PROJECTE FINAL DE CARRERA

Llicenciatura en Ciències i Tècniques Estadístiques

Títol: Supervivència de pacients amb càncer de pròstata localitzat amb models de riscos proporcionals

Autor: Mercè Viladrich Sarret

Director: Klaus Langohr

Departament: Departament d'Estadística i Investigació Operativa

Convocatòria: Juny 2015

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Facultat de Matemàtiques i Estadística

Facultat de Matemàtiques i Estadística
Universitat Politècnica de Catalunya

Projecte fi de carrera (PFC)

Supervivència de pacients amb càncer de pròstata localitzat amb models de riscos proporcionals

Mercè Viladrich

Director: Klaus Langohr

Departament d'Estadística i Investigació Operativa

Agraïments

Aquest projecte no hauria estat possible sense l'ajuda de molta gent i voldria agrair-los l'esforç dedicat al llarg d'aquests mesos.

M'agradaria agrair especialment al professor i director Klaus Langohr, la seva dedicació i implicació en aquest treball, ja que cada reunió ha representat nous coneixements i una gran dosi de motivació.

Així mateix, a la Dra. Montse Ferrer i l'Oriol Cunillera de l'IMIM per tots els coneixements que m'han trames en recerca oncològica, i al Grup Multicèntric Espanyol de Càncer de Pròstata Localitzat per cedir-me les seves dades per poder elaborar aquest projecte.

També, a Rosario Pelaez, companya de feina, per haver-se llegit el projecte i haver donat el seu punt de vista.

Finalment, agrair a la meva família tot el suport i la confiança que m'han donat en tot moment i donar-me la última empenta per acabar aquesta etapa.

A tots ells, gràcies.

Acrònims

BT	Braquiteràpia prostàtica
CP	Càncer de pròstata
CRAN	*Comprehensive R Archive Network*
EPIC	*Expanded Prostate Cancer Index Composite*
EVP	Esdeveniments per variable
FACT-G	*Functional Assessment of Cancer Therapy - general*
FACT-P	*Functional Assessment of Cancer Therapy - prostate*
GLL	Graus de llibertat
IC	Interval de confiança
IIEF	*International Index of Erectile Function*
INE	Institut Nacional d'Estadística
IPSS	*International Prostate Symptoms Score*
PSA	Antigen prostàtic específic
QV	Qualitat de vida
PR	Prostatectomia radical
RT	Radioteràpia externa

Índex

Índex de figures VI
Índex de taules VII
1. Introducció 8
1.1 Antecedents 8
1.2 Motivació 9
1.3 Càncer de pròstata 10
2. El conjunt de dades 14
3. Metodologia estadística 17
3.1 Anàlisi de la supervivència 17
3.1.1 Conceptes bàsics 18
3.1.2 Censura per la dreta 19
3.1.3 Estimació de Kaplan-Meier 21
3.1.4 Comparació de funcions de supervivència 22
3.2 Model de Cox de riscos proporcionals 24
3.2.1 L'expressió del model 24
3.2.2 Estimació dels coeficients del model 25
3.2.3 Nombre de variables incloses al model de Cox 27
3.3 Anàlisi de residus 28
3.4 Gran censura per la dreta 32
3.5 Llibreries d'R 33
4. Resultats 36
4.1 Anàlisi descriptiva 36
4.2 Anàlisi de la supervivència 37
4.3 Model de Cox de riscos proporcionals 41
4.4 Anàlisi dels residus 42
5. Discussió 47
5.1 Conclusions epidemiològiques 47
5.2 Conclusions estadístiques 48
5.3 Limitacions de l'estudi 48
Bibliografia 49
A. Codi R dels anàlisi realitzats 51
B. Corbes de supervivència addicionals 65

Índex de figures

Figura 1.1 Localització de la pròstata (Font: Cáncer de próstata: ¿Qué es el cáncer de próstata?. American Cancer Society. 2011) 10
Figura 1.2 Corva Kaplan-Meire (Font: R Documentation: Plot method for survfit objects) 22
Figura 4.1 Estimació de Kaplan-Meier i intervals de confiança del 95% de la funció de supervivència global 38
Figura 4.2 Corba de supervivència segons grup de tractament 39
Figura 4.3 Corba de supervivència segons grup de risc i grup de tractament 39
Figura 4.4 Corba de supervivència segons tractament hormonal previ i grup de tractament 40
Figura 4.5 Residu basats en la desviància 43
Figura 4.6 Residus basats en l'*score* 44
Figura 4.7 Residus de Schoenfeld per a l'estimació global del model 45
Figura 4.8 Residus de Schoenfeld per a cada variable inclosa al model 46
Figura 7.1 Corba de supervivència segons grup d'edat i grup de tractament…….65
Figura 7.2 Corba de supervivència segons *clinical T stage* i grup de tractament 65
Figura 7.3 Corba de supervivència segons tractament hormonal actual i grup de tractament 66
Figura 7.4 Corba de supervivència segons tractament hormonal de rescat i grup de tractament.....66

Índex de taules

Taula 4.1 Característiques basals de la mostra en estudi ……37
Taula 4.2 Comparació de les funcions de supervivència dels tractaments en diferents subgrups d'interès …… 40
Taula 4.3 Coeficients de correlació entre les variables d'interès..…… 41
Taula 4.4 Model de Cox amb les covariables d'interès ..…… 42
Taula 4.5 Model de Cox sense observacions influents …… 44
Taula 4.6 Proporcionalitat de riscos ..…….…….. 45

1. Introducció

1.1 Antecedents

El càncer de pròstata (CP) és la tercera causa de mort per càncer de la població masculina a l'estat espanyol[1], la segona de la Unió Europea[2] i la sisena a nivell mundial, essent la tercera en els països desenvolupats i la sisena en els subdesenvolupats[3]. La determinació de l'Antigen Prostàtic Específic (PSA) permet detectar de forma prematura molts tumors òrgan-confinats que fins fa poc temps passaven desapercebuts, fet que ha portat la migració del diagnòstic d'aquesta malaltia fins a estadis més precoços.

Els tres principals tractaments per al CP a Espanya són la prostatectomia radical (PR), la radioteràpia externa (RT) i la braquiteràpia prostàtica (BT). El primer tractament està considerat tradicionalment com el tractament preferent en els homes amb una esperança de vida superior als 10 anys, mentre que la RT es considera oportuna per aquells pacients en els que la cirurgia pot comportar un risc major, tot i que té efectes secundaris importants[4]. Un altre tractament disponible és la BT, que produeix menys danys als teixits que envolten el tumor i pot disminuir els efectes secundaris de la radioteràpia externa. Aquest fet, sumat als resultats publicats a partir de la dècada dels 90 on es conclou que els efectes d'aquest tractament en el control de la progressió del CP són comparables a la prostatectomia radical han fet que s'hagi estès el seu ús i que s'hagi plantejat la discussió sobre el criteri establert de l'ús de la prostatectomia radical com a primera opció de tractament per a aquest tipus de càncer[5].

La pràctica clínica habitual a Espanya inclou aquestes tres opcions, però l'elecció del tractament per part del pacients hi juga un paper poc decisiu. L'evidència disponible entre els diferents resultats dels tractaments, ja sigui mesurada en supervivència, símptomes, efectes secundaris o qualitat de vida, i la falta d'estudis que els comparin els tres tractaments en el context espanyol són dos dels factors més importants que juguen en contra d'aquest fet. En aquest context s'ha dut a terme un projecte multicèntric a nivell nacional per comparar l'impacte d'aquests tres tractaments en els pacients amb CP òrgan-confinat anomenat ***Efectividad a largo plazo de la prostatectomía radical, la braquiterapia y la radioteràpia conformacional externa 3D en el cáncer de próstata órgano-confinado*.**

1.2 Motivació

Interès epidemiològic i metodològic

El projecte que es duu a terme està dins del marc de l'estudi ***Efectividad a largo plazo de la prostatectomía radical, la braquiterapia y la radioterapia conformacional externa 3D en el cáncer de próstata órgano-confinado*****.** Els objectius d'aquest estudi són comparar l'efectivitat dels tres tractaments als 3 anys de seguiment en termes de qualitat de vida, supervivència lliure de progressió bioquímica i supervivència global, així com avaluar els efectes secundaris dels tractaments. La determinació de la supervivència global serà l'objectiu desenvolupat al llarg d'aquest document.

En l'àmbit epidemiològic, l'interès principal de la realització d'aquest estudi és la manca de referències prèvies a Espanya que comparin els tres tractaments principals per al CP, fet que porta a plantejar quin és el millor tractament per al control de la progressió de la malaltia. Els efectes secundaris produïts per la prostatectomia, tractament preferent fins a l'actualitat, són importants, mentre que tant els produïts per la radioteràpia com per la braquiteràpia han recuperat els nivells previs a la intervenció en 3 anys. Els resultats de l'estudi, a més, permetran realitzar comparacions amb estudis similars realitzats en països, degut a que s'utilitzen els mateixos instruments de mesura.

Malgrat l'elevada incidència de CP hi ha una manca important d'evidència científica de qualitat sobre l'efectivitat dels tractaments en tumors localitzats. Aquest fet sumat a que el càncer de pròstata es una de les principals causes de mortalitat en els homes a l'estat espanyol són les motivacions mèdiques principals que han conduït en la realització d'aquest document.

Interès personal

El fet de poder participar en un estudi d'àmbit nacional com aquest i poder comptar amb l'experiència del Grup Multicèntric Espanyol d'Investigació de Càncer de Pròstata Localitzat és una gran motivació personal per realitzar aquest treball.

En l'àmbit metodològic les principals motivacions estadístiques d'aquest projecte són la introducció i aprofundiment, per part meva, en l'anàlisi de la supervivència i tots els seus aspectes, tant teòrics com pràctics i la seva aplicació al programari lliure R, ja que resulta interessant aprofundir en les seves llibreries per poder millorar professionalment.

1.3 Càncer de pròstata

La pròstata és un glàndula exclusivament masculina que està ubicada davant del recte i per sota de la bufeta i la seva mida augmenta amb l'edat. La seva funció és produir una certa quantitat del líquid que protegeix els espermatozous[6].

El cos humà està format per milions de cèl·lules vives que creixen, es divideixen i moren d'una forma ordenada. Durant els primers anys de vida aquestes es divideixen més ràpidament per permetre el creixement. Un cop arribada l'edat adulta només es divideixen per substituir aquelles cèl·lules desgastades o mortes. El càncer s'origina quan en alguna part del cos aquestes cèl·lules comencen a créixer sense control enlloc de morir-se[7].

A la pròstata hi ha molts tipus de cèl·lules, però gairebé la totalitat dels CP es desenvolupen a partir de les cèl·lules de la glàndula, la funció de les quals és produir el líquid que s'agrega al semen. Aquest tipus de càncer s'anomena adenocarcinoma. Alguns CP poden créixer i propagar-se molt ràpid però la gran majoria es desenvolupen lentament[6].

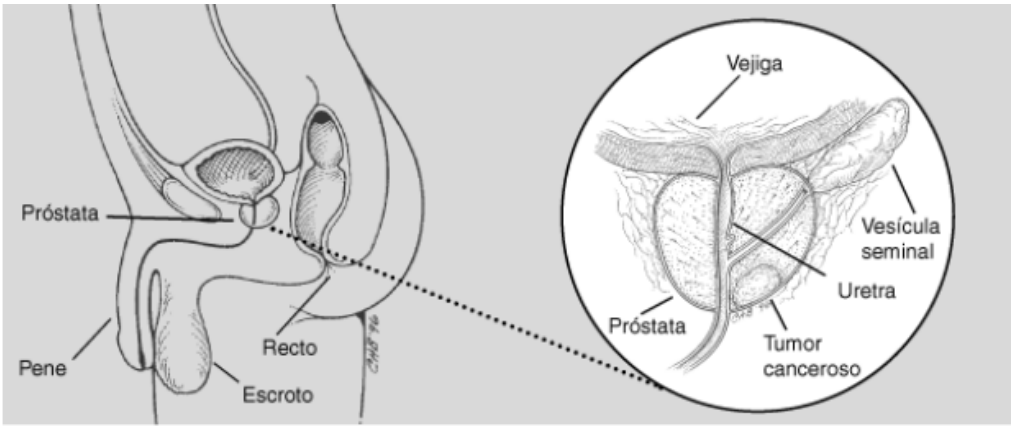

Figura 1.1 Localització de la pròstata (Font: Cáncer de próstata: ¿Qué es el cáncer de próstata?. American Cancer Society. 2011)

El càncer de pròstata és un dels tumors diagnosticats amb més freqüència en homes, amb una taxa ajustada d'incidència estimada a Espanya de 45 casos per 100.000 homes/any i a la Unió Europea de 68 casos per 100.000 homes/any[8]. Aquesta incidència està relacionada estretament amb l'edat ja que més de la meitat dels CP són diagnosticats en homes majors de 65 anys, tot i que aquests moriran per altres causes no relacionades amb el tumor[3].

La prevalença del CP a un any en la població espanyola és de 15,5% i de 15,2% a cinc anys[6]. Aquesta xifra ha augmentat considerablement en els últims anys degut a l'increment de l'esperança de vida de la població i a la utilització del PSA com a eina de diagnòstic, fet que ha permès que es detectessin tumors que haurien estat clínicament indetectables[9]. Aquest paràmetre

ens permet classificar els individus en diferents grups, aquells que tinguin nivells inferiors o iguals a 10 ng/ml pertanyen al grup de risc baix, aquells valors entre > 10 ng/ml i < a 20 ng/ml pertanyen al risc mitjà i aquells que els seus valors són superiors o igual a 20 ng/ml es consideren de risc alt.

La Prostatectomia Radical (PR) consisteix en l'extirpació de la glàndula prostàtica i les dues vesícules seminals, és una tècnica que permet revisar els ganglis limfàtics per comprovar si s'ha produït una extensió del tumor i poder realitzar l'extirpació si és el cas. Sempre que sigui possible s'intenta preservar els nervis, fet que sinó pot produir impotència a llarg termini. La radioteràpia (RT) permet aplicar dosis més altes de radiació a la pròstata i dosis més baixes a la zona que la rodeja, reduint la gravetat dels efectes secundaris. Per acabar la braquiteràpia (BT) consisteix en la implantació d'una "llavor" radioactiva permanent a la pròstata, fet que fa que el tractament sigui mínimament invasiu[10].

