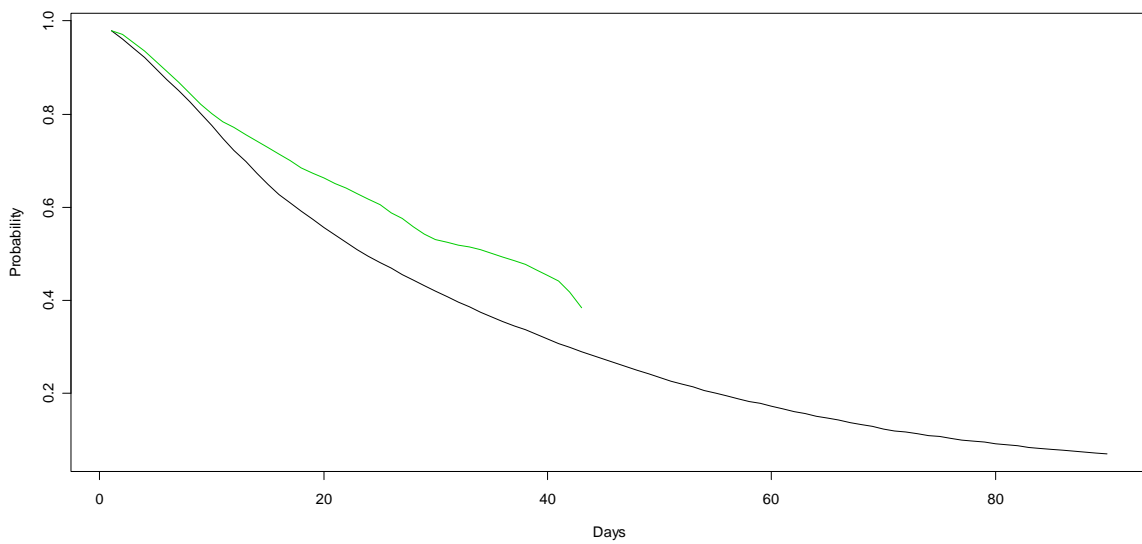
Negre: home
Verd: dona

Gràfic de la probabilitat de supervivència per el cinquè ingrès segons el sexe:



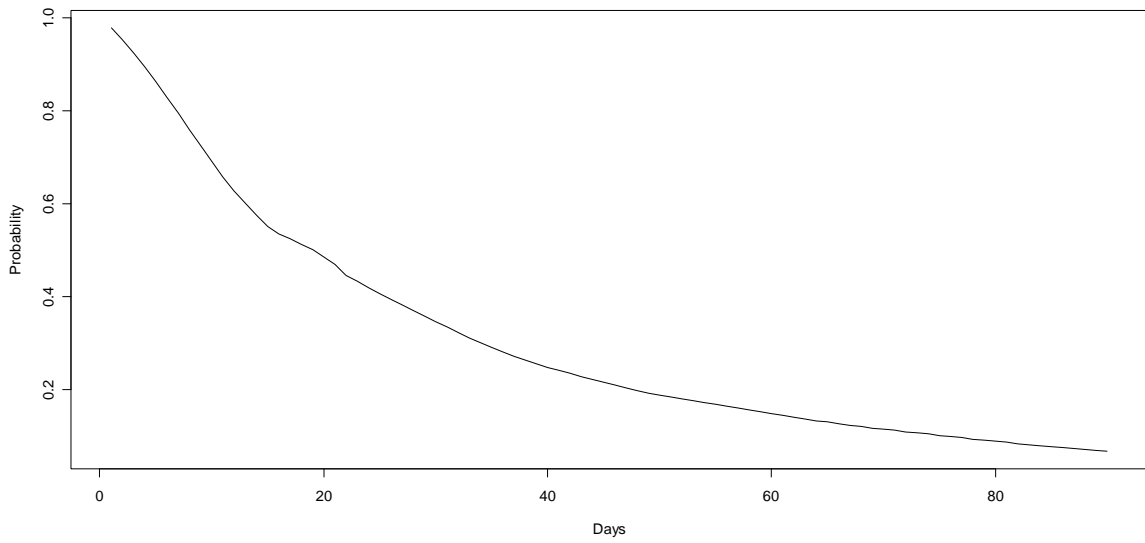
Negre: home
Verd: dona

Gràfic de la probabilitat de supervivència per el primer ingrès segons presenti o no el problema d'alcohol:



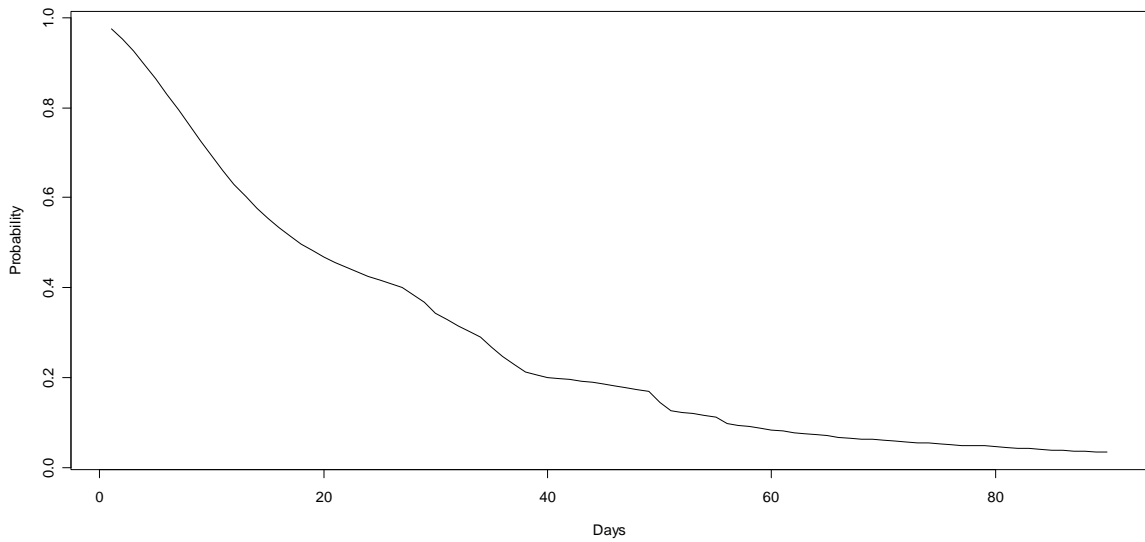
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el segon ingrès segons presenti o no el problema d'alcohol:



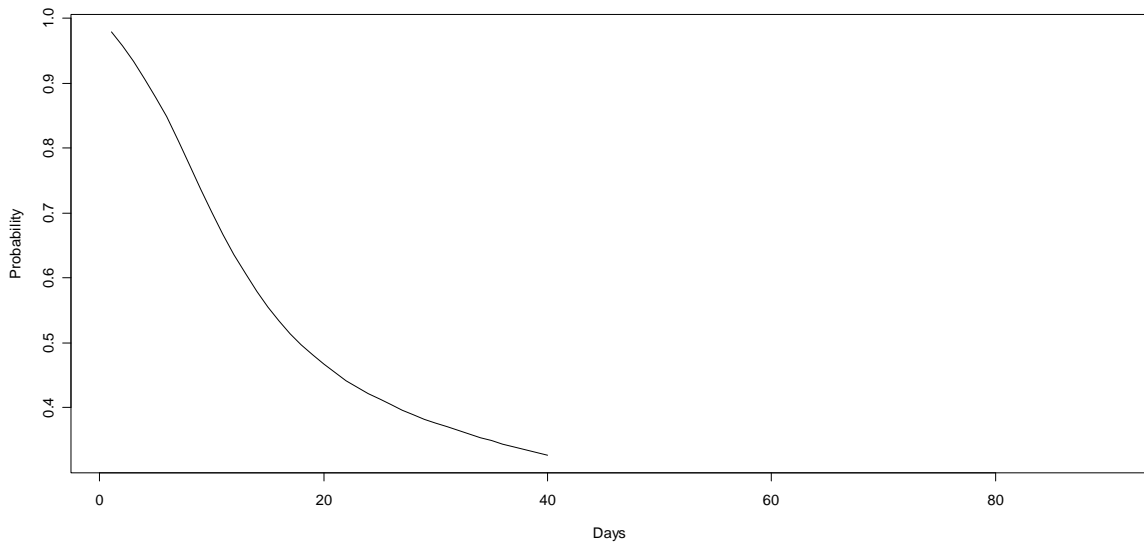
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el tercer ingrès segons presenti o no el problema d'alcohol:



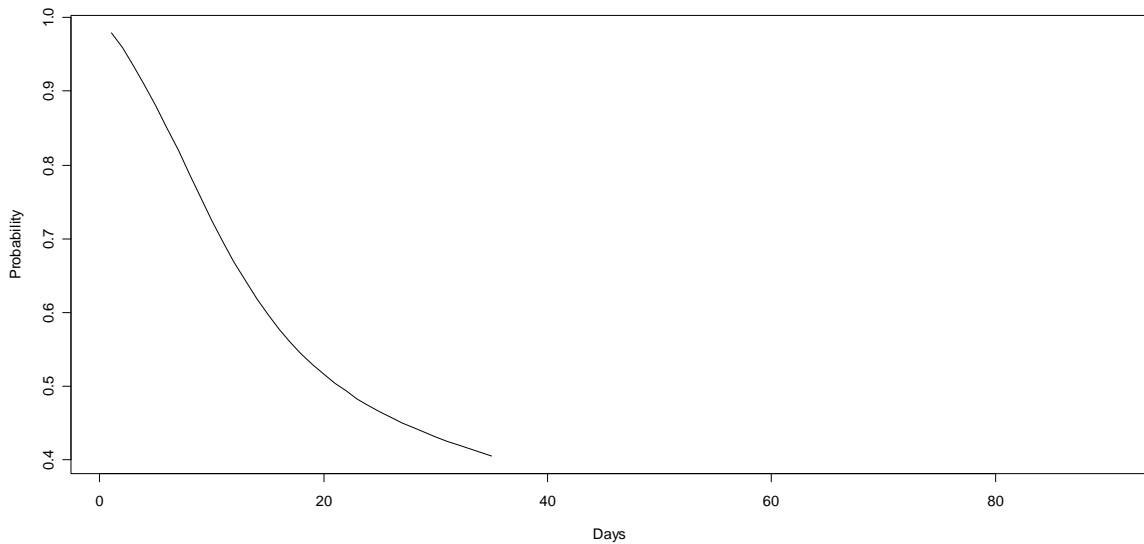
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el quart ingrès segons presenti o no el problema d'alcohol:



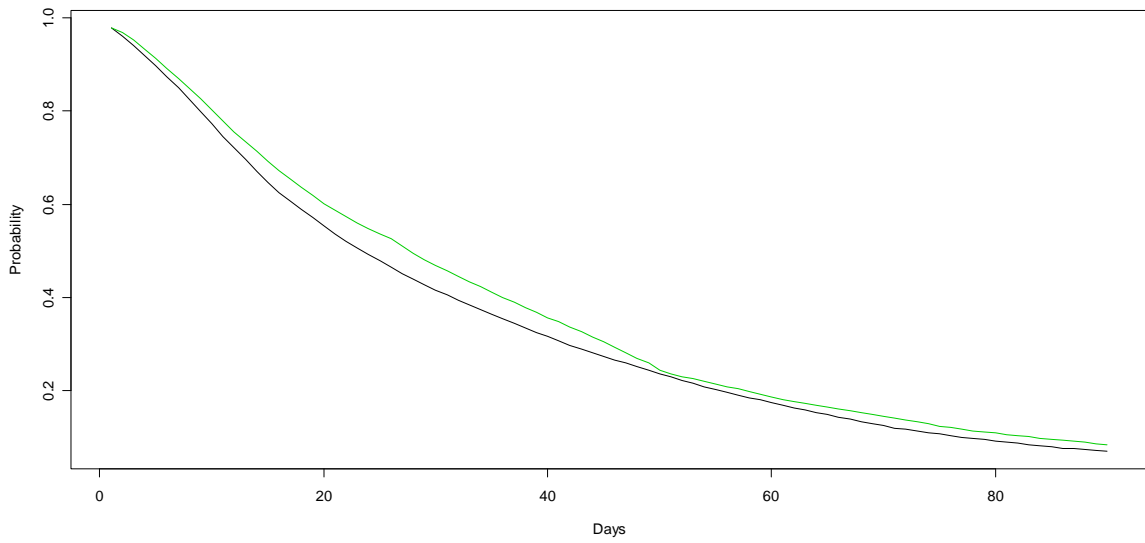
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el cinquè ingrès segons presenti o no el problema d'alcohol:



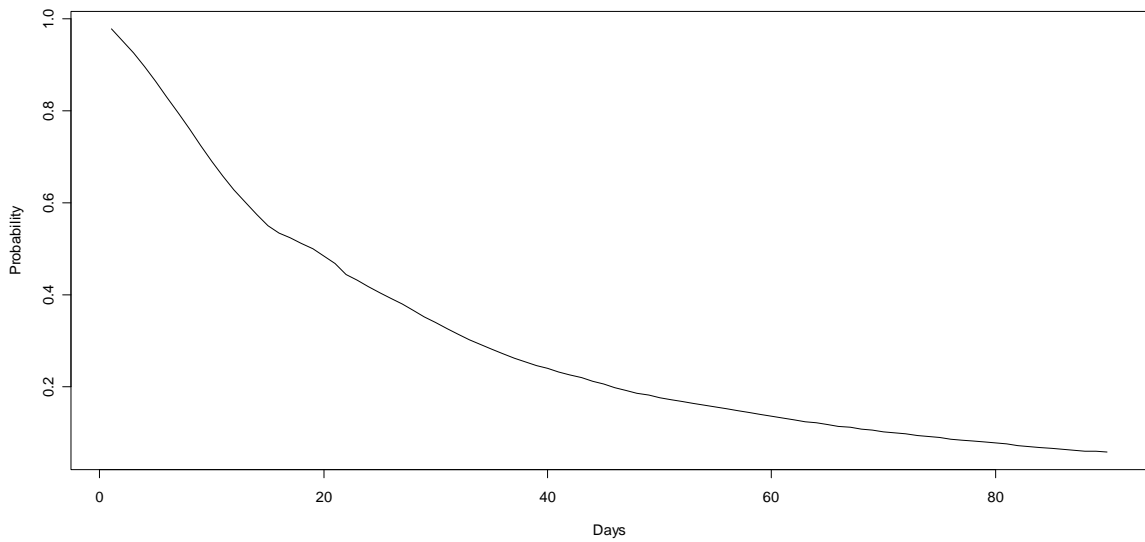
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el primer ingrès segons presenti o no el problema de MPOC:



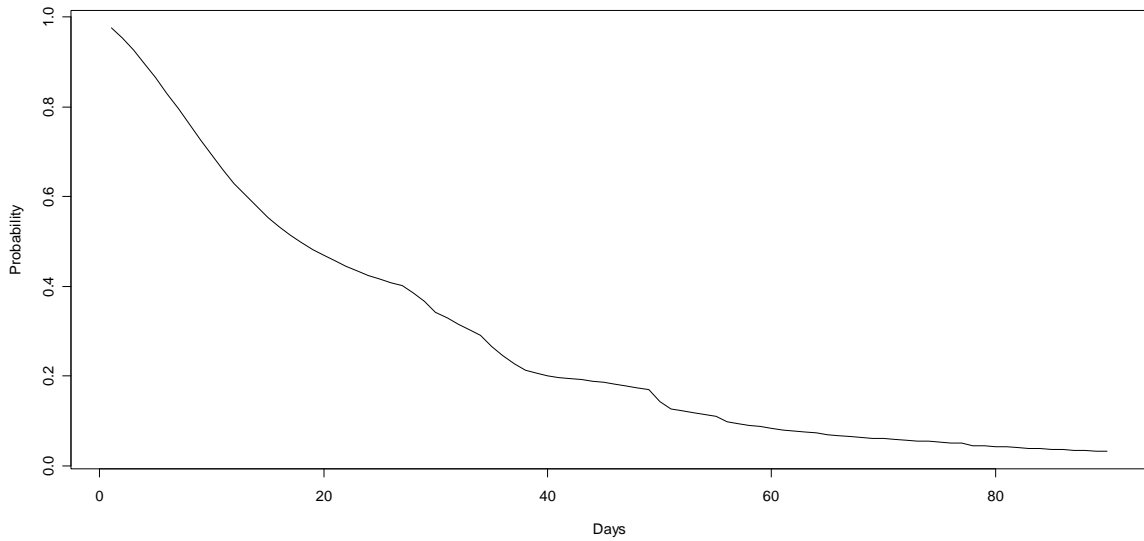
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el segon ingrès segons presenti o no el problema de MPOC:



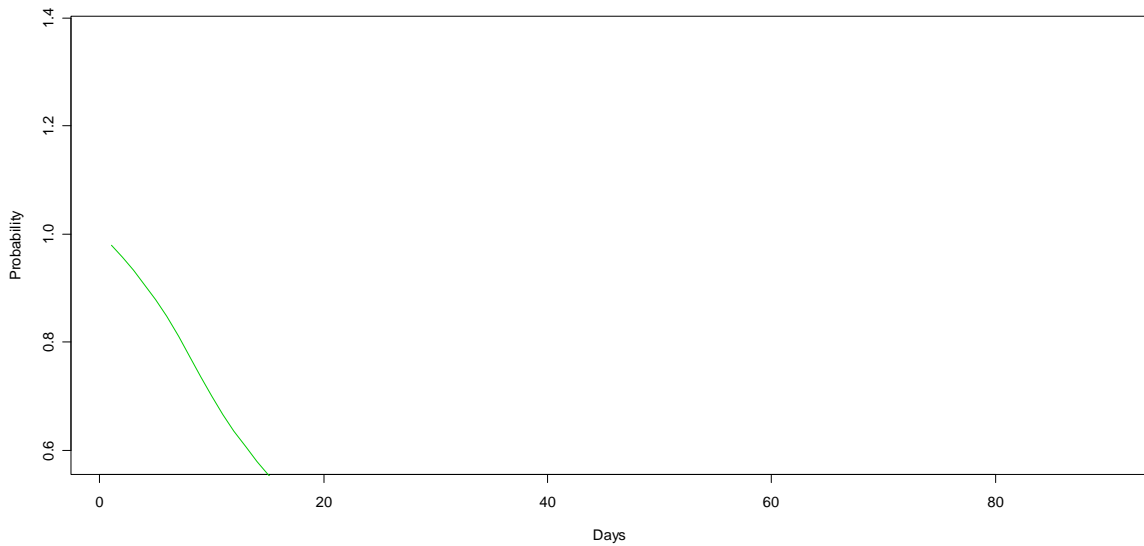
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el tercer ingrès segons presenti o no el problema de MPOC:



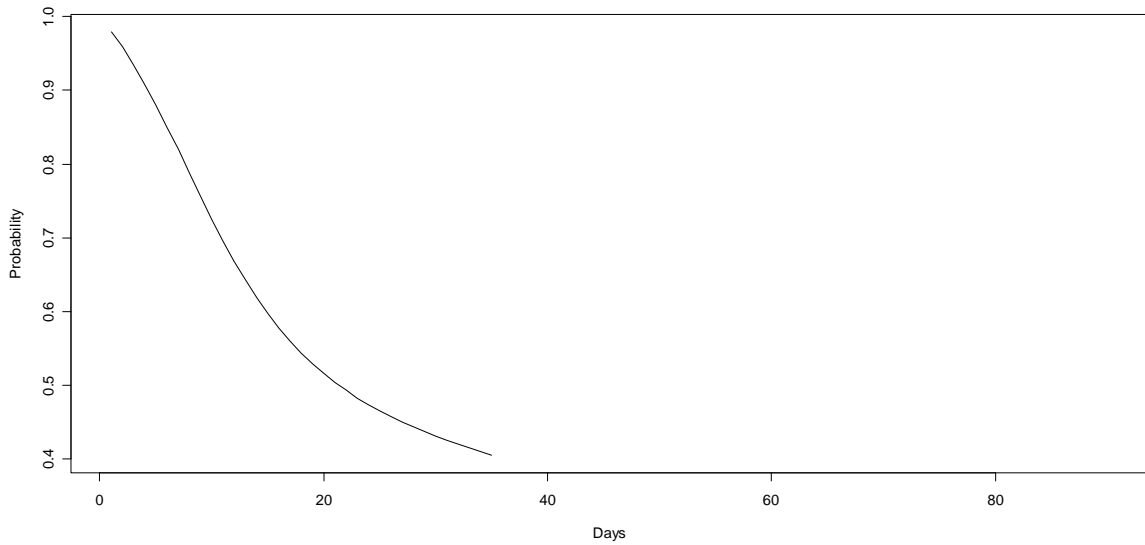
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el quart ingrès segons presenti o no el problema de MPOC:



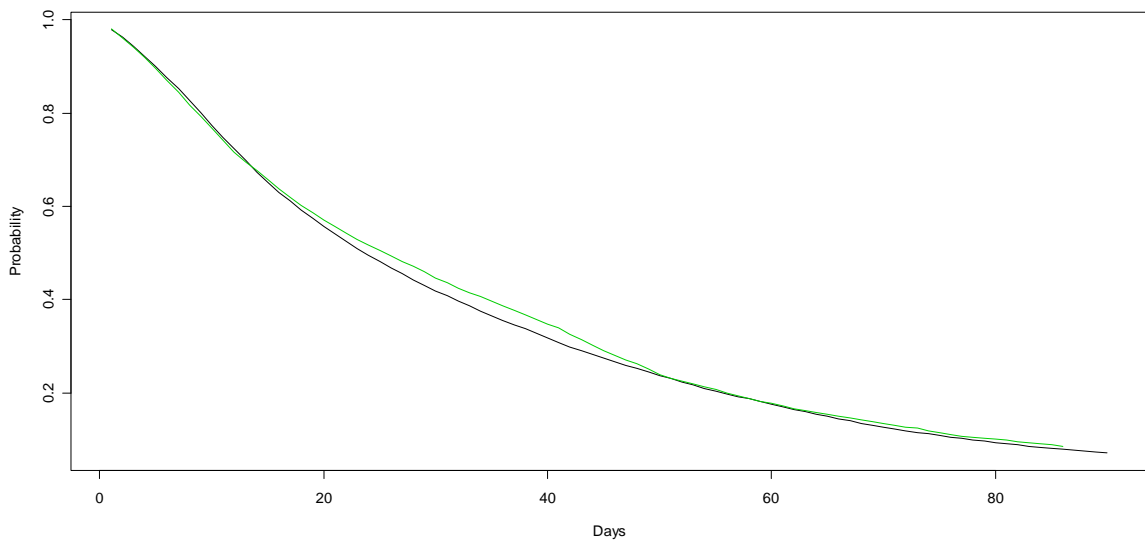
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el cinquè ingrès segons presenti o no el problema de MPOC:



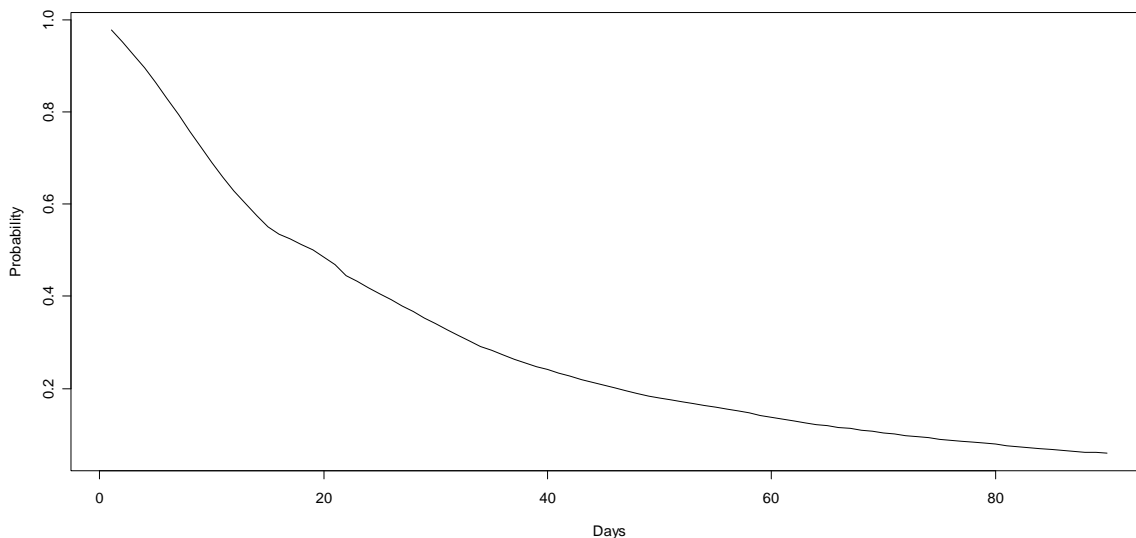
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el primer ingrès segons presenti o no el problema de hipertensió arterial:



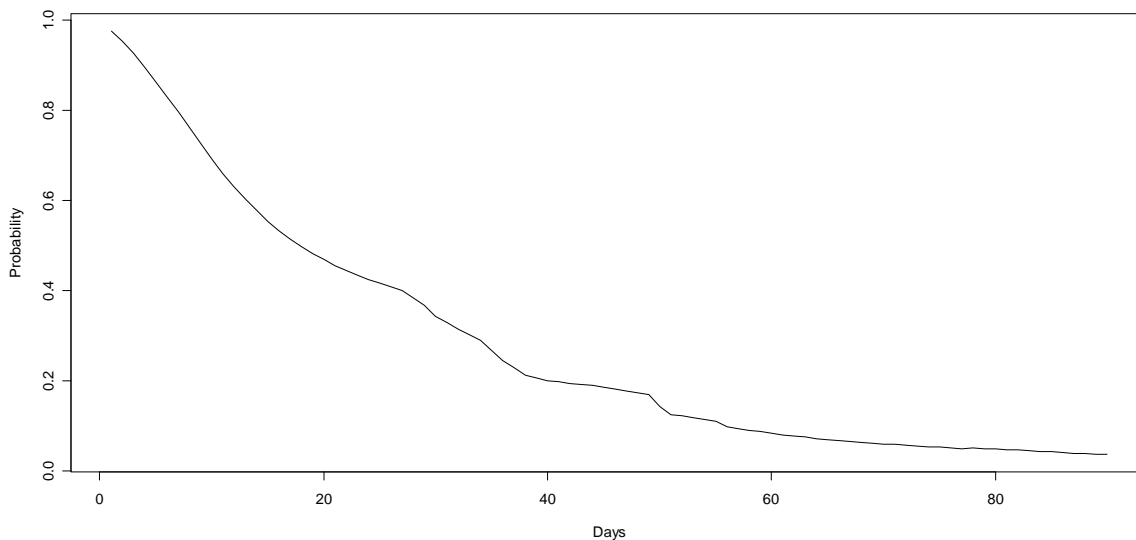
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el segon ingrès segons presenti o no el problema de hipertensió arterial:



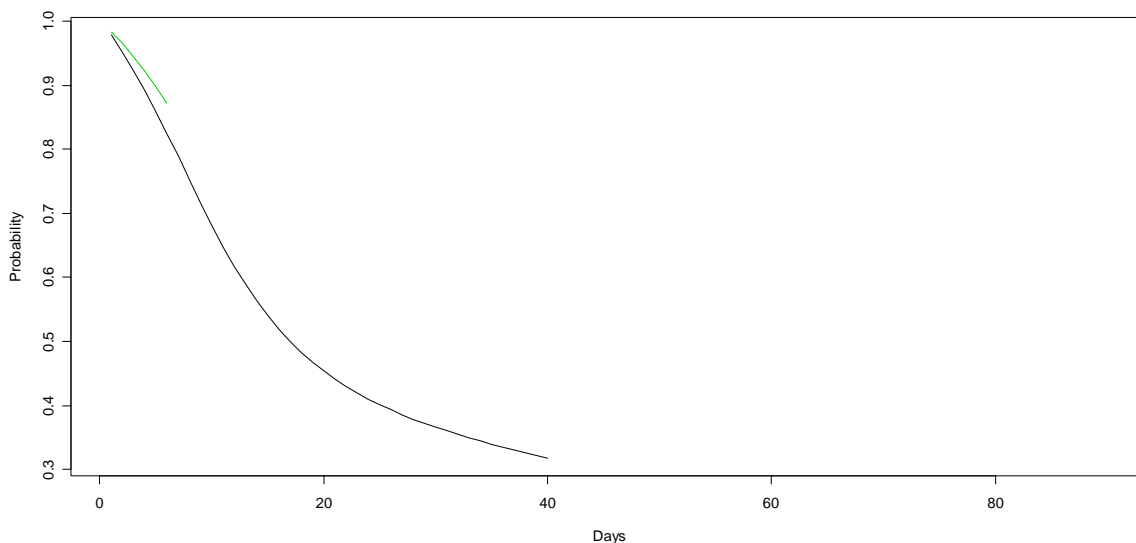
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el tercer ingrès segons presenti o no el problema de hipertensió arterial:



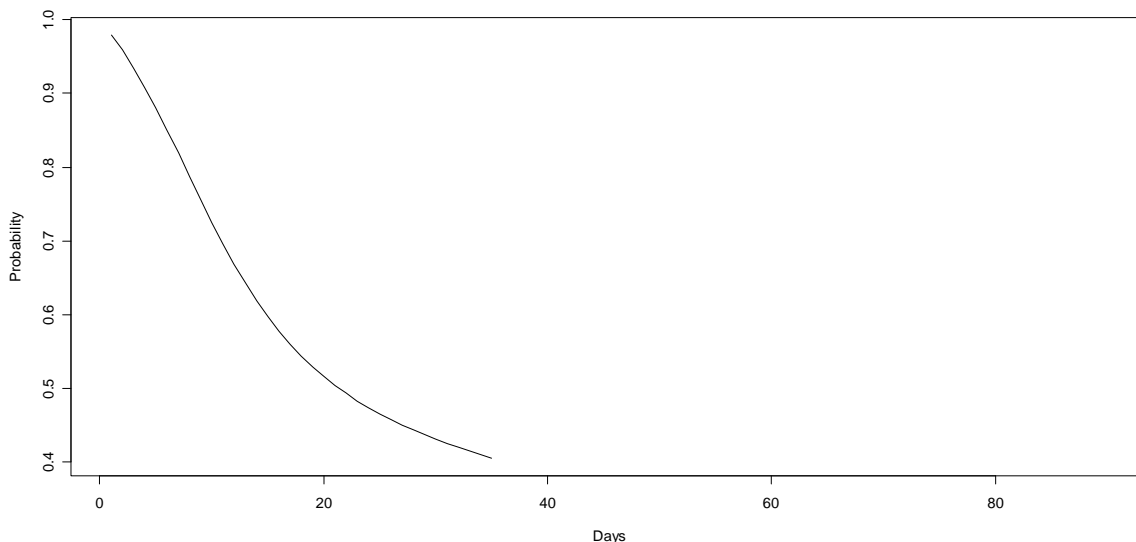
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el quart ingrès segons presenti o no el problema de hipertensió arterial:



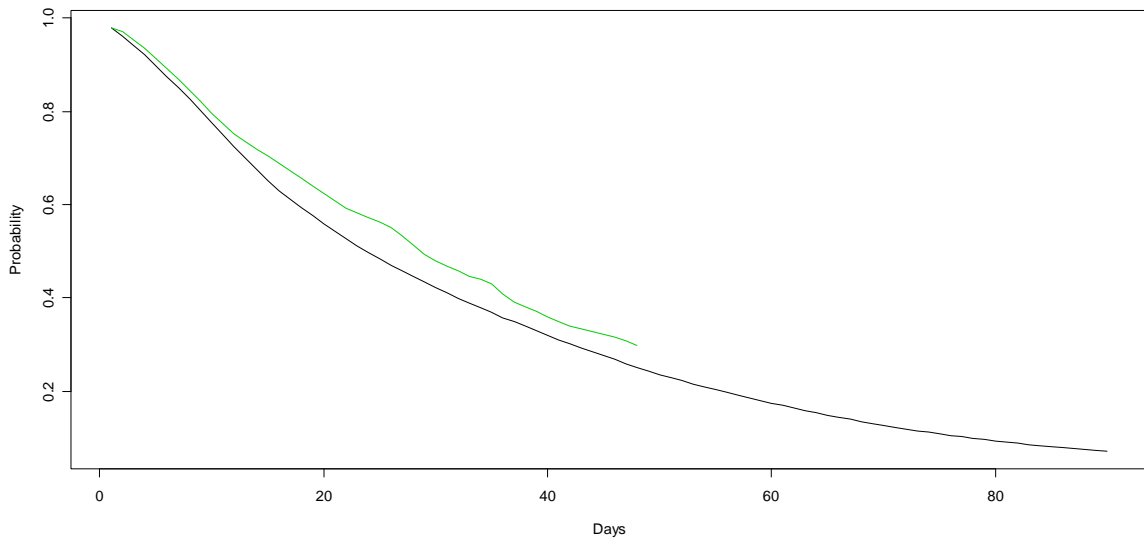
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el cinquè ingrès segons presenti o no el problema de hipertensió arterial:



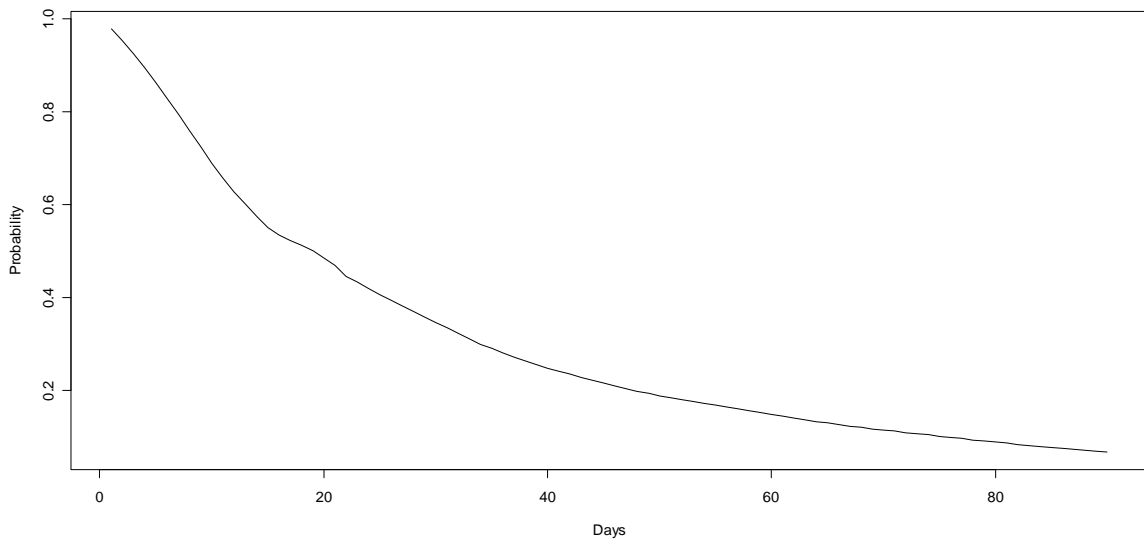
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el primer ingrès segons presenti o no el problema de insuficiència cardíaca:



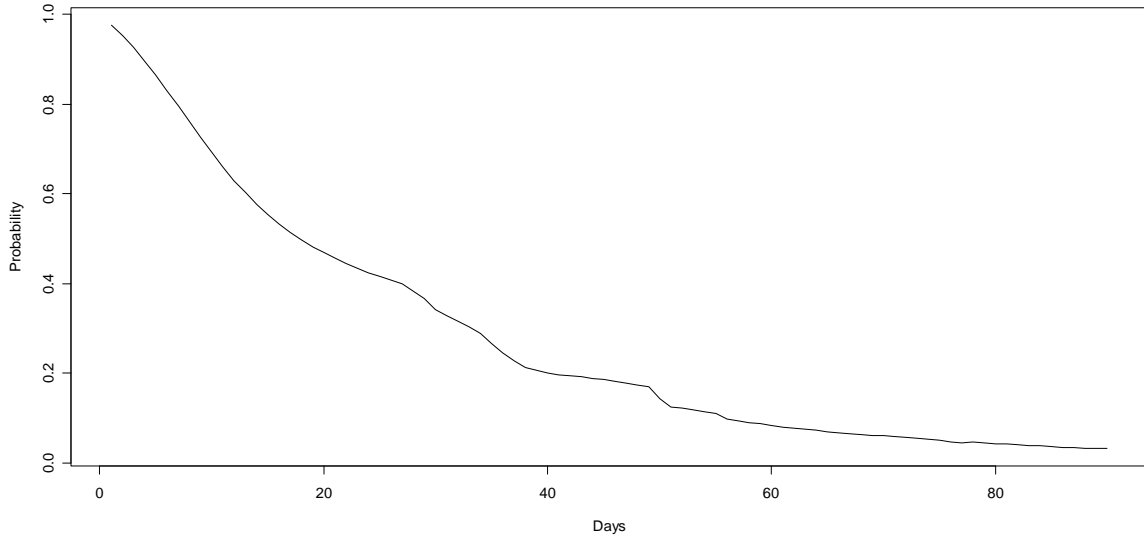
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el segon ingrès segons presenti o no el problema de insuficiència cardíaca:



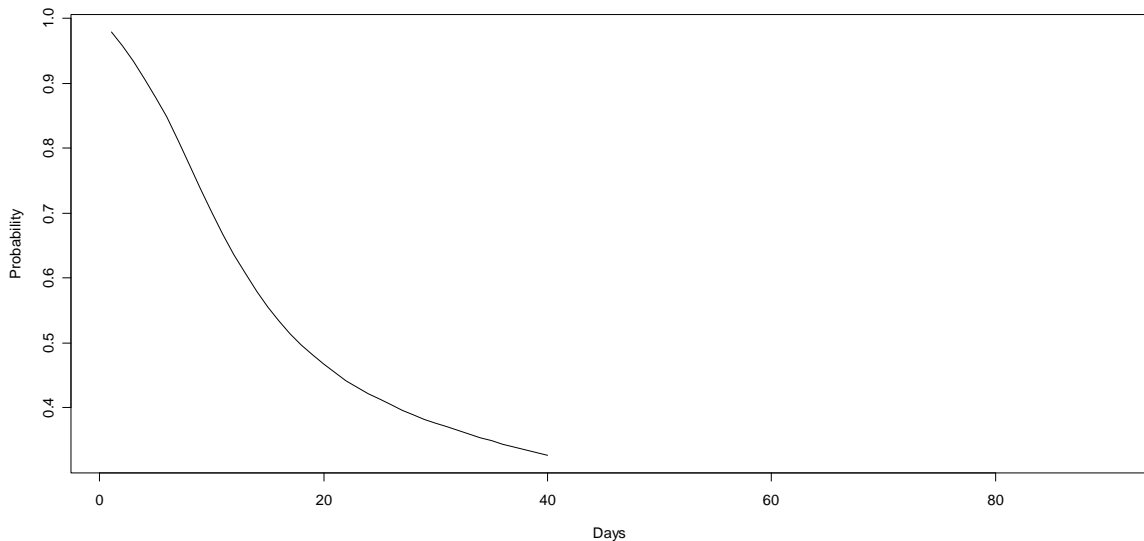
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el tercer ingrès segons presenti o no el problema de insuficiència cardíaca:



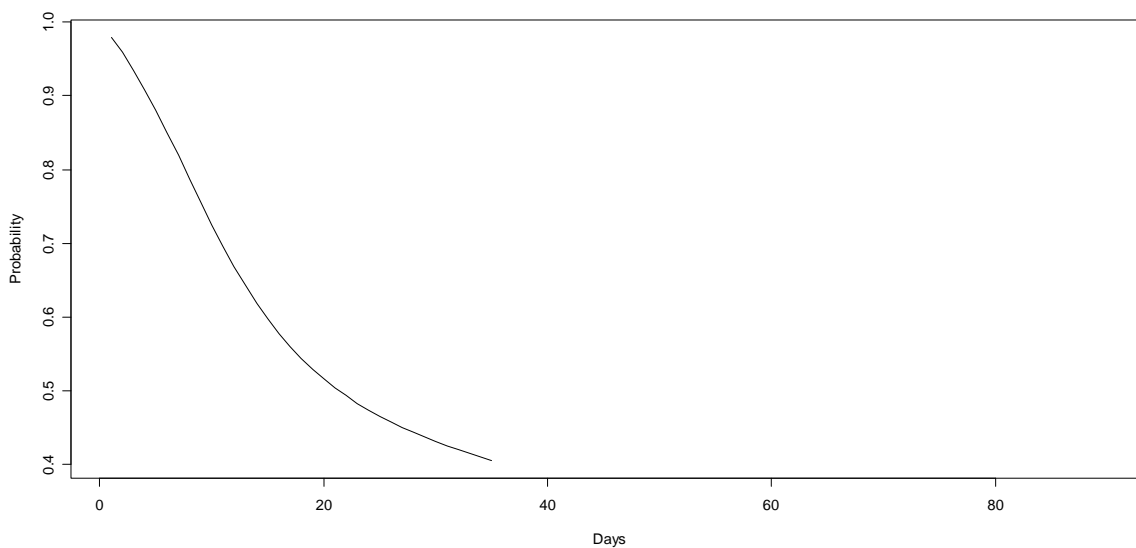
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el quart ingrès segons presenti o no el problema de insuficiència cardíaca:



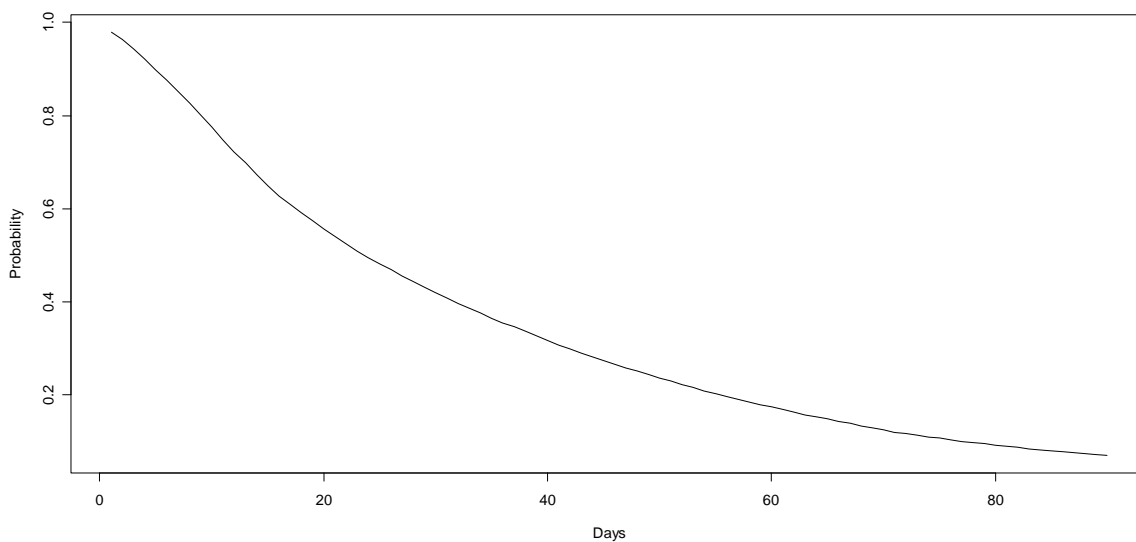
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el cinquè ingrès segons presentí o no el problema de insuficiència cardíaca:



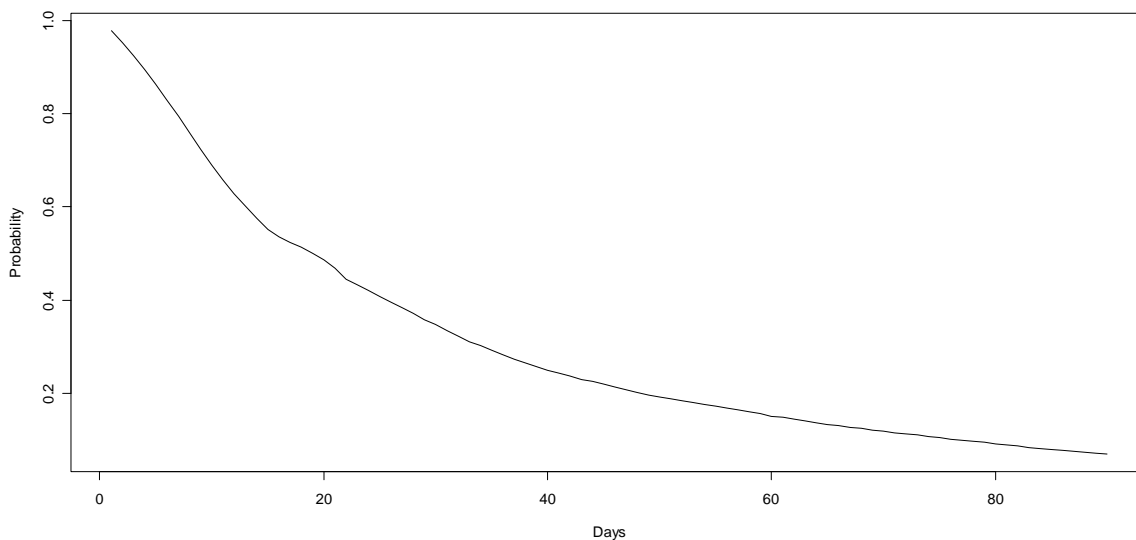
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el primer ingrès segons presentí o no el problema de dislipèmia:



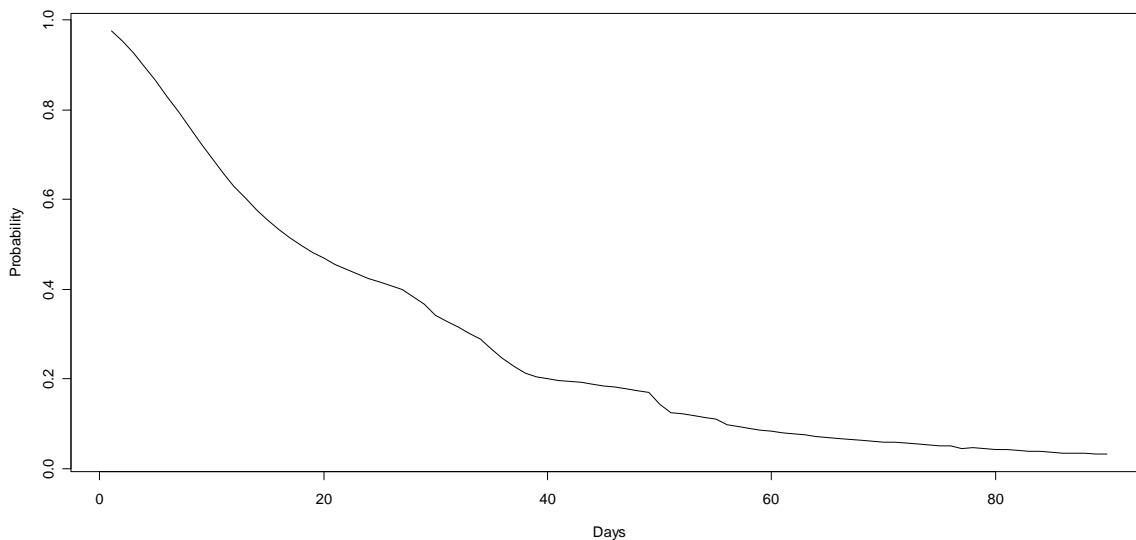
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el segon ingrès segons presenti o no el problema de dislipèmia:



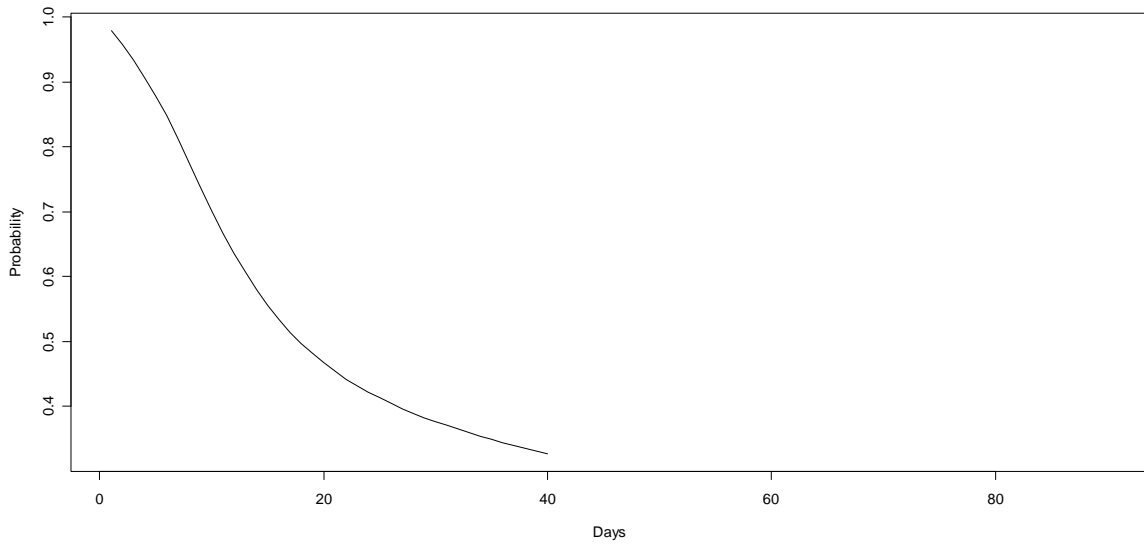
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el tercer ingrès segons presenti o no el problema de dislipèmia:



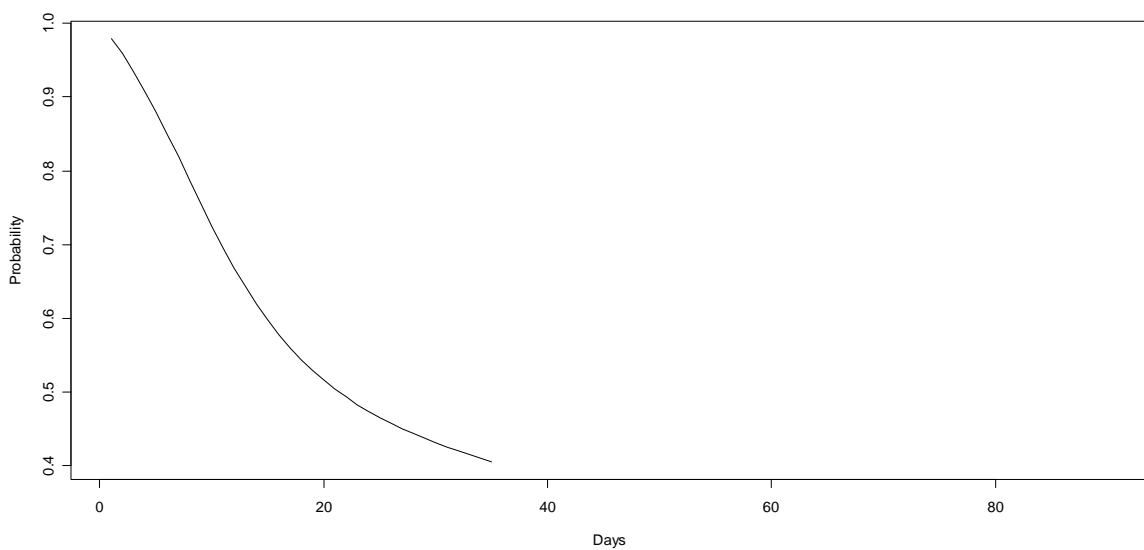
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el quart ingrès segons presenti o no el problema de dislipèmia:



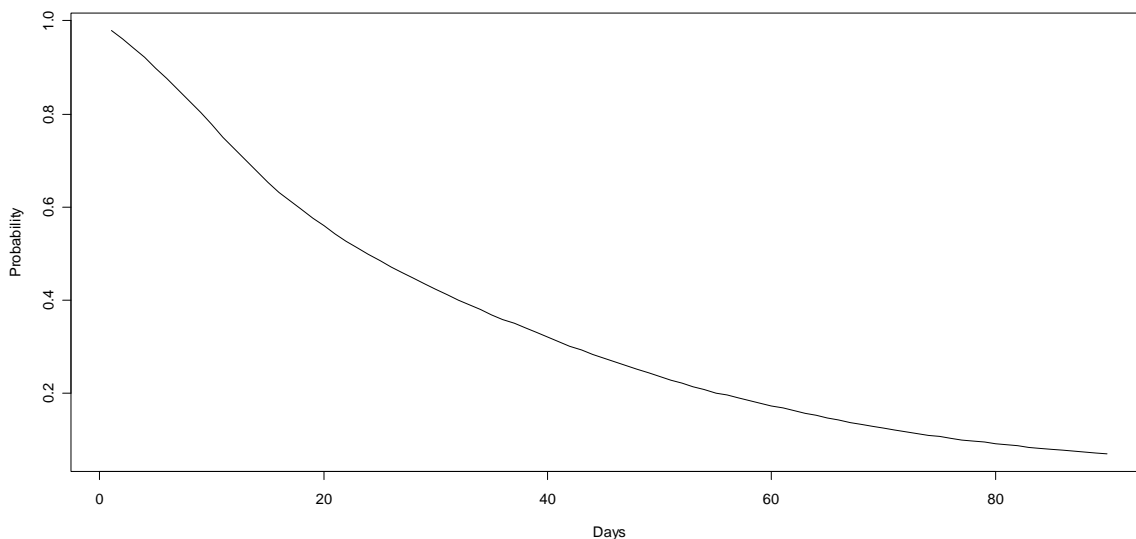
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el cinquè ingrès segons presenti o no el problema de dislipèmia:



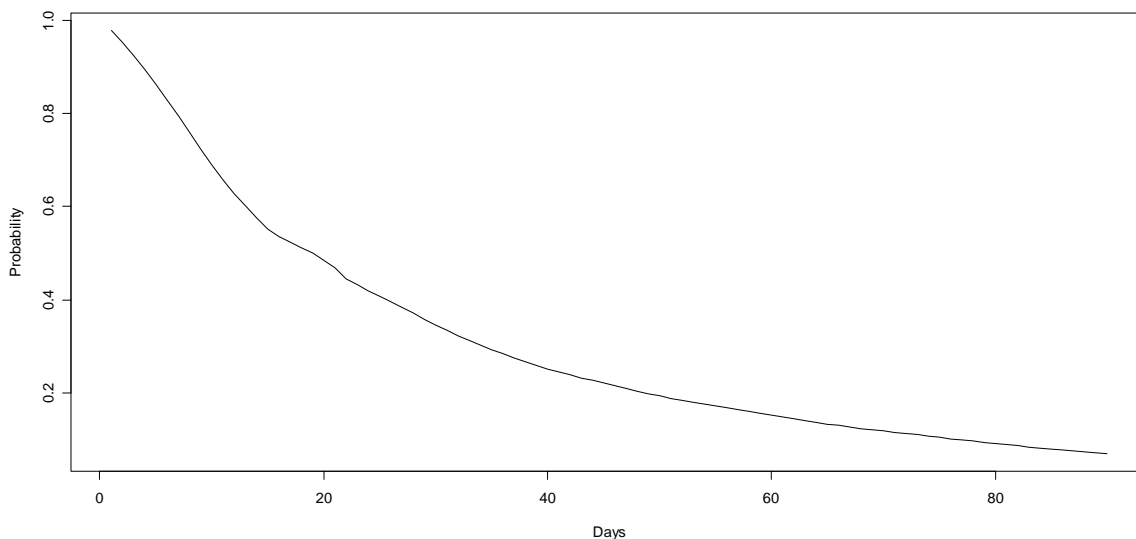
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el primer ingrès segons presentí o no el problema de diabetis mellitus 2:



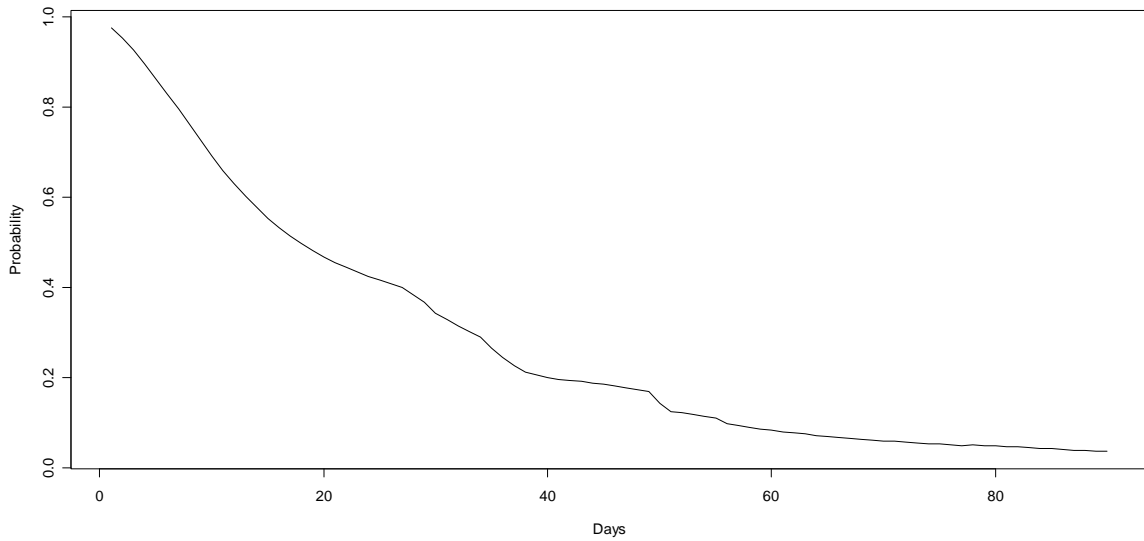
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el segon ingrès segons presentí o no el problema de diabetis mellitus 2:



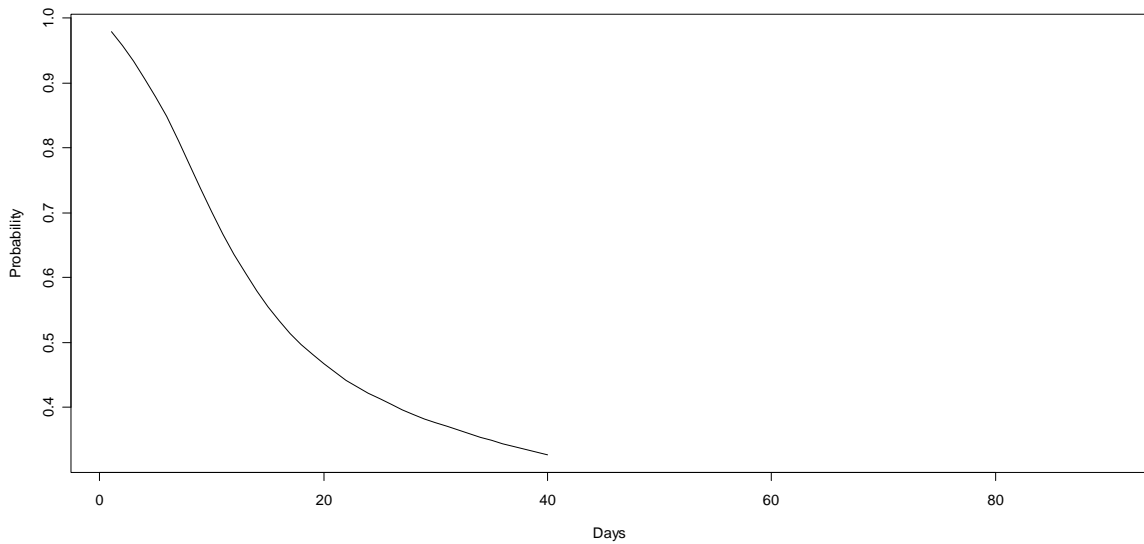
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el tercer ingrès segons presenti o no el problema de diabetis mellitus 2:



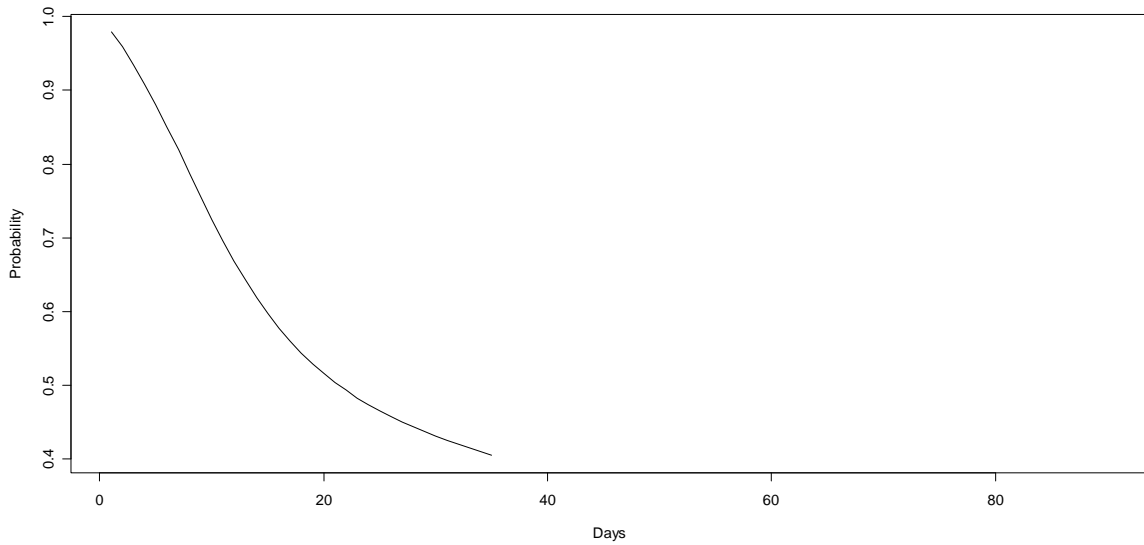
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el quart ingrès segons presenti o no el problema de diabetis mellitus 2:



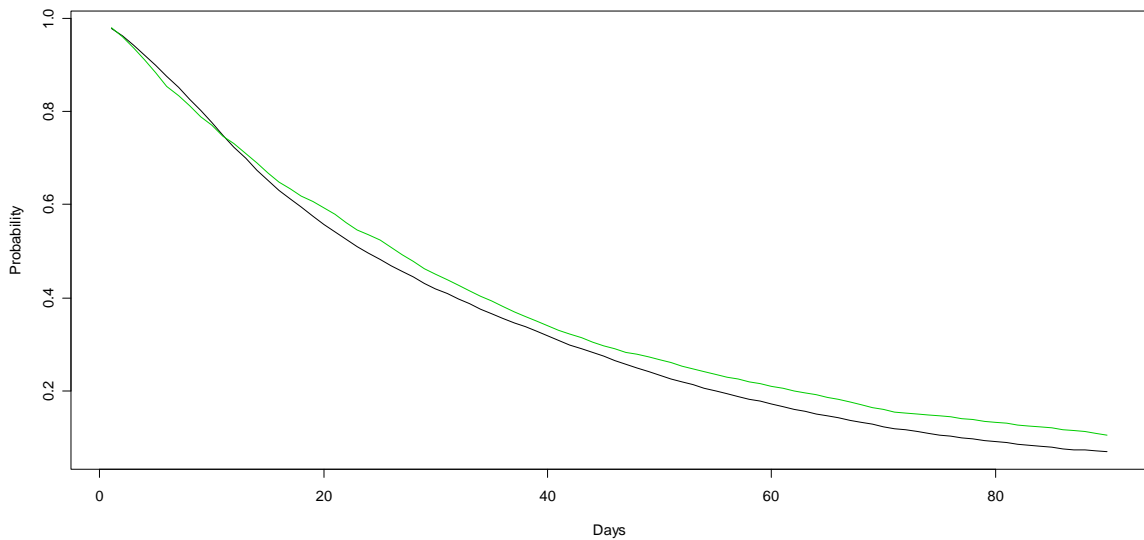
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el cinquè ingrès segons presenti o no el problema de diabetis mellitus 2:



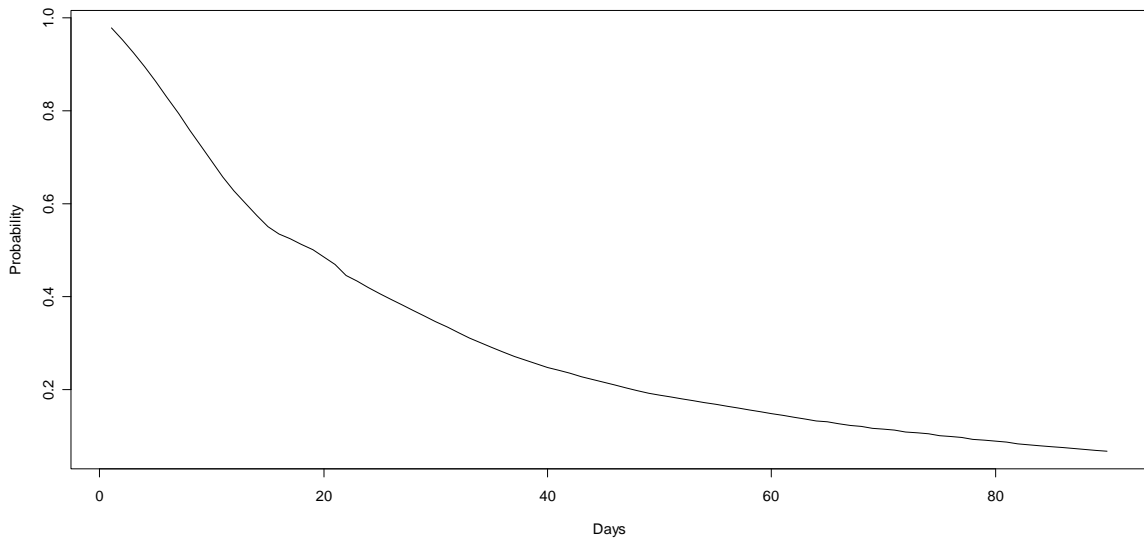
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el primer ingrès segons presenti o no el problema de tabaquisme:



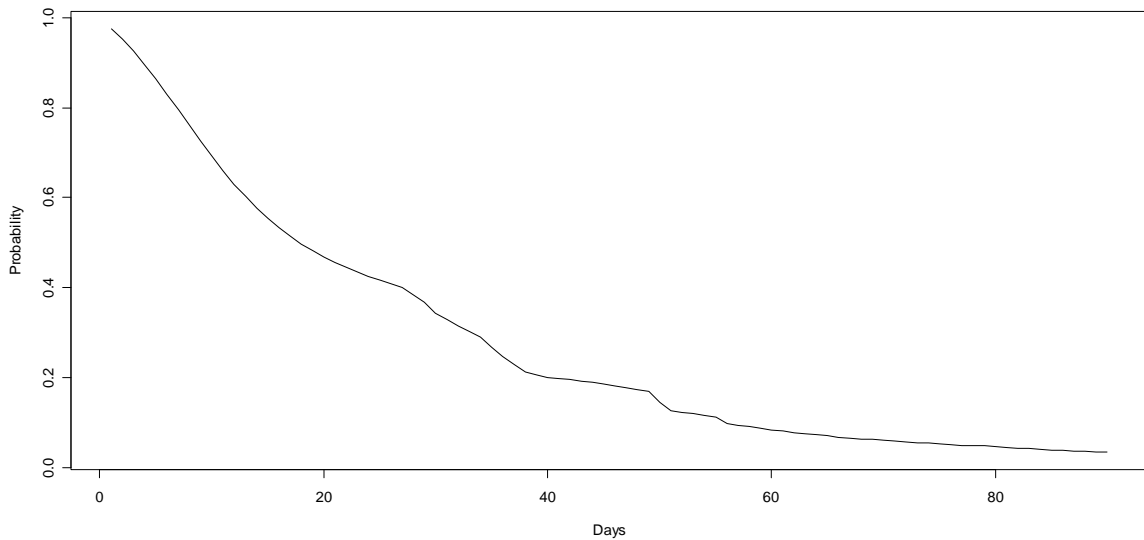
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el segon ingrès segons presenti o no el problema de tabaquisme:



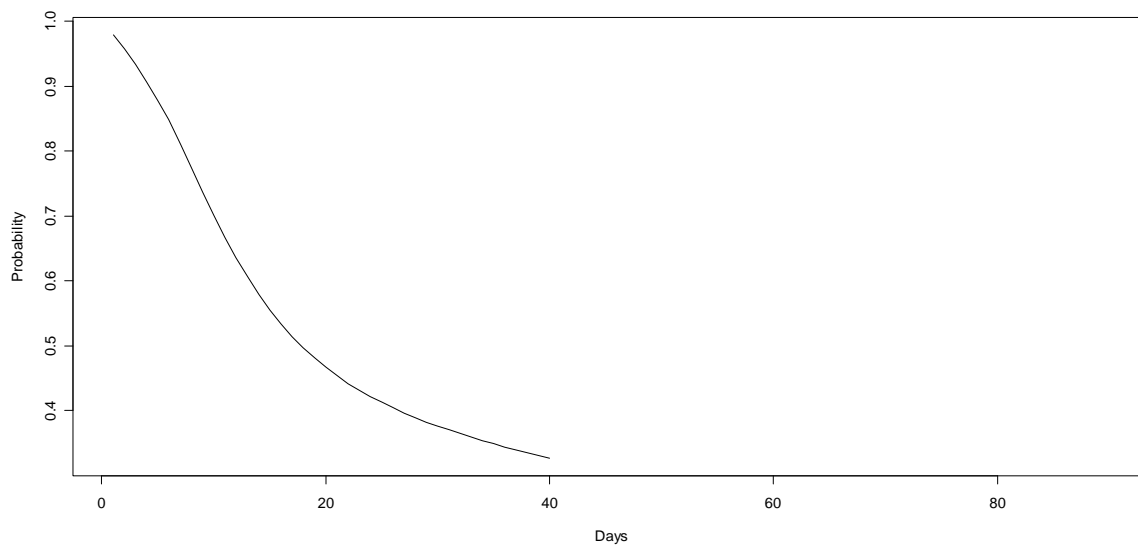
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el tercer ingrès segons presenti o no el problema de tabaquisme:



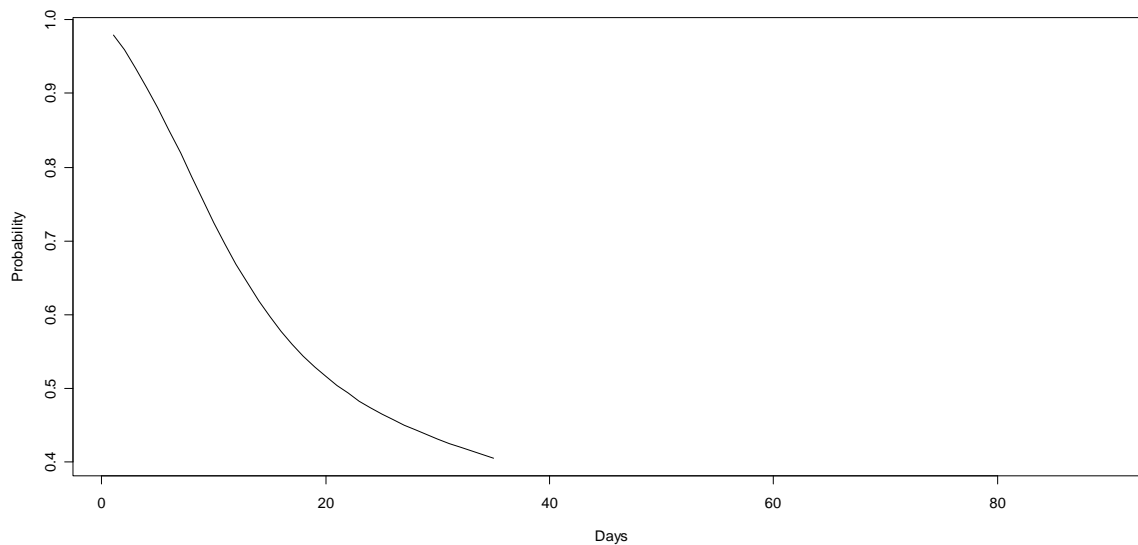
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el quart ingrès segons presenti o no el problema de tabaquisme:



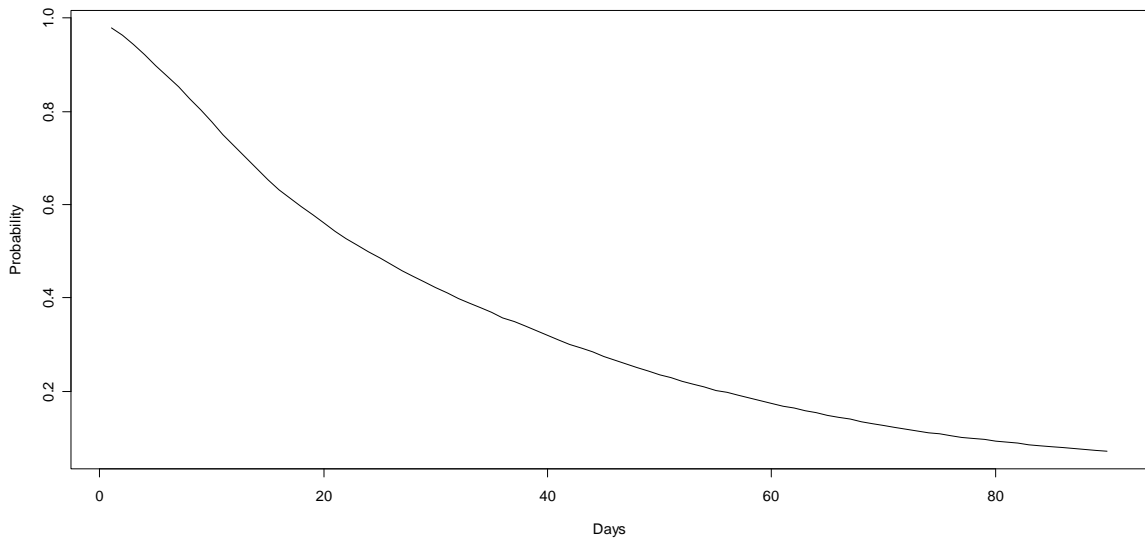
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el cinquè ingrès segons presenti o no el problema de tabaquisme:



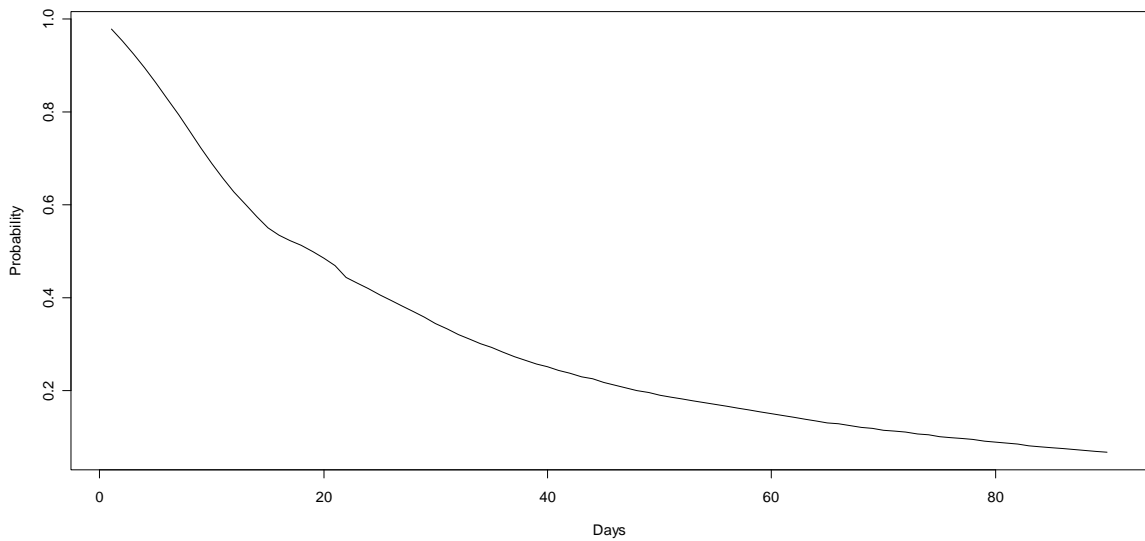
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el primer ingrès segons presenti o no el problema de extabaquisme:



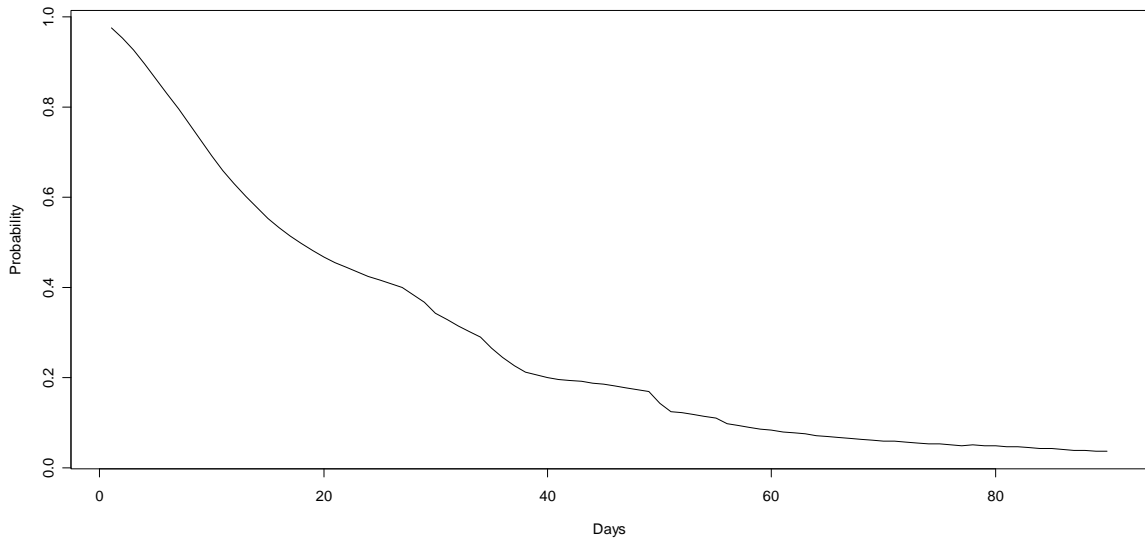
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el segon ingrès segons presenti o no el problema de extabaquisme:



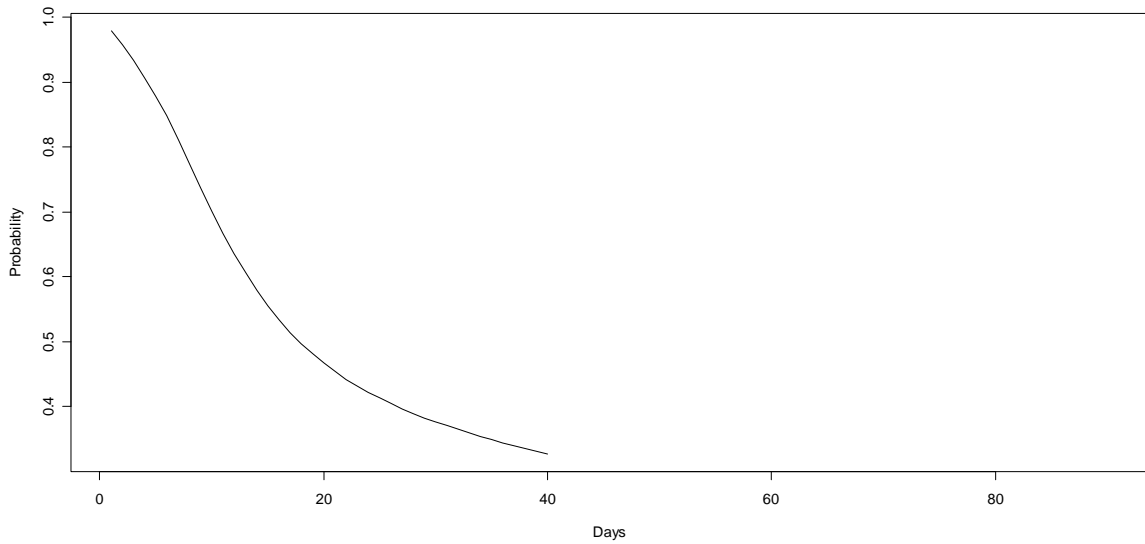
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el tercer ingrès segons presenti o no el problema de extabaquisme:



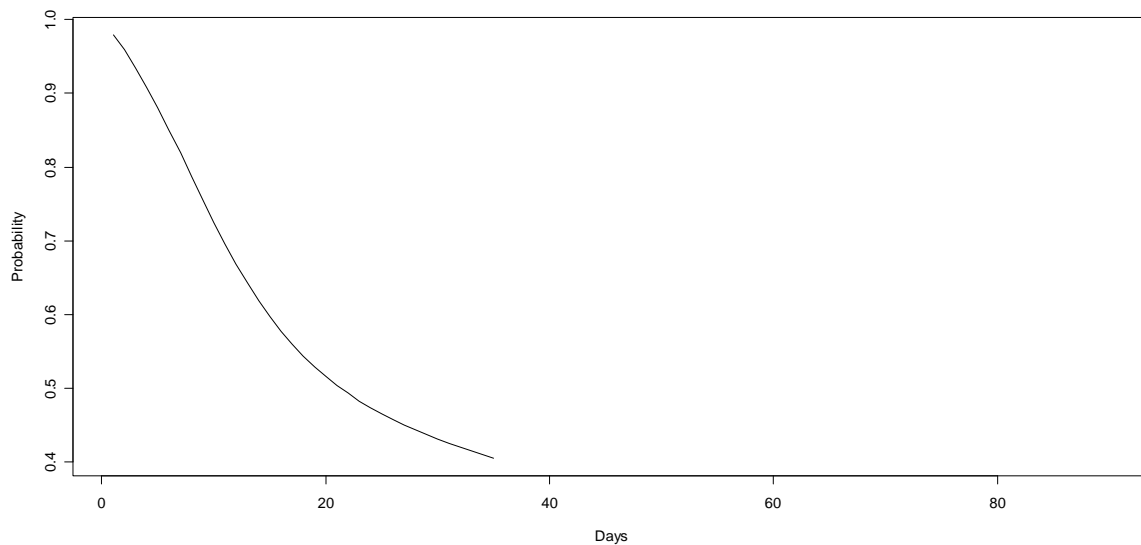
Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el quart ingrès segons presenti o no el problema de extabaquisme:



Negre: no presenta el problema
Verd: si presenta el problema

Gràfic de la probabilitat de supervivència per el cinquè ingrès segons presenti o no el problema de extabaquisme:



Negre: no presenta el problema
Verd: si presenta el problema