La manca de definicions clares, períodes variables per a l'avaluació de la toxicitat i la falta d'estudis comparatius donen com a resultat un ventall d'efectes secundaris molt ampli. Aquesta amplitud també és deguda a l'edat dels pacients, el tipus de tractament i la tècnica quirúrgica. Tot i que existeix aquest fet figura un patró que ens mostra que la cirurgia és l'opció de tractament amb més risc d'impotència i pèrdues d'orina; que la radioteràpia externa provoca més risc de problemes intestinals; i que la braquiteràpia és l'opció terapèutica amb un major risc de símptomes d'irritació i obstrucció del sistema urinari. Cap de les tres opcions de tractament és superior a les altres en el control del càncer clínicament, per tant a priori no es pot decidir quina és millot o pitjor, l'única forma de fer-ho es mesurant la intensitat i repercussió dels efectes secundaris, per això a les nombroses guies existents s'indica que l'elecció del tractament per part del pacient en CP de baix risc és primordial[5].

En el CP precoç la majoria d'estudis publicats comparen només dos tractaments o dues conductes. En l'estudi de Bill-Axelson et al. (2005)[11] es comparen resultats de PR amb la conducta expectant. Es van incloure 347 homes als quals se'ls va realitzar una PR i 348 amb els quals es va tenir una actitud expectant. Pel que fa a l'evolució de la progressió es mostren diferencies estadísticament significatives entre els dos grups als cinc anys de seguiment (8,1% en PR vs 27,2% en conducta expectant) i augmenta al llarg del seguiment (19,2% en PR vs 44,3% en conducta expectant). En quant a la mortalitat per CP als cinc anys de seguiment va ser del 2,3% en els homes amb PR i del 4,3% en els de conducta expectant. Als 10 anys de seguiment aquestes xifres van augmentar al 9,6% i 14,9% respectivament.

Si es volen veure resultats dels tres tractaments que s'estudiaran en les publicacions de referència només és donen dades sobre PSA. En l'estudi D'Amico (1998)[12] l'objectiu era estudiar el control del PSA en els tres tractaments estudiats en el nostre estudi i per poder donar

resposta s'agrupaven els pacients en risc baix, moderat i alt segons els factors pronòstic establerts per la literatura: nivell inicial de PSA, el Gleason Score en la biòpsia i el AJCC T-Stage de 1992. Els pacients amb risc baix després d'implantar-los el tractament s'estima que tenen els mateixos beneficis als 5 anys en els tres grups de tractament, a més a més l'addició d'un tractament neodajuvant en la BT no mostra cap efecte beneficiós addicional. En segon lloc els pacients amb risc moderat després d'implantar-los el tractament es van mostrar resultats pitjors en els tractats amb BT en comparació amb els tractats amb PR, RT o BT amb suplement de tractament neoadjuvant. Aquestes diferències es veuen gràficament, però l'estudi te una limitació en la potència estadística que no es permet detectar aquestes diferències. Pel que fa als pacients amb un risc alt als 5 anys de seguiment observem que hi ha pitjors resultats en els tractats amb BT ja sigui amb suplement neodajuvant o no enfront als pacients tractats amb PR o RT. Una recomanació d'aquest estudi és poder realitzar un seguiment molt més llarg, per poder observar què passa amb aquests pacients després dels cinc primers anys d'haver-los implantat el seu tractament per al CP. Tot i que recomana que es tracti com un estudi preliminar suggereix que als pacients amb un alt risc, que tenen una major necessitat de tractament, el tractament amb BT, ja sigui amb suplement o no de tractament neoadjuvant, pot ser més ineficaç que el tractament amb PR o RT. D'Amico (1998)[12] obre la porta a continuar estudiant l'efecte dels tres tractaments tant a nivell de PSA com en tots els aspectes clínics del CP.

En l'estudi de Stokes (2000)[13] l'objectiu era avaluar a llarg termini la supervivència lliure de progressió estratificat per nivell de risc inicial de PSA en pacients tractats amb PR, RT i BT sense addició de tractament neodajuvant. Els resultats d'aquest estudi mostren que entre els pacients amb risc baix o mitjà no existeixen diferències en la supervivència lliure de progressió, aquests resultats van en la línea de l'estudi D'Amico (1998)[12]. En canvi, s'obtenen millors resultats entre els pacients amb risc alt tractats amb PR en comparació amb la RT i la BT.

En la revisió de bibliògrafa podem veure que hi ha un gran nombre d'estudis que avaluen la supervivència però deixen de banda la resta d'aspectes clínics. En aquest punt els estudis de qualitat de vida (QV) permeten avaluar els efectes del tractament a partir de la perspectiva del pacient. En aquest sentit, en l'avaluació de la QV al mes de l'inici del tractament els pacients tractats amb BT i RT presentaven una pitjor qualitat de vida en contrapartida a l'any de seguiment que els pacients tractats amb PR presentaven pitjors resultats en aquest sentit[5]. A partir d'aquí es plantejava el problema de la manca d'un període més llarg de seguiment per detectar millor l'impacte dels efectes secundaris, fet que s'ha intentat donar resposta en el marc global de l'estudi. Els resultats que s'han publicat fins al moment mostren que, a nivell basal, només s'observaren diferències entre tractaments en l'índex EPIC (*Expanded Prostate Cancer Index Composite*) en la puntuació referent a l'aparell intestinal i als dos anys de seguiment s'observaren diferencies entre tractaments en la disfunció sexual, urinària i intestinal en els

ítems del mateix qüestionari[14]. Tot i la importància d'aquesta mesura, aquest fet no serà objecte d'estudi en el present document.

Pel que fa a la supervivència global, raó d'ésser d'aquest document, en l'assaig aleatori realitzat per Holmberg et al. (2002)[15] on es compara el tractament amb PR i la conducta expectant a 8 anys de seguiment no s'observen diferències estadísticament significatives en la supervivència global dels pacients inclosos en ambdós grups. En l'estudi d'Amico et al. (2004)[16] on s'estudia l'efectivitat del tractament amb RT comparat amb aquest mateix tractament però amb 6 mesos de supressió de teràpia amb andrògens s'observà una supervivència global significativament major en el segon grup en comparació amb el primer a 5 anys de seguiment 88% (IC 95%, 80%-95%) vs 78% (IC 95%, 68% - 88%), respectivament. El fet de no disposar de bibliografia que compari la supervivència global dels tres tractaments en estudi és la raó principal per desenvolupar aquest document.

L'estructura d'aquest document consta, en primer lloc l'apartat del conjunt de dades de l'estudi on es fa un breu resum de les dades obtingudes i les variables recollides. En el següent apartat de metodologia estadística es presenten tots els conceptes que s'han utilitzat per dur a terme l'anàlisi de les dades. A continuació es presenten els resultats obtinguts, tant l'anàlisi descriptiva de la mostra, com les diferents corbes de supervivència d'interès per als clínics, així com el model de Cox estudiat i l'anàlisi dels seus residus. Per acabar es presenten les conclusions tant a nivell d'estudi com a nivell personal.

2. El conjunt de dades

La practica clínica habitual per al CP a Espanya inclou la PR, la RT i la BT. En el moment del disseny de l'estudi no hi havia cap estudi d'àmbit nacional que comparés aquests tres tractaments. Aquest fet va motivar el disseny d'un estudi prospectiu d'àmbit nacional de seguiment als 10 anys després de l'inici del tractament de pacients diagnosticats de CP òrgan-confinat per comparar l'impacte d'aquests tractaments en la qualitat de vida i en el temps de supervivència dels pacients amb CP.

L'objectiu general d'aquest estudi multicèntric és comparar l'efectivitat dels tractaments per al CP als 3, 5 i 10 anys de seguiment. Com a objectius específics destaquem la comparació de la qualitat de vida en els tres grups de tractament, la presència dels principals efectes adversos en els tres grups de tractament i la comparació de la supervivència global i la supervivència lliure de progressió bioquímica en els tres grups de tractament.

L'estudi consta d'una visita pre-tractament, una visita al mes 1, 3, 6, 12 i 24 després de l'inici de tractament i als 3, 5 i 10 anys de seguiment on es recullen l'estat del pacient, els valors de PSA, les complicacions urinàries i intestinals, el tractament per la CP, aparició de recidives, tipus i tractament d'aquestes i una bateria de qüestionaris de qualitat de vida: qüestionari de salut general S-F36, qüestionari de l'impacte funcional del tractament oncològic (FACT-G i FACT-P), el qüestionari EPIC, la Puntuació Internacional de Símptomes Prostàtics (IPSS) i l'Índex Internacional de la Funció Erèctil (IIEF). El fet de recollir aquestes variables fa que els resultats d'aquest estudi es puguin comparar amb resultats internacionals mitjançant l'estudi de referència SPIRIT[17] que compara la PR amb la BT i així poder comparar els resultats a nivell espanyol amb resultats d'altres països com ara Estats Units i Canadà entre d'altres.

Els pacients inclosos en l'estudi van ser reclutats seqüencialment entre l'abril del 2003 i el març del 2005 en consultes externes d'11 departaments hospitalaris espanyols: dos d'Andalusia, sis de Catalunya, un de Galícia, un de Madrid i un del País Basc.

Per poder obtenir dades fiables de la mortalitat dels pacients en estudi dels quals s'havia perdut el seguiment es demana un creuament amb una base de dades de mortalitat del *Insituto Nacional de Estadística* (INE). El problema és que aquesta base de dades s'actualitza amb més d'un any de retràs i les dades del 2013 s'han fet públiques al febrer del 2015[18]. Aquest interval de temps, per aquells pacients que són pèrdua de seguiment es supleix confrontant-los un per un amb les

dades de la seguretat social i així ens permet tenir el màxim d'actualitzat el nombre de pacients que esdevenen èxitus.

A continuació detallem les variables disponibles per l'anàlisi:

- Identificador del pacient: és un valor únic que ens permet identificar el pacient amb les dades procedents del seu centre
- Última data de seguiment disponible
- Indicador de censura: variable que pren valor 1 si el pacient va morir i 0 si està viu al final de l'estudi
- Volum de la pròstata a l'inici del estudi
- Puntuació de Gleason, basada en la biòpsia o la cirurgia
- PSA: Anàlisi de l'antigen prostàtic específic a l'inici del estudi
- Categoria T de l'estadiatge TNM: valor referent a la classificació del tumor primari. Pot prendre valors del 0 al 4 (0 = sense evidència de tumor primari, 1 = tumor no evident clínicament, 2 = tumor confinat a la pròstata, 3 = tumor que s'estén més enllà de la càpsula de la pròstata, 4 = tumor fixe o que envaeix estructures adjacents diferents de les vesícules seminals.
- Número d'història clínica
- Tractament que ha rebut el pacient per al càncer de pròstata. Pren valors de 1 a 3 (1 = Prostatectomia radical, 2 = Radioteràpia, 3 = Braquiteràpia)
- Data d'inici de tractament, o en el cas de la prostatectomia data de la cirurgia
- Grup de risc: variable calculada per la investigadora principal a partir de la puntuació de Gleason, el valor de PSA i l'estadiatge T:
    - Grup de risc baix: T1c o T2a, PSA$\leq$10 i Gleason<6
    - Grup mig: T2b, PSA entre 11 i 20 i Gleason=7
    - Grup alt: T2c, PSA>20 i Gleason>7
- Temps des de l'inici del tractament fins a la mort en el cas dels èxitus i temps des de l'inici del tractament fins a la última data de seguiment per als pacients vius
- Neoadjuvant: tractament hormonal previ a l'inici del tractament en estudi

- Tractament hormonal de rescat: tractament rebut pel pacient després de la detecció d'una recidiva a partir de l'inici del tractament en estudi
- Adjuvant: tractament hormonal com a complement del tractament en estudi
- Data d'inici del tractament hormonal de rescat
- Edat del pacient a l'inici de l'estudi

3. Metodologia estadística

En aquest punt es detallen els conceptes bàsics que hem utilitzat per poder dur a terme l'anàlisi del projecte. En els apartats de l'anàlisi de la supervivència i del model de Cox per a riscos proporcionals ens hem basat en els llibres Anàlisi de supervivència[19] i *Survival analysis. A self-learning text*[20,]i ho hem complementat amb altres lectures d'interès[21,22,23]

Al final del capítol s'expliquen les funcions del software R[24] que hem utilitzat per analitzar les dades.

3.1 Anàlisi de la supervivència

Els orígens de l'anàlisi de la supervivència es remunten a l'ús de les taules de vida, en les que es representaven el temps de mort en una població en funció de l'edat. Les primeres aplicacions es troben en estudis actuarials tot i que a partir de la segona meitat del segle XX s'ha desenvolupat al voltant de les aplicacions clíniques per poder analitzar la duració entre dos successos. També trobem nombroses aplicacions en l'àmbit industrial, anàlisi de fiabilitat, i en l'àmbit de l'economia i la sociologia, anàlisi de duracions, entre d'altres. En resum, l'anàlisi de supervivència estudia el temps fins un esdeveniment d'interès i els factors que hi intervenen.

El principal ús de l'anàlisi de la supervivència és representar la duració entre dos successos. El cas més senzill és considerar el naixement com a primer esdeveniment i la mort com a segon. La característica més important és que la resposta és el temps, i per tant no es mesura igual que la resta de variables. Donat que el temps entre dos successos es mesura seqüencialment és possible que no s'arribi a mesurar el temps fins a l'esdeveniment i com a conseqüència la informació sobre la supervivència està incompleta per alguns individus. En aquests casos es parla de dades censurades. La majoria dels mètodes utilitzats en l'anàlisi de la supervivència assumeixen que els individus censurats estan subjectes a la mateixa probabilitat de presentar l'esdeveniment que els individus que resten a l'estudi, es a dir, que la distribució del temps residual entre els individus censurats i no censurats sigui la mateixa, aquesta condició s'anomena condició de no informativitat.

A continuació es detallen els conceptes més importants de l'anàlisi de la supervivència i els models més utilitzats en aquests tipus d'anàlisis.

3.1.1 Conceptes bàsics

Denotarem T el temps fins al succés d'interès, ε. Aquest esdeveniment en el nostre estudi és la mort, tot i que generalment pot ser el desenvolupament d'una malaltia, l'embaràs, etc. Formalment, T és una variable aleatòria no negativa que correspon a una població homogènia. La distribució o el model per a T es pot caracteritzar per diferents funcions: la funció de supervivència (*S(t)*), la funció de distribució (*F(t)*), la funció de densitat de probabilitat (*f(t)*), la funció de risc ($\lambda(t)$), la funció de riscos acumulats ($\Lambda(t)$) o la vida mitja residual (*vmr(t)*). Cadascuna de les funcions anteriors serveix per poder il·lustrar diferents aspectes de T. Si coneixem qualsevol d'aquestes funcions les altres 5 es poden determinar de forma unívoca. En aquest punt ens centrarem amb al funció de supervivència i la funció de risc, ja que són les més utilitzades.

Funció de supervivència

La funció de supervivència (*S(t)*) es defineix com la probabilitat de que un individu sobrevisqui més de t unitats de temps, és a dir, la probabilitat de que el succés ε passi després de t:

$$S(t) = Prob(T>t), \text{ per a } t\geq 0.$$

Les propietats bàsiques d'aquesta funció són:

1. S(0)=1 i S(∞)=0
2. *S(t)* és una funció monòtona decreixent
3. Si T es contínua, *S(t)* es contínua i estrictament decreixent
4. *S(t) = 1-F(t)*

Funció de risc

La funció de risc ($\lambda(t)$) es pot interpretar com la probabilitat de que passi un succés en un interval de temps petit, donat que no hagi passat l'esdeveniment a l'inici del mateix. Si T és una variable contínua es definirà formalment com:

$$\lambda(t) = lim_{\Delta t \to 0} \frac{1}{\Delta t} Prob(t \leq T < t + \Delta t | T \geq t).$$

Aquesta funció expressa com el risc instantani canvia amb el temps i conté la mateixa informació que la funció de supervivència però en termes de la seva velocitat de canvi, quan la funció de supervivència decreix ràpidament el risc és alt mentre que si la funció de supervivència és constant el risc és 0.

Algunes propietats bàsiques d'aquesta funció són:

1. $\lambda(t)$ és una funció no negativa
2. Si T és contínua $\lambda(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} lnS(t)$

La funció de risc pot prendre moltes formes, un risc creixent correspon a una població que envelleix, com ara el temps de vida de pacients amb una leucèmia que no responen al tractament. Un risc decreixent correspon a una població els individus de la qual s'enforteixen amb el pas del temps, com ara el temps de vida dels pacients després d'una intervenció quirúrgica. Una funció de risc amb forma de banyera és pròpia de les poblacions que es segueixen des del naixement. La funció de risc amb forma de gepa correspon a una funció que creixen molt al principi i al llarg del temps comencen a decréixer, com ara la supervivència després d'un trasplantament. Quan la funció de risc és constant es denomina taxa de risc.

3.1.2 Censura per la dreta

Una de les característiques més importants dels estudis de supervivència és que és possible que al final del període d'observació no tots els pacients hauran presentat el succés objecte d'estudi, en el nostre cas la mort. També s'hi afegeix el fet que hi hagi hagut pèrdues de seguiment per alguns individus per causes diverses, com pot ser el canvi d'hospital, de país... Aquest fet l'anomenem censura i és una de les conseqüències de que el temps es mesuri seqüencialment. Els individus censurats seran analitzats des del moment de l'entrada a l'estudi fins l'última visita disponible i en la que no havien presentat el succés.

En els estudis que es vol explicar el temps fins al succés els individus són seguits fins que succeeix l'esdeveniment d'interès ε. Si aquest succés s'observa quan passa, el temps fins a ε és conegut. En canvi, si per un individu el succés ε no ha ocorregut durant el seguiment de l'estudi, direm que aquest individu està censurat per la dreta. Aquesta censura es pot donar per diverses causes, com ara la finalització de l'estudi en un moment determinat, la pèrdua de seguiment dels individus, la interrupció del tractament d'alguns individus, l'esdeveniment per una causa diferent a la d'interès de l'estudi (aquestes dues últimes opcions no es poden donar en el nostre cas).

Quan els individus entren en l'estudi en moments diferents E_i però s'estableix una data en la que es tancarà l'estudi es coneix aquest fenomen com a censura per la dreta de tipus I generalitzada. Aquest tipus de censura es el que esdevé en el nostre estudi.

Altres tipus de censura per la dreta són la censura de tipus I fixa, que es dona quan per cada individu, l'esdeveniment és observat si ocorre abans que un temps especificat prèviament. La censura de tipus I progressiva que s'observa quan el temps de censura és diferent per a cada individu. Finalment, la censura de tipus II es dona quan es decideix acabar l'estudi després que hagin ocorregut un nombre determinat d'esdeveniments.

El truncament, en canvi, és una condició que consisteix en filtrar la presència de certs individus de manera que l'investigador no en coneix la seva existència. En aquest cas, la inferència s'haurà de realitzar condicionada a aquest fet, ja que a diferència de la censura, que ens informa sobre el temps de l'esdeveniment, amb el truncament no coneixem l'existència dels individus que el presenten. En el nostre estudi no tenim individus truncats.

El tipus de censura, el tipus de dades i el truncament s'han de tenir en compte a l'hora de construir la funció de versemblança, ja que aquests dos factors definiran la contribució de cada individu a la funció de versemblança. També s'ha de tenir en compte la independència entre el temps d'interès i la censura. La versemblança per a una mostra d'individus s'escriu com un producte de la contribució per a cada individu.

Definim C com la variable aleatòria dels temps de censura, llavors

$$Y = min(T,C)\, i\, \delta = \begin{cases} 1\ Y = T \\ 0\ Y = C \end{cases}$$

on Y és la variable aleatòria que defineix el temps observat i δ la variable indicadora de si és un individu censurat o no.

Donada una mostra (Y_1,δ_1), ..., (Y_n,δ_n) i suposant la independència entre T i C, definirem la funció de versemblança com

$$\mathcal{L} = \prod_{i=1}^{n} \left\{ [f(y_i)]^{\delta_i} [S(y_i)]^{1-\delta_i} \right\}$$

el que és equivalent a

$$\mathcal{L} = \prod_{i=1}^{n} \left\{ [\lambda(y_i)]^{\delta_i} [S(y_i)] \right\},$$

ja que $f(y)=\lambda(y)S(y)$.

3.1.3 Estimació de Kaplan-Meier

Els mètodes estadístics més utilitzats en l'anàlisi de la supervivència són els no paramètrics, com ara l'anàlisi actuarial o el mètode de Kaplan-Meier. En aquest punt ens centrarem en el mètode de Kaplan-Meier, ja que és el que utilitzarem per analitzar les nostres dades. Aquest mètode el van proposar Kaplan i Meier al 1958 i es basa en la descomposició de la funció de supervivència en un producte de probabilitats condicionades.

Sigui T_1, T_2,..., T_n una mostra de la població d'interès i C_1, C_2,..., C_n variables aleatòries independents i idènticament distribuïdes amb funció de distribució G. Suposem que la censura no és informativa, és a dir, que els temps de censura C_1, C_2,..., C_n no estan relacionats amb els temps potencials de mort T_1, T_2,..., T_n. En altres paraules, suposem que tots els individus censurats tenen el mateix risc de presentar una mort futura que tots aquells individus vius i no censurats.

Donat el parell de dades observades Y_i i δ_i descrites a l'apartat 3.1.2 i n_i el nombre d'individus vius just abans de Y_i, definim l'estimador de la funció de supervivència de Kaplan-Meier en el cas que no hi hagi empats, és a dir, totes les dades observades Y_i són diferents, com:

$$\hat{S}(t) = \prod_{i:Y(i)\leq t} \left(\frac{n - i}{n - i + 1} \right)^{\delta_{(t)}}$$

En el cas que hi hagi empats entre les observacions censurades i les no censurades considerarem que les no censurades han ocorregut abans que les censurades. Definim d_i com el nombre d'individus que moren en el moment Y_i i a partir de l'expressió anterior podem definir l'estimador de la funció de supervivència de Kaplan-Meier de forma general per tot $t \leq t_{max}$ en el cas que hi hagi empats com:

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

La variància de l'estimador de Kaplan-Meier ve donada per la fórmula de Greenwood:

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

Els estimadors proposats per la variància, en general, tendeixen a estimar-la per mostres petites i moderades.

Entre les propietats més interessants de l'estimador de Kaplan-Meier podem destacar que si no hi ha censura, la funció de supervivència es resumeix a la funció de supervivència empírica i

sota certes condicions de regularitat l'estimador de Kaplan-Meier és l'estimador de màxima versemblança no paramètric generalitzat. A més a més l'estimador de Kaplan-Meier és consistent, és a dir, asimptòticament no té biaix i la variància tendeix a 0.

La forma més habitual de representar l'estimador Kaplan-Meier és gràficament. Com hem comentat en el punt 3.1.1 la funció de supervivència és decreixent. La funció de supervivència estimada és una funció esglaonada on cada salt ve donat per l'aparició d'un esdeveniment, en el nostre cas la mort. Si aquest esdeveniment ocorre després de t_1 es produirà el salt de la funció en t_2. Una altra peculiaritat d'aquesta corba, és que si la darrera dada és una observació censurada la corba mai arribarà al 0 i per tant no sabrem com es distribuirà la cua de la nostra funció de supervivència. La Figura mostra un exemple d'una funció de supervivència estimada.

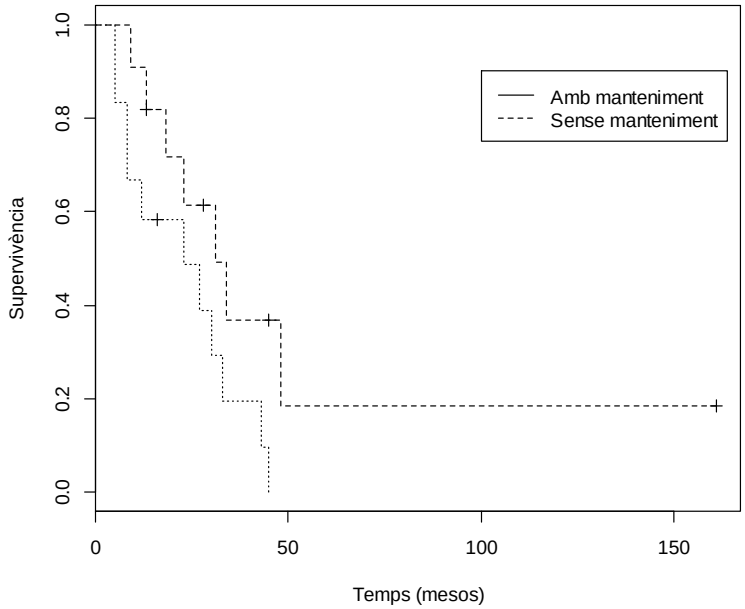

Figura 1.2 Corva Kaplan-Meire (Font: R Documentation: Plot method for survfit objects)

3.1.4 Comparació de funcions de supervivència

En el cas que en l'estudi hi hagi una o més variables d'interès, un dels objectius és comprar les corbes de supervivència per les diferents poblacions. Quan a la mostra estudiada no hi ha censura és poden utilitzar mètodes no paramètrics clàssics per comparar dues o més distribucions, com ara la prova de Wilcoxon per a la comparació de dues poblacions o la prova de Kruskal-Wallis per comparar més de dues poblacions.

Ara bé, en el cas que en la mostra estudiada hi hagin observacions censurades s'acostumen a utilitzar tests ponderats, que es basen en ponderar en cada temps exacte d'ocurrència les

diferències entre el número d'esdeveniments i el nombre esperat d'esdeveniments sota la hipòtesis nul·la d'igualtat de supervivència entre els grups. Sigui R_i el nombre d'individus en risc abans de t_i, definim l'estadístic de contrast com:

$$Z_j(\tau) = \sum_{i=1}^{D} W(t_i) \left\{ d_{ij} - R_{ij} \frac{d_i}{R_i} \right\}$$

Per a mostres grans tots aquests tests segueixen una distribució χ^2 amb 1 grau de llibertat. Diferents eleccions de la funció de ponderació (W) proporcionen proves estadístiques diferents. Segons la ponderació utilitzada s'emfatitzaran unes o altres diferències en la funció de risc. La ponderació que utilitzarem depèn de l'alternativa que es vol detectar, per exemple, si els pesos al principi són grans i la mitjana que transcorre en el temps disminueix, els valors estadístics emfatitzaran les diferències primerenques. Possibles eleccions dels pesos són:

$$W(t_i) = \begin{cases} 1 & \textit{correspon a la prova log-rank} \\ R_i & \textit{correspon a la prova de Gehan} \\ \sqrt{R_i} & \textit{correspon a la prova de Tarone i Ware} \\ \tilde{S}(t_i) & \textit{correspon a la prova de Peto} \\ \hat{S}(t_i - 1)^p \left(1 - \hat{S}(t_i - 1)\right)^q & \textit{correspon a la familia de Fleming i Harrington} \end{cases}$$

On R_i és el nombre d'individus a risc abans d'ocórrer t_i, $\tilde{S}(t_i) = \prod_{t_i \le t} \left(1 - \frac{d_i}{R_i + 1}\right)$, i $\hat{S}$ és l'estimador de Kaplan-Meier per a la mostra combinada. En tots els casos, la hipòtesi nul·la sempre és que no hi ha diferències entre les funcions de supervivència.

Una de les proves estadístiques més utilitzades per comparar dues o més corbes de supervivència és la prova no paramètrica del log-rank, on la hipòtesis nul·la assumeix que els grups provenen de la mateixa població. Aquesta prova té una potència òptima per detectar diferències entre grups quan les funcions de risc són proporcionals.

La prova de Gehan considera pesos que tenen una forta dependència del moment en que ocórrer l'esdeveniment i la distribució de la censura. En canvi, la prova de Tarone-Ware assigna més pes a la diferència entre el nombre d'esdeveniments observats i esperats en aquells punts on hi ha més dades, és a dir, aquesta prova considera que els pesos depenen fortament del moment en que ocórrer l'esdeveniment i de la distribució de la censura. Aquestes dues proves ens poden portar a resultats falsos si els patrons de censura de cada grup són diferents. Altrament la prova de Peto depèn de la funció de supervivència de les mostres combinades i no es veu afectada per a les possibles diferencies en els patrons de censura. Per acabar, la família de proves de Fleming i Harrington variant els valors de p i q permet atansar-se al tipus de distribució desitjat:

$$\begin{cases} p = q = 0 & ens\ porta\ \ a\ la\ prova\ log-rank \\ p = 1, q = 0 & ens\ porta\ a\ la\ prova\ de\ Gehan \\ p > 0, q = 0 & ens\ porta\ a\ una\ prova\ per\ det ectar\ diferéncies\ primerenques \\ p = 0, q > 0 & ens\ porta\ a\ una\ prova\ per\ det ectar\ diferéncies\ tardanes \end{cases}$$

3.2 Model de Cox de riscos proporcionals

El model Cox, o de riscos proporcionals, és un dels models de regressió més utilitzats en l'anàlisi de la supervivència. Els treballs de Cox són pioners en aquesta metodologia i en particular *Regression Models and Life Tables*[22], és un dels més estudiats.

El model de Cox és un model semiparamètric que permet investigar la relació que hi ha entre un conjunt de variables pronòstic i el temps de supervivència.

3.2.1 L'expressió del model

Continuant amb la notació dels apartats anteriors, la variable aleatòria que indica el temps fins al succés ε (en el nostre cas la mort) la denotarem T i als factors pronòstic d'interès, o variables explicatives recollides a l'inici de l'estudi, els denotarem mitjançant $\boldsymbol{Z} = (Z_1, Z_2, ..., Z_p)$ i ens hi referirem com a perfil del individu.

El model bàsic de Cox estableix la regressió multivariant entre el temps de supervivència i les variables d'interès com:

$$\lambda(t|Z) = \lambda_0(t) exp(\beta_1 Z_1 + \beta_2 Z_2 + \cdots + \beta_p Z_p)$$

on λ_0 és la funció de risc basal, que correspondria a un individu que tingues valor 0 en totes les covariables d'interès ($\boldsymbol{Z=0}$).

Una de les premisses d'aquest model és que la raó entre les funcions de risc és manté constant al llarg del temps, ja que es verifica:

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

on el terme de la dreta només depèn de les covariables i no del temps. El factor que defineix la part de la dreta de l'equació correspon al risc relatiu d'un individu amb perfil $\boldsymbol{Z} = (Z_1, Z_2, ..., Z_p)$ respecte a un individu teòric amb factors pronòstic nuls ($\boldsymbol{Z_i=0}$). Així doncs, el risc relatiu ens

dóna la informació de quantes vegades és més gran el risc instantani de presentar l'esdeveniment amb aquell perfil concret.

A partir de la fórmula anterior podem observar que la funció de risc basal λ_0 pot prendre qualsevol forma, per això diem que el model de Cox és semiparamètric.

Aplicant la transformació logarítmica a l'equació anterior obtenim l'anomenat índex pronòstic o *risk score*, una expressió equivalent on el terme de la dreta no conté la constant:

$$ln\left(\frac{\lambda(t|Z)}{\lambda_0(t)}\right) = \beta_1 Z_1 + \beta_2 Z_2 + \cdots + \beta_p Z_p$$

Entre les propietats del model de Cox destaquem les següents:

- El logaritme del quocient de les funcions de risc està relacionat linealment amb els factors pronòstic.
- El logaritme del quocient de les funcions de risc i el quocient de les funcions de risc se suposen constants independentment del moment en el que es calculen. Aquesta propietat és la que els dóna el nom de model de riscos proporcionals.
- La funció de risc basal λ_0 pot prendre qualsevol forma, per aquest fet es diu que el model de Cox és semiparamètric
- El model de Cox es podria interpretar com un model de regressió logística si entenem que la probabilitat de morir $p=\lambda(t|Z')$ en l'interval $(t, t+\Delta t)$ donat que en el moment t l'individu està viu y suposant que p és molt petita:

$$ln\left(\frac{p}{1-p}\right) = \alpha + \beta_1 Z_1 + \beta_2 Z_2 + \cdots + \beta_p Z_p$$

3.2.2 Estimació dels coeficients del model

L'anàlisi estadística del model de Cox es basa en l'estimació dels paràmetres regressors β_j mitjançant la maximització de la versemblança parcial $L(\beta)$ i l'estimació no paramètrica de la funció de risc basal mitjançant la versemblança condicionada als valors estimats dels paràmetres regressors.

Els estimadors que obtenim per als coeficients β_j compleixen, en general, bones propietats característiques del mètode de la màxima versemblança, però la polèmica d'aquesta

metodologia ve donada per la definició de la funció de versemblança parcial, ja que la funció de versemblança no només depèn de β_j, que tenim parametritzats, sinó que també depèn de $\lambda_0(t)$.

Denotem Γ el conjunt que té tota la informació de la mostra, $\Gamma = \{(Y_i,\delta_i,Z_i),\ i=1,2,\ ...,\ n\}$ i definim la funció de versemblança parcial com:

$$L(\beta_1,\ldots,\beta_p)=\prod_{j=1}^{n} P\{e_j = i \mid \Gamma_j\}=\prod_{j=1}^{n} P\{Z_{(j)} = z_{(j)} \mid \Gamma_j\}$$

on z_j és el vector de covariants. Aquesta funció es pot interpretar com el producte, per cada temps de l'esdeveniment, de la probabilitat condicionada de que l'individu presenti l'esdeveniment en el moment t_j sabent que ha passat l'esdeveniment entre els n_j individus a risc en l'instant $t_{(j)}$. Basant-nos en aquesta interpretació i en el fet que la funció de versemblança parcial elimina la dependència de la funció de risc basal $\lambda_0(t)$, podem reescriure la funció de versemblança parcial de manera que el numerador només depèn de la informació del individu que experimenta l'esdeveniment i el denominador només depèn de tots aquells individus que encara no l'han experimentat:

$$L(\beta_1,\ldots,\beta_p)=\prod_{j=1}^{n}\frac{exp\left\{\sum_{k=1}^{p}\beta_k z_{(j)}k\right\}}{\sum_{l\in R(t_{(j)})} exp\left\{\sum_{k=1}^{p}\beta_k z_{jk}\right\}}$$

Finalment, maximitzant la funció de versemblança obtenim l'estimador dels coeficients del model, aconseguint un estimador no esbiaixat, eficient i normal.

Per resoldre la prova d'hipòtesi $H_{0:}\ \beta = \beta_0$ versus $H_{1:}\ \beta \neq \beta_0$ podem utilitzar la prova de Wald, la prova de la raó de versemblança i la prova del *score*. La prova de Wald utilitza directament el valor de l'estimador de la màxima versemblança i es basa en la seva aproximació a una distribució Normal. Aquesta prova te l'avantatge que ens permet presentar les conclusions en funció dels paràmetres d'interès facilitant la interpretació, però té l'inconvenient que sota diferents reparametritzacions no és invariant i, per tant, si la funció de versemblança té una forma inusual, pot donar respostes sense sentit.

Les altres proves proposades són invariants sota diferents parametritzacions, fet que ens permet extreure conclusions independentment de la parametrització utilitzada. Per una banda, la prova de la raó de versemblances utilitza el valor de la màxima versemblança proporcionat per les dades i el compara amb el valor que tindria la funció donat que H_0 fos certa; i per una altra, la prova del *score* utilitza el gradient del logaritme de la versemblança en H_0 i es basa en el fet que sota H_0 el vector *score* segueix aproximadament una distribució Normal.

Es recomana l'ús del test de la raó de versemblances ja que és invariant sota diferents parametritzacions i l'aproximació a la Normal és més ràpida que amb el test del *score.*

Per poder dur a terme la construcció d'un model utilitzant el model de Cox hem de tenir sempre en compte els aspectes no estadístics, les diferents transformacions que poden prendre les variables i la consideració de les tendències no lineals, com podria ser el fet d'afegir termes quadràtics o cúbics i observar la reducció de la desviància.

La comprovació d'una hipòtesi específica ajustada per possibles variables confusores i la predicció de la distribució del temps fins al succés al partir d'un conjunt de covariables són dos dels problemes de l'anàlisi de regressió.

La interpretació dels paràmetres estimats pel model de Cox no es fa directament sobre aquests sinó sobre la seva exponencial, exp($\hat{\beta}$). La interpretació de les covariables dicotòmiques és diferent que la de les variables continues. Per a cada covariable dicotòmica, exp($\hat{\beta}$) és un estimador de la raó de riscos (*hazard ratio*) i s'interpreta com el factor per el qual augmenta el risc instantani amb la presencia de la covariable en relació a l'absència de la covariable. Per al cas de covariables contínues, exp($\hat{\beta}$) representa la raó de riscos (el *hazard ratio*) al incrementar en una unitat la covariable.

3.2.3 Nombre de variables incloses al model de Cox

L'interès del nostre estudi es basa en veure quina relació tenen un seguit de variables amb sentit clínic en el temps de supervivència en pacients amb càncer de pròstata òrgan confinat. Quan ens trobem amb un problema d'aquestes magnituds o volem identificar quines variables poden ajudar als investigadors a explicar el temps de supervivència ens trobem amb el handicap de quin és el nombre màxim de variables que podem incloure al nostre model.

Diversos autors han estudiat aquest problema, ja que es un tema de discussió en la pràctica habitual. En l'article de Peduzzi (1996)[25] es presenta un estudi de simulació de Monte Carlo per avaluar l'efecte del nombre d'esdeveniments per variable (EPV) analitzats en una regressió logística, on s'estudia la incorporació de 7 variables explicatives en el model amb una població on s'han observat 252 esdeveniments. Dins de l'estudi de simulació es generen 500 mostres de 673 individus, els mateixos que els inclosos en l'anàlisi, i es consideren els diferents escenaris incloent 2, 5, 10, 15, 20 i 25 EPV en el model. Els autors arriben a la conclusió que per valors superiors a 10 EPV no s'observen grans problemes però que en el cas de tenir 10 o menys EPV

l'estimació dels coeficients es esbiaixada, es sobreestima la variància global del model i s'infraestima la variància dels coeficients de regressió, entre altres problemes.

Altres estudis afirmen que 10 EPV, o inclús fins a 5 EPV, són suficients per obtenir bons resultats en l'estimació del model de Cox[26,27]. La seva avaluació es basa en una revisió sistemàtica de publicacions amb models de Cox on avaluen l'interval de confiança per als coeficients del model, l'error de tipus I de la prova d'hipòtesi H_0: $\beta_1=0$ i el biaix de l'estimació dels coeficients del model obtenint bons resultats. No obstant, tots els autors coincideixen en la premissa d'interpretar els resultats amb cura quan es té un nombre baix d'EPV.

Un cop escollides les variables que es volen incloure en el model, és important comprovar que no estan gaire relacionades entre elles. Per observar la correlació entre les variables contínues s'utilitza el coeficient de correlació de Pearson. El valor de l'índex de correlació va de -1 a 1, on -1 indica una correlació negativa perfecta, 0 la no existència de relació lineal i 1 la presència d'una correlació positiva perfecta. Per estudiar la correlació entre dues variables categòriques s'empra la V de Cramer que pren valors entre 0, que indica que no hi ha associació entre les dues variables estudiades, i 1, que indica una major associació entre variables. Habitualment valors superiors a 0,3 ens indiquen una possible relació entre les variables estudiades. Per acabar, per quantificar la correlació entre una variable categòrica i una continua s'utilitza l'Eta quadrat que s'interpreta com la proporció de variància de la variable continua que està explicada per la variable categòrica.

3.3 Anàlisi de residus

Una vegada construït el model de Cox per al temps de supervivència, hem de verificar que es compleixen les premisses en les quals es basa el model, és a dir, comprovar la linealitat dels regressors respecte a la variable resposta i la bondat d'ajust. La validesa d'aquest model i l'anàlisi dels residus tenen la mateixa importància en el model de Cox com en la regressió lineal.

En la regressió de Cox la definició dels residus no és unívoca, en aquest model tenim diferents definicions que ens portaran a diferents exàmens dels residus relacionats amb la comprovació del model.

Al llarg d'aquesta secció es tindrà en compte la terminologia detallada a continuació. Considerem, com en la resta del manuscrit, el conjunt de dades observades $D=\{Y_i, \delta_i, Z_i, i=1,2,..,n\}$ i suposem que s'ha ajustat el següent model de Cox $\lambda(t|Z_i)=\exp\{\beta' Z_i\}\lambda_0(t)$ obtenint

els estimadors dels paràmetres del model, la funció de risc basal i la funció de risc acumulada basal. D'aquesta forma, la funció de risc estimada per l'i-èssim individu ve donada per la següent expressió:

$$\hat{\lambda}(t|Z_i) = \exp\{\hat{\beta}' Z_i\}\hat{\lambda}_0(t)$$

A continuació es descriuran els diferents tipus de residus i el diagnòstic del model derivat a cada un d'ells. Concretament, s'estudiaran els residus de Cox i Snell, els basats en martingales, en l'*score*, en la desviància i els residus de Schoenfeld.

Residus de Cox i Snell

Definim els residus de Cox i Snell per a l'individu i-èssim com:

$$rc_i = \exp\{\hat{\beta}' Z_i\}\hat{\Lambda}_0(y_i),\ i = 1,\cdots,n$$

on $\hat{\Lambda}_0(t_i)$ és l'estimador de Nelson-Aalen. Una de les propietats d'aquests residus és que no estan distribuïts simètricament al voltant de 0 i no poden ser negatius.

L'ús dels residus de Cox i Snell ens serveix per comprovar l'ajust global del model mitjançant un gràfic de la funció de risc acumulada per la mostra rc_i contra cadascun dels residus rc_i. Si l'ajust del model és correcte obtindrem un gràfic molt similar a una recta que passi per l'origen amb una pendent de 45º.

El major inconvenient d'aquests residus és que es distribueixen aproximadament com una exponencial independentment de la validesa del model i aquest fet fa que alguns autors no recomanin el seu ús[28]. Un altre dels desavantatges es que no ens donen informació del tipus de llunyania del model quan el gràfic no és lineal. Per últim remarcar que la llunyania del model també pot venir donada per la imprecisió en l'estimació dels paràmetres, ja que per construir aquest tipus de residus substituïm els valors dels paràmetres pels seus estimadors.

Residus basats en martingales

Els residus basats en martingales es basen en una petita modificació dels residus de Cox i Snell i es defineixen per cada individu com:

$$r_{M_i} = \delta_i - r_{C_i} = \begin{cases} 1 - r_{C_i} & \textit{si la observació no està censurada} \\ -r_{C_i} & \textit{si la observació està censurada} \end{cases}$$

Els residus basats en martingales no es distribueixen asimptòticament al voltant de 0 i prenen valors entre -∞ i 1 per les observacions no censurades i -∞ i 0 per les observacions censurades. La seva interpretació es pot fer com la diferència entre el nombre d'esdeveniments observats i els esperats sota el model de Cox.

Entre les propietats d'aquests residus destaquem que el valor esperat de cada residu quan s'avalua el valor verdader de β es 0 $\left(E\left(r_{M_i}\right)=0\right)$, la suma de tots els residus calculats en $\hat{\beta}$ és 0. A més a més els residus calculats amb el verdader valor de β no estan correlacionats, però els residus calculats amb $\hat{\beta}$ estan correlacionats negativament.

Una de les particularitats dels residus basats en martingales és que detecten molt bé aquells individus que viuen més del que s'espera, però no tant bé els que viuen menys de l'esperat, i es pot comprovar examinant els valors observats menys els valors esperats. Aquests residus també són molt útils per determinar la millor transformació per cada covariable de manera que expliqui de forma òptima el seu efecte sobre el temps de supervivència. Per tal de poder trobar la millor transformació que s'adequa a cada covariable se suposa que es coneix la millor transformació de la resta de covariables incloses al model, s'ajusta el model de Cox basant-se en les p-1 covariables i es calcula els residus basats en martingales per aquest model. Després es grafiquen els valors dels residus obtinguts contra els valors de la covariable per a la i-èssima observació: si el núvol de punts és lineal no és necessari fer cap tipus de transformació.

Residus basats en l'*score*

Definirem els residus basats en l'*score* per la covariable k-èssima i l'individu i-èssim com:

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

$$\overline{Z}_k(t)=\frac{\sum_{l=1}^n J_l(t)Z_{lk}\exp\{\beta' Z_l(t)\}}{\sum_{l=1}^n J_l(t)\exp\{\beta' Z_l(t)\}}$$

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

on $J_i(t)$ pren valor 1 si l'individu *i* està en risc abans de t i $N_i(t)$ és l'indicador de si l'individu i-èssim ha presentat el succés. Aquesta definició ens permet expressar les covariables dependents del temps.

El conjunt de residus basats en l'*score* forma una matriu n x p on cadascun dels residus (Z_{ij}) estableix la diferencia entre el valor observat i el valor esperat de la covariable *j* corresponent a l'individu i-èssim. Podríem dir que el resultat que obtenim és un càlcul aproximat per comparar l'estimador de β que s'obté utilitzant totes les dades amb l'estimador de β que obtindríem utilitzant totes les dades menys les de l'individu j-èssim.

Els residus basats en l'*score* són útils per comprovar la hipòtesi de riscos proporcionals i el gràfic d'aquests residus versus les covariables ens ajuda a determinar la influència de l'individu i-èssim en l'estimació dels coeficients.

Residus basats en la desviància

Els residus basats en la desviància són una transformació dels residus basats en martingales i es defineixen com:

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

S'observa que quan r_{Mi} és 0, r_{Di} també és 0 i que l'efecte del logaritme en la definició d'aquests residus és inflar els valors propers a 1 disminuir els valors negatius grams.

Entre les propietats dels residus basats en la desviància destaquem que estan simètricament distribuïts al voltant de 0 i s'aproximen a una distribució Normal quan el model és correcte.

Els residus basats en la desviància serveixen per avaluar l'efecte de cada individu en el model estimat, per poder-ho avaluar es planteja un gràfic d'aquests residus contra el *risk score*. Si el model es correcte, quan el nombre d'individus censurats no es excessiu aquests residus s'haurien de comportar com una distribució Normal, en canvi, es té un gran nombre d'individus censurats, com es el nostre cas, s'observa un gran nombre de valors al voltant de 0 que distorsionen la interpretació del gràfic. Aquests residus també son bons detectors tant d'individus que viuen més de l'esperat com els que viuen menys de l'esperat pel model.

Un cop es te el model definitiu els residus basats en la desviància ajuden a decidir sobre la bondat d'ajust del model, concretament a observar els valors atípics.

Residus de Schoenfeld

Definim el residu de Schoenfeld per al k-èssim succés i el i-èssim individu quan no hi ha empats com:

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

Aquests residus són útils per detectar anomalies per a cada una de les variables que intervenen al model. En cas que sigui necessari, també ens suggereixen possibles transformació a aplicar a les dades. D'altra banda, ens permeten determinar la influència que te cada individu en l'estimació dels coeficients i comprovar la hipòtesis de riscos proporcionals.

3.4 Gran censura per la dreta

En estudis de supervivència el seguiment dels individus molts cops està censurat per la dreta degut a la finalització de l'estudi o de la pèrdua de seguiment. La finalització de l'estudi és, en general, fixada per l'investigador i és un cas particular de censura per la dreta.

La gran censura (*Heavy censoring*) per la dreta apareix amb freqüència en l'anàlisi de supervivència degut a un seguiment insuficient. Quan el període de seguiment no és molt llarg, no s'observa l'esdeveniment estudiat per un gran nombre de casos, aquests individus sense esdeveniment són censurats per la dreta. Un segon cas que produeix gran censura per la dreta són els supervivents a llarg termini, en aquests individus no s'observa l'esdeveniment durant l'estudi però en un futur el presentaran, aquests casos també seran censurats per la dreta al final de l'estudi.

Quan s'estima un model de Cox amb una gran censura per la dreta ens plantegem un seguit de qüestions sobre la qualitat de l'ajust del model i dels seus estimadors: (1) les propietats dels estimadors continuen sent valides? (2) quan tenim una gran censura per la dreta és correcte analitzar el temps de supervivència mitjançant un model de riscos proporcionals? (3) en quins escenaris no és recomanable analitzar les dades mitjançant un model de Cox?

La tesi doctoral de López (2014)[29] estudia mitjançant simulació, l'efecte que tenen diferents nivells de censura per la dreta en l'estimació del risc relatiu en models de riscos proporcionals. Pressuposa un model fix de censura per la distribució de la censura i un model de Cox per al temps de supervivència. Les simulacions assumeixen com a variable explicativa una variable binària i un percentatge fix de censura (nivell de censura). Per tal d'avaluar l'efecte d'aquest nivell de censura en el risc relatiu estudia el biaix, la variància i l'error quadràtic mig dels estimadors obtinguts.

López conclou que en nivells de censura entre el 70% i el 90%, el model de Cox és adequat si la grandària mostral és superior o igual a 500. L'estudi verifica que el comportament del risc relatiu, en termes de l'error quadràtic mig, és millor si les mostres estan balancejades. En grandàries mostrals inferiors a 500 individus, si la censura és molt gran no hem d'utilitzar el

model de Cox o l'hem d'utilitzar amb cura. En aquest cas ens podem trobar amb problemes de convergència o podem obtenir una estimació errònia del risc relatiu, ja que tendeix a la sobreestimació dels paràmetres.

3.5 Llibreries d'R

Una eina de consulta molt útil per trobar informació sobre paquets d'R per a l'anàlisi de la supervivència és el *Task View* per a l'anàlisi de supervivència que ens posa a disposició el CRAN (http://cran.r-project.org/web/views/Survival.html), on es pot trobar un recull exhaustiu de tots els paquets desenvolupats per dur a terme aquest anàlisi. A continuació detallem el conjunt de paquets especialitzats que hem utilitzat en aquest projecte.

Paquet `survival`

Un dels paquets més utilitzats en l'anàlisi de la supervivència mitjançant R és el `survival`, actualment està disponible la seva versió 2.38-1, es una llibreria desenvolupada per Thomas Lumley a partir del codi inicialment desenvolupat per Terry Therneau[30,31]. A continuació es detallen les funcions utilitzades per dur a terme l'anàlisi.

- **`Surv`**. Per poder utilitzar les funcions d'aquest paquet s'ha de definir el temps de supervivència i la censura, que es pot fer mitjançant la funció *`Surv(time,event)`* que ens permet crear objectes del tipus *`survival`*, on *`time`* es el temps fins al succés i event la variable censura, que pren valor 0 per les dades censurades i 1 que s'observa el succés.

- **`survfit`**. Ens permet obtenir l'estimació de la funció de supervivència utilitzant el mètode de Kaplan i Meier, que és l'opció per defecte, o de Nelson i Aalen. També permet predir la funció de supervivència per a models de Cox.

- **`Survdiff`**. Ens permet efectuar contrastos d'hipòtesi per verificar la igualtat o diferència de dues o més corbes de supervivència, basats en les famílies de proves G-rho proposades per Harrington i Fleming (1982).

- **`coxph`**. Ens permet ajustar models de regressió de Cox, models amb variables dependents del temps, models estratificats, models de múltiples esdeveniments per individu i altres extensions derivades des del punt de vista en els processos de

comptatge. En aquest estudi s'utilitza aquesta funció per fer inferència sota un model de riscos proporcionals obtenint la prova de Wald, de la raó de versemblança i de l'*score* per poder prendre decisions. Un altre aspecte important d'aquesta funció és que, en el cas que hi hagi empats, ens permet distingir entre la versemblança d'Efron, opció per defecte, la de Breslow i l'exacta.

- **`cox.zph`**. Ens permet dur a terme el contrast d'hipòtesi de riscos proporcionals. Sota la hipòtesis nul·la s'assumeix que els riscos són proporcionals, associat a que els paràmetres són 0. La sortida directa d'aqueta funció presenta el contrast global i per cadascuna de les covariables incloses al model.

- **`residuals`**. Aquesta funció es pot aplicar als models generats mitjançant les funcions coxph i survreg i ens permet calcular els residus per dur a terme el diagnòstic del model.

Paquet `FHtest`

El paquet `FHtest` ens permet dur a terme un gran nombre de comparacions de dues o més corbes de supervivència a partir del test de Fleming i Harrington per dades censurades per la dreta i a partir d'una extensió del mateix test de Fleming i Harrington per dades censurades en intervals. Actualment està disponible la seva versió 1.2, és una llibreria desenvolupada per Ramon Oller i Klaus Langohr[32]. A continuació es detalla la funció utilitzades per dur a terme l'anàlisi.

- **FHtestrcc**. Ens permet dur a terme el contrast de Fleming i Harrington per dades censurades per la dreta mitjançant un procés de comptatge. Especificant diferents valors per rho i lambda ens permet obtenir el resultat de diferents tests, per exemple amb valor de rho igual a 0 i lambda igual a 0 obtenim els resultats de la prova log-rank, amb valor de rho igual a 1 i lambda igual a 0 els de la prova de Peto-Peto. Si volem donar més pes a les observacions tardanes utilitzarem un de rho igual a 0 i de lambda igual a 1 i si volem donar més pes als valors propers a la mediana utilitzarem rho igual a 1 i lambda igual a 1.

Paquet `rms`

El paquet `rms` proposa una versió modificada de la funció survfit. És un paquet que pot treballar qualsevol tipus de model de regressió, però esta dissenyat bàsicament per a l'anàlisi de models de regressió binaris o ordinals i models de Cox. Actualment està disponible la seva

versió 4.3-1, és una llibreria desenvolupada per Frank E. Harrell Jr[33]. A continuació es detalla la funció utilitzada per dur a terme l'anàlisi.

- **`Survplot`**. Aquesta funció ens permet graficar corbes estimades de supervivència i conté una opció que ens permet incloure en el gràfic el nombre d'individus en situació de risc en cada interval de temps.

4. Resultats

4.1 Anàlisi descriptiva

Inicialment es van reclutar 763 pacients amb càncer de pròstata, dels quals se'n van excloure 59 (7,7%) per presentar un risc alt en la classificació de la malaltia. Dels 704 pacients inclosos en l'anàlisi, a un 27,4% se'ls va realitzar una prostatectomia radical, un 27,6% van rebre tractament radioteràpic i un 45,0% van rebre tractament braquiteràpic, quedant així distribuïts els grups d'interès.

A la Taula 4.1 es mostren les característiques basals de la mostra en estudi, on s'observa que la mediana d'edat dels pacients amb PR (64,7 anys) és inferior que per la resta de tractaments, seguits pels pacients amb BT (67,5 anys) i els pacients amb RT (70,9 anys). Si s'observa la categorització de l'edat, més de la meitat dels pacients amb PR tenen menys de 65 anys (52,1%), més de la meitat dels pacients amb RT (58,0%) i un 40,3% dels pacients amb BT tenen més de 70 anys. S'observa que tant els valors de PSA com la puntuació de Gleason entre els tres grups són força semblants.

També s'observa un major percentatge de pacients amb BT que presenten un valor baix en l'índex T (81,4%), seguit dels pacients amb PR (67,9%) i dels pacients amb RT (58,2%).

Pel que fa al tractament hormonal, destacar que només un 8,8% dels pacients amb PR van rebre tractament hormonal previ al tractament en estudi, a diferència dels pacients amb RT i BT que en van rebre un 31,4% i un 33,1% respectivament. Subratllar també que a un 16,0% dels pacients amb RT se'ls va administrar tractament hormonal complementari al tractament en estudi, reduint-se aquest valor al 1,0% i 1,9% en els pacients tractats amb PR i BT respectivament. Per acabar ressaltar que un 15,5% dels pacients tractats amb RT va rebre tractament hormonal de rescat, enfront al 9,8% dels pacients tractats amb BT i al 6,2% dels pacients tractats amb PR.

Als 5 anys de l'inici de l'estudi el grup que va presentar més èxitus va ser el de la RT (12,4%), seguit de la BT (9,8%) i de la RT (6,2%).

Taula 4.1: Característiques basals de la mostra en estudi

	Prostatectomia radical	**Radioteràpia**	**Braquiteràpia**	**Global**
Participants, n (%)	193 (27,4%)	194 (27,6%)	317 (45,0%)	704 (100,0%)
Característiques clíniques				
Edat				
mitjana (DE)	64.2 (5.5)	70.1 (5.3)	67.5 (6.4)	67.3 (6.3)
mediana	64.7	70.9	67.5	68.2
(Q1, Q3)	(60.0 , 68.4)	(67.7 , 73.7)	(63.5 , 72.1)	(63.5 , 71.9)
(Mín, Màx)	(44.8 , 74.8)	(55.26 , 83.1)	(46.48 , 83.2)	(44.8 , 83.2)
< 65 anys	100 (52,1%)	32 (16,6%)	94 (30,0%)	226 (32,4%)
65 – <70 anys	64 (33,3%)	49 (25,4%)	93 (29,7%)	206 (29,5%)
≥ 70 anys	28 (14,6%)	112 (58,0%)	126 (40,3%)	266 (38,1%)
PSA (ng/mL),				
mitjana (DE)	7.62 (2.92)	8.06 (3.38)	6.96 (2.23)	7.45 (2.81)
mediana	7.21	7.60	6.63	6.91
(Q1, Q3)	(5.33 ,)	(5.70 , 9.35)	(5.50 , 8.25)	(5.50 , 8.70)
(Mín, Màx)	(2.30 ,)	(0.08 , 19.00)	(0.82 , 18.3)	(0.08 , 19.00)
Gleason score, m (DE)				
mitjana (DE)	6.3 (0.7)	5.9 (1.1)	5.5 (0.9)	5.8 (1.0)
mediana	6.0	6.0	6.0	6.0
(Q1, Q3)	(6.0 , 7.0)	(6.0 , 7.0)	(5.0 , 6.0)	(6.0 , 6.0)
(Mín, Màx)	(4.0 , 7.0)	(2.0 , 7.0)	(2.0 , 7.0)	(2.0 , 7.0)
Clinical T Stage, n (%)				
T1	131 (67,9%)	113 (58,2%)	258 (81,4%)	502 (71,3%)
T2	62 (32,1%)	81 (41,8%)	59 (18,6%)	202 (28,7%)
Tx	0 (0,0%)	0 (0,0%)	0 (0,0%)	0 (0,0%)
Grup de risc, n (%)				
Baix	92 (47,7%)	107 (55,1%)	283 (89,3%)	482 (68,5%)
Intermig	101 (52,3%)	87 (44,8%)	34 (10,7%)	222 (31,5%)
Tractament hormonal anterior a l'inici del tractament, n (%)	17 (8,8%)	61 (31,4%)	105 (33,1%)	183 (26,0%)
Complement de tractament hormonal (adjuvant), n (%)	2 (1,0%)	31 (16,0%)	6 (1,9%)	39 (5,5%)
Tractament hormonal de rescat, n (%)	13 (6,7%)	30 (15,5%)	16 (5,0%)	59 (8,4%)
Èxitus, n (%)	12 (6,2%)	24 (12,4%)	31 (9,8%)	67 (9,5%)

4.2 Anàlisi de la supervivència

Tal i com s'observa en la funció de supervivència global (Figura 4.1), l'eix vertical és la probabilitat de supervivència estimada del pacient i l'eix horitzontal és el temps que comprèn des de l'inici de l'estudi fins al tall per analitzar els 5 primers anys de seguiment dels pacients. En aquest gràfic es pot veure que la probabilitat estimada de sobreviure als 5 anys de seguiment és de 0,93.

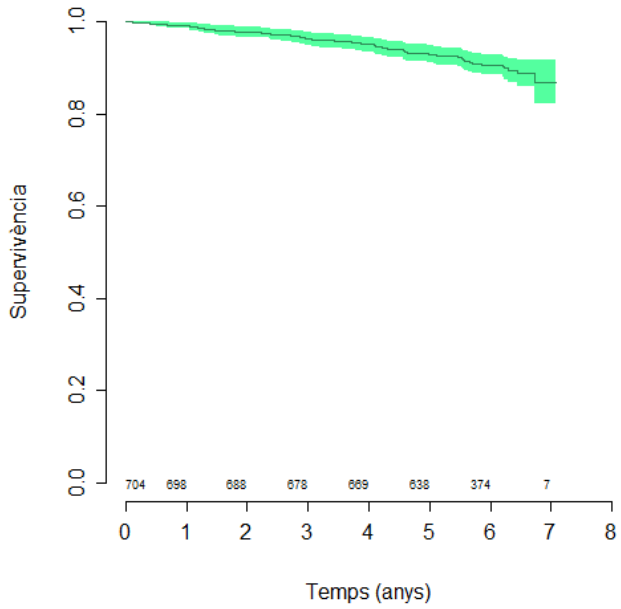

Figura 4.1 Estimació de Kaplan-Meier i intervals de confiança del 95% de la funció de supervivència global

Com hem comentat en l'apartat de mètodes un dels objectius clínics d'aquest treball és estudiar diferents corbes de supervivència per diferents subgrups d'interès. A continuació es presenten les corbes més rellevants, la resta de corbes estudiades es mostren en l'apèndix B (Corbes de supervivència addicionals).

Com es pot observar en la Figura 4.2, la probabilitat estimada d'èxitus dels pacients tractats amb PR als 5 anys és superior que els tractats amb RT i BT (0,97, 0,91 i 0,90 respectivament). També es pot observar que la BT té el major nombre d'individus a risc als cinc anys, seguit de la RT i la PT (173, 105 i 96 respectivament).

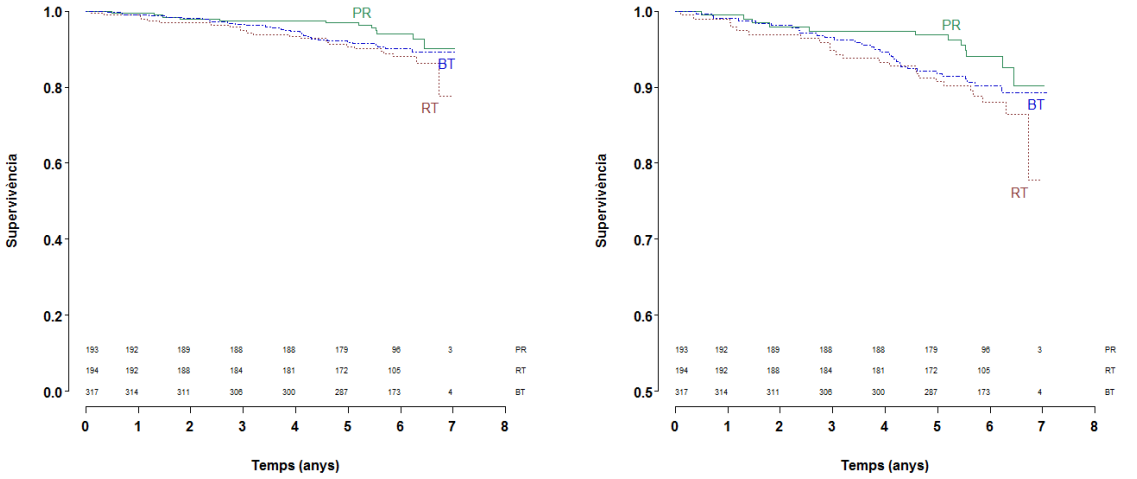

Figura 4.2 Corba de supervivència segons grup de tractament[1]

Si observem el comportament per separat dels pacients amb risc baix i risc moderat (Figura 4.3) es pot veure que els pacients amb risc baix es comporten de la mateixa manera, presentant una probabilitat estimada de supervivència als 5 anys de 0,96 en el cas de la PR, de 0,93 en el cas de la RT i 0,92 en el cas de la BT. En canvi, en el grup de pacients de risc moderat s'observa una probabilitat estimada de supervivència als 5 anys de 0,98, 0,90 i 0,82 respectivament.

Notem que els pacients amb tractament hormonal previ tractats amb PR tenen una supervivència als 5 anys de 1, seguit del pacients BT (0,94) i dels pacients RT (0,91). En el grup de pacients sense tractament hormonal previ s'observa una supervivència als 5 anys de 0,97 en els pacients tractats amb PR, 0,91 en els pacients BT i 0,90 en els pacients RT. (Figura 4.4).

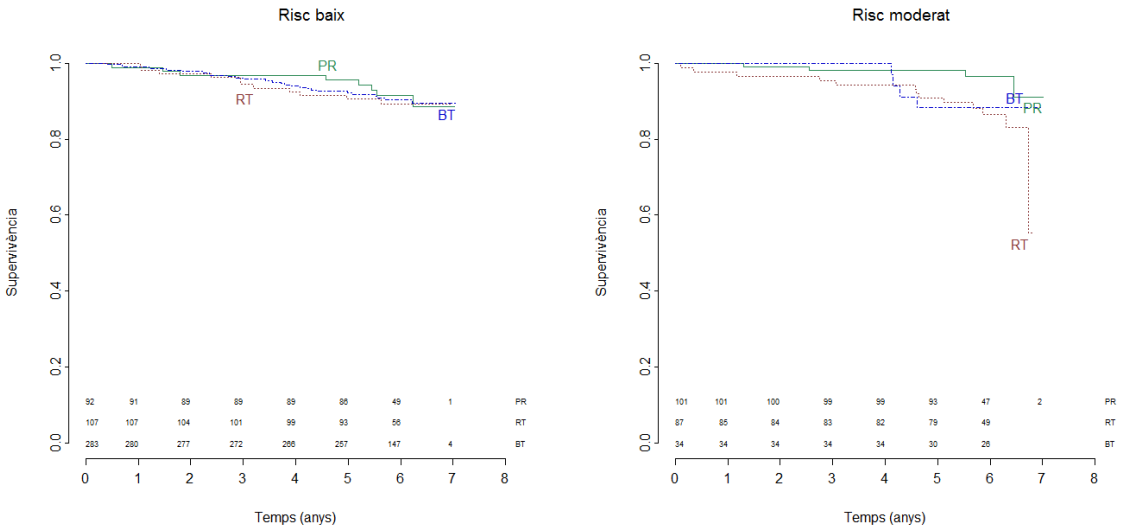

Figura 4.3 Corba de supervivència segons grup de risc i grup de tractament

[1] El gràfic de la dreta mostra les mateixes dades que el gràfic de l'esquerra canviant l'eix de les ys

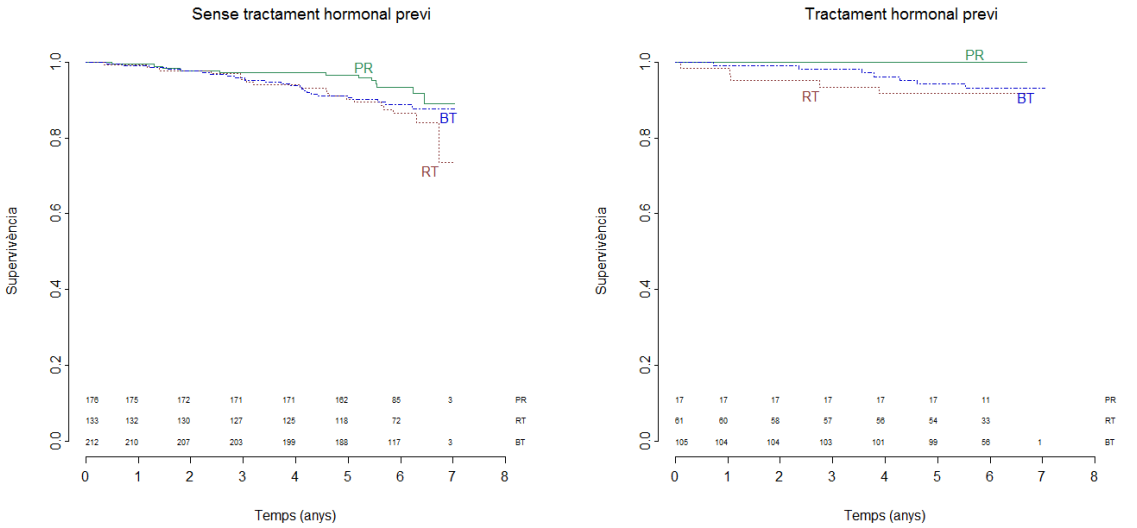

Figura 4.4 Corba de supervivència segons tractament hormonal previ i grup de tractament

La Taula 4.2 a continuació presenta un resum de totes les proves estadístiques realitzades (log-rank, Peto-Peto i Fleming i Harrington) per comparar les funcions de supervivència dels tres tractament en les diferents submostres d'interès. Segons es pot veure, comparant els temps de supervivència entre els grups de tractament mitjançant la prova del log-rank obtenim diferències estadísticament significatives ($p<0.05$) en els pacients entre 65 i 70 anys, en els pacients amb risc moderat i en els pacients amb tractament hormonal previ. Aquestes diferències es mantenen amb la prova de Peto-Peto. En canvi, la prova de Fleming i Harrington amb $\rho=1$ i $\lambda=1$ no s'observen diferències estadísticament significatives per cap subgrup d'interès.

Taula 4.2: Comparació de les funcions de supervivència dels tractaments en diferents subgrups d'interès

	Log-rank			**Peto-Peto**			**Fleming i Harrington (ρ=1 λ=1)**		
	χ^2	GLL	p-valor	χ^2	GLL	p-valor	χ^2	GLL	p-valor
Global	4.1	2	0.132	4.1	2	0.128	2.3	2	0.310
Edat: <65 anys	0.6	2	0.727	0.7	2	0.717	0.7	2	0.710
Edat: 65 – <70 anys	8.1	2	0.0174	8.2	2	0.0165	4.4	2	0.113
Edat: ≥ 70 anys	1.4	2	0.499	1.4	2	0.486	0.8	2	0.672
Clinical T stage: T1	1.6	2	0.451	1.6	2	0.439	0.7	2	0.719
Clinical T stage: T2	2.8	2	0.245	2.8	2	0.243	1.7	2	0.434
Grup de risc: Baix	0.2	2	0.919	0.2	2	0.907	0	2	0.99
Grup de risc: Moderat	6.3	2	0.0427	6.3	2	0.0435	5	2	0.0829

	Log-rank			**Peto-Peto**			**Fleming i Harrington (ρ=1 λ=1)**		
	χ^2	GLL	p-valor	χ^2	GLL	p-valor	χ^2	GLL	p-valor
Tractament hormonal previ: No	4.2	2	0.125	6.5	2	0.039	5.6	2	0.0626
Tractament hormonal previ: Si	6.4	2	0.04	6.5	2	0.039	5.5	2	0.0626
Tractament hormonal adjuvant: No	4.9	2	0.868	4.9	2	0.842	2.9	2	0.234
Tractament hormonal adjuvant: Si	4	2	0.136	4	2	0.136	2	2	0.376
Tractament hormonal de rescat: No	2.4	2	0.3	2.5	2	0.285	0.6	2	0.725
Tractament hormonal de rescat: Si	5.6	2	0.0618	5.7	2	0.0585	2.5	2	0.286

4.3 Model de Cox de riscos proporcionals

Com hem comentat anteriorment l'interès del nostre estudi es basa en veure quina relació tenen el tractament per al càncer, l'edat, el nivell de PSA, el tractament hormonal complementari al tractament quimioteràpic i les dues variables que quantifiquen el risc (*clinical T stage* i *Gleason score*) en la supervivència dels pacients amb càncer de pròstata. En el nostre cas volem incloure 5 variables d'interès clínic al model i s'han observat 67 èxitus, el que ens porta a obtenir 13 EPV, valor acceptat per tota la literatura consultada.

Abans d'introduir les variables al model es comprova que no estiguessin correlacionades entre elles. A la Taula 4.2 s'observa que no hi ha cap parell de variables que tinguin un coeficient de correlació que indiqui una forta relació entre elles.

Taula 4.3: Coeficients de correlació entre les variables d'interès

	Edat	PSA	Tractament hormonal adjuvant	Gleason score	Clinical T stage
Tractament	0.1191	0.0275	0.2820	0.3120	0.2170
Edat	-	0.0970	0.0126	0.0001	0.0004
PSA		-	0.0403	0.0067	0.0034
Tractament hormonal adjuvant			-	0.1920	0.0110
Gleason score				-	0.2270

Un cop construït el model, podem observar que el tractament amb RT o BT tenen un signe positiu en l'estimació dels seus paràmetres, fet que es tradueix en un major risc instantani a morir estimat en el model de Cox. Els pacients tractats amb RT tenen 1.5 vegades més de risc de morir que els pacients amb PR i els pacients tractats amb BT 1.1 vegades més. També podem veure que el PSA, el fet de rebre tractament hormonal adjuvant, el Gleason score, i el clinical stage T2 són factors protectors en el risc de mort. Però, tot això només és vàlid a nivell de la mostra. Utilitzant un nivell de significació del 0,05, no podem fer cap afirmació a nivell poblacional, ja que tots els p valors d'aquestes variables són superiors a 0,15. Cal destacar que la variable edat es la única estadísticament significativa ($p<0.05$) amb un hazard rati estimat de 1,061 indicant que el risc de morir augmenta amb l'edat.

Taula 4.4: Model de Cox amb les covariables d'interès

Variable	$\hat{\beta}$	$\exp(\hat{\beta})$	$se(\hat{\beta})$	$IC(e^{\beta};95\%)$	p-valor
Tractament - RT	0.405	1.499	0.3902	(0.698 , 3.220)	0.2998
Tractament - BT	0.069	1.072	0.369	(0.520 , 2.208)	0.8508
Edat	0.059	1.061	0.024	(1.013 , 1.111)	0.0125
PSA	-0.027	0.974	0.048	(0.887 , 1.069)	0.5735
Tractament hormonal adjuvant	-0.893	0.409	0.736	(0.097 , 1.733)	0.2250
Gleason score	-0.170	0.844	0.128	(0.6567 , 1.083)	0.1821
Clinical T satge	-0.190	0.827	0.291	(0.467 , 1.465)	0.5155

4.4 Anàlisi dels residus

Com hem comentat en l'apartat 3.3 (Anàlisi de residus), analitzarem la verificació dels supòsits del model de Cox mitjançant la metodologia proposada per Therneau i Grambsch[28].

A la Figura 4.5 podem observar els residus basats en la desviància, on s'observa que el nostre model s'ajusta força bé a les dades però que hi ha un determinat nombre d'individus que viuen més de l'esperat, com es pot veure en el núvol de punts superior.

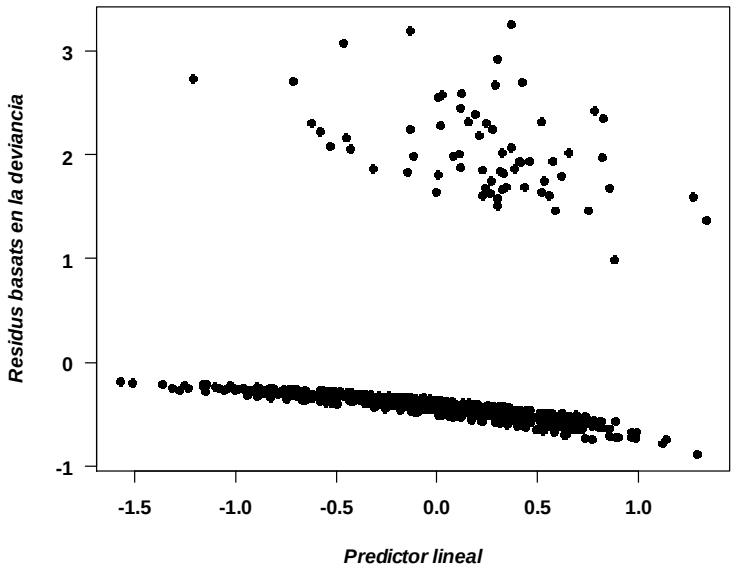

Figura 4.5 Residu basats en la desviància

Mitjançant els residus basats en l'*score* podem veure que hi ha dues observacions influents en el tractament hormonal adjuvant (Figura 4.6). En concret són les observacions 108 i 536, ambdós pacients reben tractament adjuvant i tenen un temps de supervivència de 4,6 i 2,8 anys respectivament. Si observem el temps de vida dels pacients que reben tractament hormonal adjuvant podem veure que aquestes dues observacions es corresponen amb els valors mínims d'aquest subgrup.

Hem realitzat el model de Cox excloent les dues observacions influents i obtenim coeficients molt semblants en l'estimació del model per a totes les variables menys per al tractament hormonal adjuvant en el qual tenim un coeficient de -0.893 utilitzant totes les variables i un coeficient de -17.245 excloent les dues observacions influents detectades. Aquest valor tant exageradament gran indica un problema de separació a causa dels pocs casos de mort al grup de tractament hormonal adjuvant.

Taula 4.5: Model de Cox sense observacions influents

Variable	$\hat{\beta}$	$\exp(\hat{\beta})$	$se(\hat{\beta})$	$IC(e^{\beta};95\%)$	p-valor
Tractament - RT	0.337	1.400	0.394	(0.647 , 3.030)	0.393
Tractament - BT	0.009	1.009	0.370	(0.488 , 2.086)	0.980
Edat	0.065	1.067	0.024	(1.018 , 1.119)	0.007
PSA	-0.050	0.952	0.051	(0.862 , 1.051)	0.328
Tractament hormonal adjuvant	-17.245	0.00	2598.968	(0.000 , -)	0.995
Gleason score	-0.202	0.817	0.129	(0.634 , 1.053)	0.119
Clinical T satge	-0.217	0.805	0.299	(0.448 , 1.445)	0.467

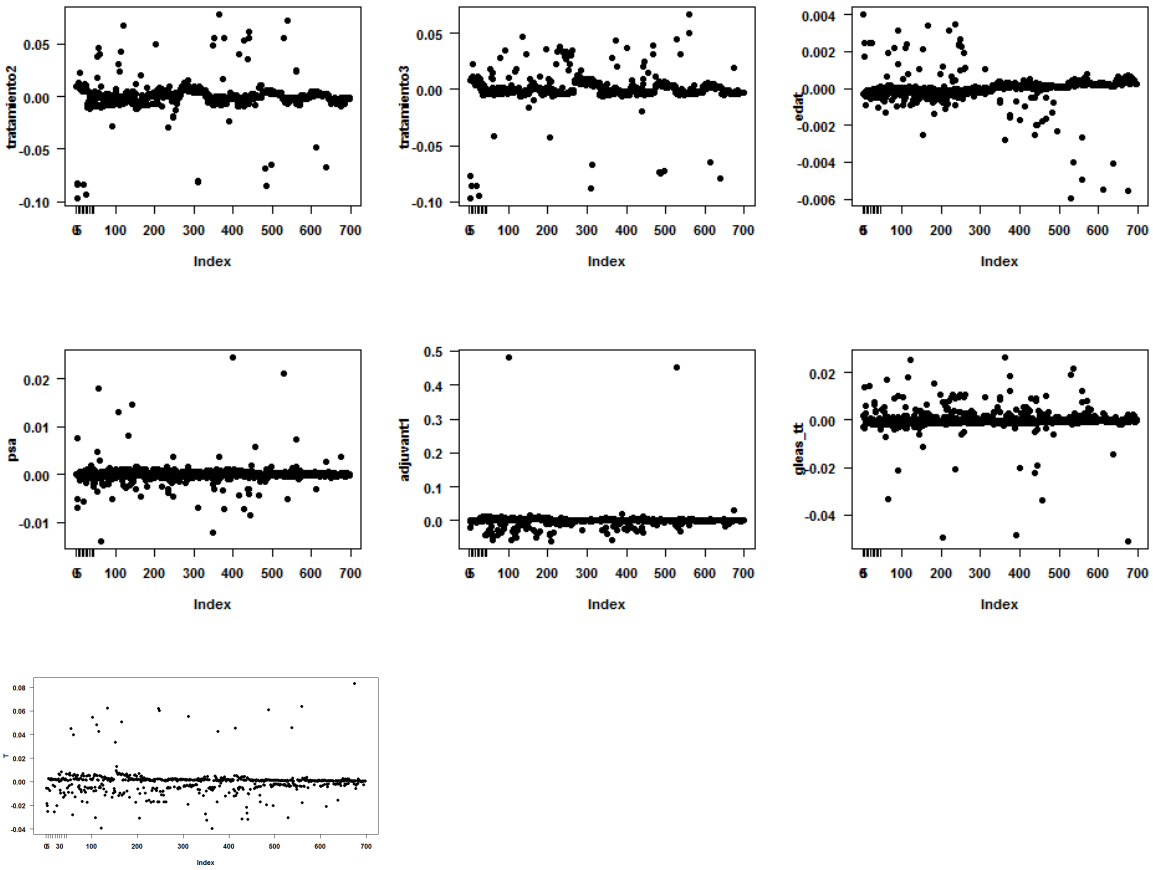

Figura 4.6 Residus basats en l'*score*

A continuació es presenta la comprovació de riscos proporcionals com podem observar en la Taula 4.4, les figures 4.7 i 4.8, on podem concloure que globalment no rebutgem la hipòtesis nul·la de la proporcionalitat de riscos per al nostre model (p=0.136; Taula 4.4). A més a més en la Figura 4.8 es torna a confirmar la presencia dels dos outliers comentats anteriorment.

Taula 4.6: Proporcionalitat de riscos

Variable	ρ	χ^2	p-valor
Tractament - RT	-0.151	1.470	0.2256
Tractament - BT	-0.140	1.270	0.2591
Edat	0.079	0.295	0.5873
PSA	0.115	1.360	0.2428
Tractament hormonal adjuvant	-0.003	0.0007	0.9794
Gleason score	-0.158	2.160	0.1420
Clinical T stage	0.337	7.150	0.0075
Ajust Global	NA	11.100	0.1363

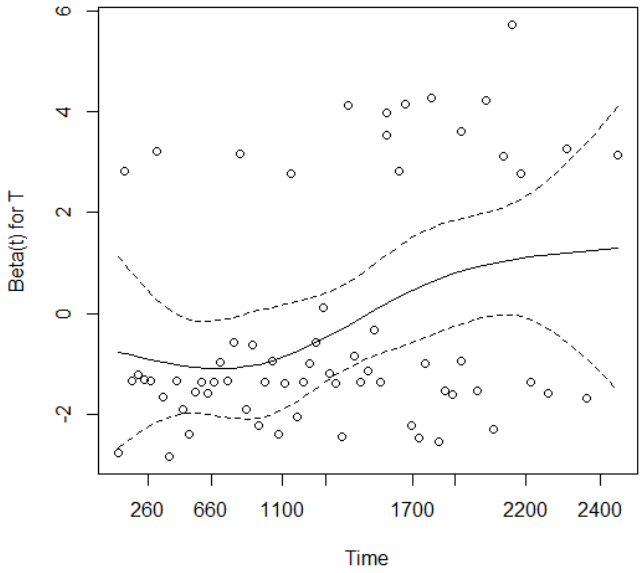

Figura 4.7 Residus de Schoenfeld per a l'estimació global del model

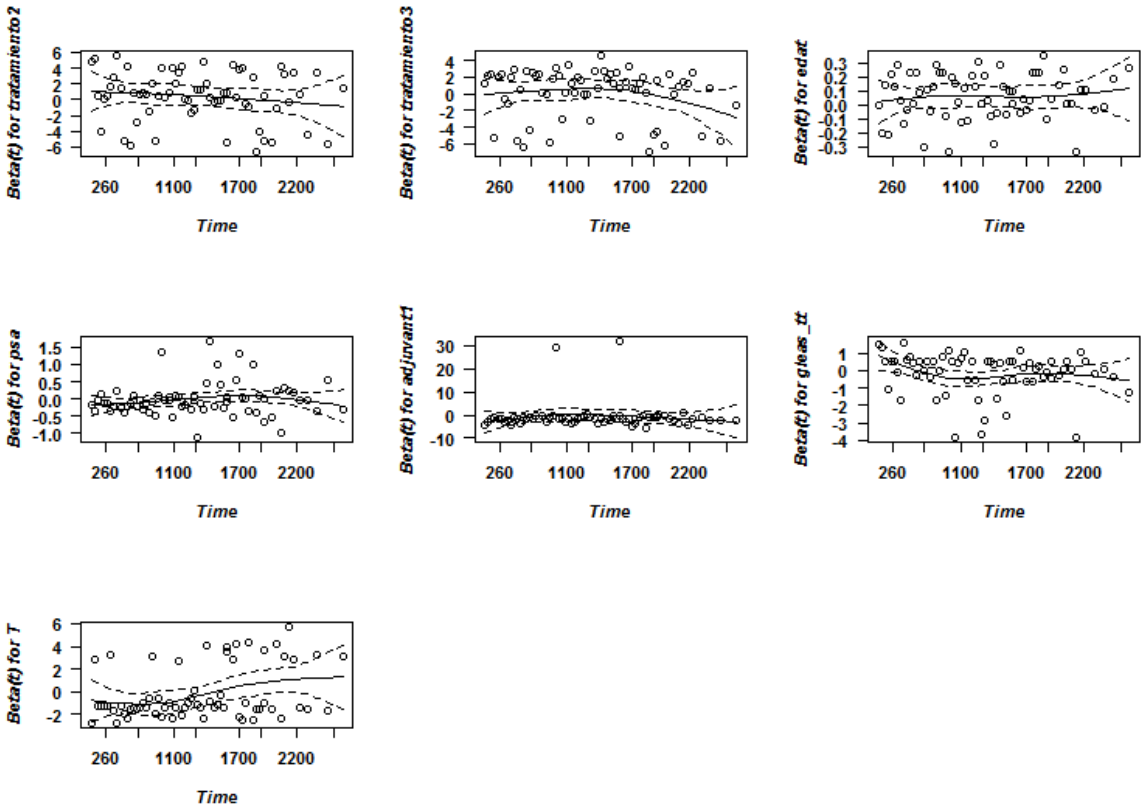

Figura 4.8 Residus de Schoenfeld per a cada variable inclosa al model

Mitjançant l'anàlisi de residus que hem dut a terme observem que els supòsits del model de Cox es semblen complir raonablement bé. El fet de no haver pogut detectar factors de risc de la supervivència, amb l'excepció de l'edat, amb un nivell de significació de 0,05, pot tenir a veure amb l'elevat percentatge de censura.

5. Discussió

Sota el títol d'aquest projecte, Supervivència de pacients amb càncer de pròstata localitzat amb models de riscos proporcionals, hem donat suport a la Dra. Montse Ferrer en l'anàlisi de la supervivència en pacients amb càncer de pròstata per als diferents subgrups d'interès clínic. A més a més hem volgut enriquir el document amb l'estimació d'un model de riscos proporcionals per al temps de supervivència dels pacients inclosos en l'estudi incloent-hi diverses variables d'interès clínic i comprovant les hipòtesis del model mitjançant l'anàlisi de residus.

La metodologia estadística que hem utilitzat per a la descripció de la mostra és l'anàlisi univariant i bivariant. Aquestes tècniques s'han estudiat tant en la diplomatura com en la llicenciatura i el màster i ens són útils per tenir una visió general de com es comporta la mostra en estudi.

Tant l'anàlisi de la supervivència, com el model de riscos proporcionals i l'anàlisi dels seus residus són tècniques que coneixia anteriorment però en les que no hi havia aprofundit ja que no havia realitzat l'assignatura en qüestió. Aquesta part del projecte m'ha permès aprofundir en la recerca dels mètodes més adients en l'anàlisi de la mostra en estudi.

A continuació mencionem tant les conclusions epidemiològiques com estadístiques que hem extret del projecte, així com les pròpies limitacions de l'estudi i les millores plantejades.

5.1 Conclusions epidemiològiques

Els pacients amb càncer de pròstata tenen una probabilitat estimada de supervivència als 5 anys de més del 90%, si desglossem aqueta probabilitat per grup de tractament, els pacients als quals se'ls ha realitzat una prostatectomia radical tenen una major probabilitat estimada de sobreviure, 97%, seguit dels pacients tractats amb raidioteràpia i braquiteràpia (91% i 90% respectivament). Pel que fa al model de Cox, destaquem que el fet de rebre tractament hormonal adjuvant és el factor més protector en el temps de supervivència dels pacients amb càncer de pròstata.

5.2 Conclusions estadístiques

L'objectiu principal d'aquest estudi, a banda de descriure la supervivència dels pacients amb CP, estadísticament era realitzar l'anàlisi dels residus, ja que corrobora la validesa del model i molts cops, en la pràctica habitual no s'acostuma a fer. Al seu favor hem de dir que el fet d'analitzar els residus no ens dona només validesa al model, sinó que ens permet detectar observacions influents en la mostra estudiada.

Un punt que hagués estat d'interès, si haguéssim disposat de més temps, és corroborar els resultats obtinguts en la tesis de López per a una variable explicativa de 3 categories, com és el cas de la nostra variable tractament. Hem assumit els seus resultats com a vàlids per realitzar el nostre anàlisi suposant que obtindríem un comportament semblant.

5.3 Limitacions de l'estudi

El disseny de l'estudi, el fet que l'estudi sigui observacional no ens garanteix que els pacients es distribueixin igual en els tres grups de tractament, per exemple la BT es prescriu a pacients amb menor risc i la PR en els pacients més joves

El fet de no haver finalitzat l'estudi i que la causa principal de mort dels pacients amb CP no és el càncer en si, sinó l'edat o altres malalties fa que tinguem un nombre mol elevat de censures i que no obtinguem un resultats clars en l'anàlisi de la supervivència. Es d'esperar que al finalitzar l'estudi, als 10 anys de seguiment es puguin extreure unes conclusions més concretes referents a la supervivència d'aquests pacients.

Bibliografia

[1]INE. Defunciones según la causa de muerte. Año 2009. Datos nacionales. Notas de prensa. 2012.

[2] EUROSTAT. Health statistics – Atlas on mortality in the European Union. 2009 edition.

[3] American Cancer Society. Global Cancer Facts & Figures 2nd Edition. 2008.

[4] Tratatamientos para el cáncer de próstata. Disponible en: www.myprostatecanceroptions.com

[5] Alonso, J; Cunillera, O; Ferrer, M., Garín, O.; Pont, A.; Boladeras, A.; et al. Efectividad a largo plazo de la prostatectomía radical, la braquiterapia y la radioterapia conformacional externa 3D en el cáncer de próstata órgano-confinado. AATRM Núm. 2006/08. Informes, estudios, e investigación. Ministerio de Sanidad y Consumo.

[6] Cáncer de próstata: ¿Qué es el cáncer de próstata?. American Cancer Society. 2011. Disponible en: www.cancer.org

[7] Cáncer de próstata: ¿Qué es el cáncer?. American Cancer Society. 2011. Disponible en: www.cancer.org

[8] Granado de la Orden, S., Saá Requejo, C. i Quintaná Viqueria, A.. Situación epidemiológica del cáncer de próstata en España. Actas Urol Esp. 2006, vol.30, n.6, pp. 574-582. Disponible en: http://scielo.isciii.es

[9] Cáncer de próstata: detección temprana. American Cancer Society. 2010. Disponible en: www.cancer.org

[10] Tratatamientos para el cáncer de próstata. Disponible en: www.myprostatecanceroptions.com

[11] Bill-Axelson, A., Holmberg, A., i Ruutu, M., Häggman, M.; Andersson, S-O.; Bratell, S.; et al. Radical prostatectomy versus watchful waiting in early prostate cancer. N ENGL MED, 2005; Vol. 352, No. 19, pp 1977-1984.

[12] D'Amico, A. V., Whitington, R., Malkowicz, S. B., Schultz, D.; Blank, K.; Broderick, G. A.; et al. Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinicaly localized prostate cancer. JAMA, 1998. Vol. 280, No. 11, pp 969-974.

[13] Stokes, S. H. Comparison of biochemical disease-free survival of patients with localized carcinoma of the prostate undergoing radical prostatectomy, transperienal ultrasound-guided radioactive seed implantations, or definitive external beam irradiation. I. J. Radiation Oncology Biol. Phys., 2000. Vol 47, No 1, pp 129-136.

[14] Ferrer, M., Suárez, J.F., Guedea, F., Fernández, P.; Macías, V.; Mariño, A.; et al. Health-related quality of life 2 years after treatment with radical prostatectomy, prostate brachyterapy, or external beam radiotherapy in patients with clinically localized prostate cancer. I. J. Radioaton Oncology Biol. Phys., 2008. Vol.72, No. 2, pp 421-432.

[15] Holmberg, L.; Bill-Axelson, A.; Helgesen, F.; Salo, J. O.; Folmerz, P.; Häggman, M.; et al. A randomized trial comparing radical prostatectomy with watchful waiting in early prostate cancer. N Engl J Med, 2002. Vol. 347, No. 11, pp 781-789.

[16] D'Amico, A. V.; Manola, J.; Loffredo, M.; Renshaw, A. A.; DellaCroce, A.; Kantoff, P. W. 6-Month Androgen Suppression Plus Radiation Therapy vs Radiation Therapy Alone for Patients With Clinically Localized Prostate Cancer. A randomized controlled trial. JAMA, 2004. Vol. 292, No. 7, pp 821 -- 827.

[17] Crook J.M.;, Gomez-Iturriaga A.; Wallace K.; Ma C.; Fung S.; Alibhai S.; et al..Comparison of health-related quality of life 5 years after SPIRIT: Surgical Prostatectomy Versus Interstitial Radiation Intervention Trial. J Clin Oncol. 2011 Feb 1;29(4):362-8

[18] INE. Defunciones según la Causa de Muerte Año 2013. Notas de prensa, 27 de febrero de 2015. Disponible en: www.ine.es/prensa/prensa.htm.

[19] Gómez, G. Análisis de Supervivencia. Universitat Politècnica de Catalunya, Barcelona 2004.

[20] Kleinbaum DG. Survival analysis. A self-learning text. New York: Springer-Verlag, 1996.

[21] Cox, D.R. Regression models and life-tables. Journal of the Royal Statistical Society. Series B (Methodological), Vol. 34, No. 2. 1972, pp. 187-220.

[22] Ahmed, F.E., Vos, P.W., Holbert, D. Modeling survival in colon cancer: a methodological review. Molecular Cancer 2007, 6:15.

[23] Hanley, J.A.; McNeil, B.J. The meaning and use of the area under a receicer operating characteristic (ROC) curve. Radiology, April 1982, Vol 143. No.1: 29-36.

[24] R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

[25] Peduzzi, P.; Concato, J.; Kemper, E.; Hol ford, T.R.; Feinstein, A. R. A Simulation Study of the Number of Events per Variable in Logistic Regression Analysis. J Clin Epidemiol Vol. 49, No. 12, pp. 1373-1379, 1996.

[26] Vittinghoff, E.; McCulloch, C.E.. Relaxing the Rule of Ten Events per Variable in Logistic and Cox Regression. Am J Epidemiol 2007;165:710–718

[27] Harrell, F.E.; Lee, K.L: Mark, D.B. Tutorial in biostatistics multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in medicine, vol. 15,361-387 1996.

[28] Therneau, T.M. y Grambsch, P.M. Modeling Survival Data. Extending the Cox Model. New York: Springer. (2000)

[29] López, L. Survival data analysis with heavy-censoring and long-term survivors. Universitat Politécnica de Catalunya, Barcelona 2014.

[30] Therneau, T. M. and Grambsch, P. M.. Modeling Survival Data: Extending the Cox Model. Springer, New York, 2000.

[31] Therneau T (2015). _A Package for Survival Analysis in S_. version 2.38, URL: http://CRAN.R-project.org/package=survival.

[32] Ramon Oller and Klaus Langohr (2015). FHtest: Tests for Right and Interval-Censored Survival Data Based on the Fleming-Harrington Class. R package version 1.2

[33] Frank E Harrell Jr (2015). rms: Regression Modeling Strategies. R package version 4.3-1. URL: http://CRAN.R-project.org/package=rms

Apèndix A. Codi R dels anàlisi realitzats

```
#Carrega de les llibreries
library(Hmisc)
library(FHtest)
library(rms)
library(Epi)
library(mvtnorm)
library(vcd)
library(SparseM)
library(catspec)
library(leaps)
library(splines)
library(heplots)

#Lectura de dades
datos<-read.table("F:\\TFM\\Analisi\\bdd_(21_07_10).txt", header=T,dec='.')

## Comprovant que s'ha fet be/ Visualitzant les dades
str(datos)
dim(datos)
names(datos)
View(datos)
Label(datos)

#Podem els labels i levels pertinents a cada variable
label(datos$Ident)<-'Identificador'
label(datos$data)<-'Data'
label(datos$surv)<-'Censura'
label(datos$vol_inic)<-'Prostate volumen'
label(datos$gleas_tt)<-'Gleason'
label(datos$psa)<-'PSA'
label(datos$T)<-'Clinical t stage'
label(datos$tratamiento)<-'Tractament'
label(datos$inici_t)<-'Data inici tractament'
label(datos$time)<-'Temps fins al succes'
label(datos$PRE_EREC)<-'Nerve-sparing'
label(datos$edat)<-'Edat'
label(datos$gedat_4c)<-'Edat 4 grups'
label(datos$gedat_3c)<-'Edat 3 grups'
label(datos$adjuvant)<-'Tractament hormonal com a complement'
label(datos$neoadjuvant)<-'Tractament hormonal anterior a l,inici del
tractament en estudi'
label(datos$horm_rescat)<-'Tractament hormonal de rescat'

levels(datos$surv)<-c('Cens','Mort')
levels(datos$tratamiento)<-c('PR','RT','BT')
levels(datos$grisc)<-c('Baix risc','Risc intergmig')
levels(datos$gedat_4c)<-c('< 60','60 - 65','65-70','>= 70')
levels(datos$gedat_3c)<-c('< 65','65-70','>= 70')
```

```
#Veig que a l'article les etiquetes estan invertides, de moment ho canvio fins
que no parli amb l'Oriol
#levels(datos$adjuvant)<-c('Si','No')
#levels(datos$neoadjuvant)<-c('Si','No')
#levels(datos$horm_rescat)<-c('Si','No')
levels(datos$adjuvant)<-c('No','Si')
levels(datos$neoadjuvant)<-c('No','Si')
levels(datos$horm_rescat)<-c('No','Si')

#Categoritzem el clinical T Stage segons les indicacions de la Dra. Montse
Ferrer
#t1 i 2 ->T1
#t3 i 4 ->T2
datos$T [datos$T == 2] <- 1
datos$T [datos$T >= 3] <- 2

dim(datos)
head(datos)
summary(datos)
#describe(datos)
str(datos)

#Anàlisis descriptiu
summary(datos)
sd(datos)
with(datos,ctab(table(surv)))
with(datos,ctab(table(tratamiento)))

with(datos,ctab(table(gedat_3c,tratamiento),digits=2, type=c("n", "row",
"column", "total"),style="long", row.vars=NULL,
col.vars=NULL,percentages=TRUE,addmargins=TRUE))
with(datos,ctab(table(T,tratamiento),digits=2, type=c("n", "row", "column",
"total"),style="long", row.vars=NULL,
col.vars=NULL,percentages=TRUE,addmargins=TRUE))
with(datos,ctab(table(grisc,tratamiento),digits=2, type=c("n", "row",
"column", "total"),style="long", row.vars=NULL,
col.vars=NULL,percentages=TRUE,addmargins=TRUE))
with(datos,ctab(table(PRE_EREC,tratamiento),digits=2, type=c("n", "row",
"column", "total"),style="long", row.vars=NULL,
col.vars=NULL,percentages=TRUE,addmargins=TRUE))
with(datos,ctab(table(neoadjuvant,tratamiento),digits=2, type=c("n", "row",
"column", "total"),style="long", row.vars=NULL,
col.vars=NULL,percentages=TRUE,addmargins=TRUE))
with(datos,ctab(table(adjuvant,tratamiento),digits=2, type=c("n", "row",
"column", "total"),style="long", row.vars=NULL,
col.vars=NULL,percentages=TRUE,addmargins=TRUE))
with(datos,ctab(table(horm_rescat,tratamiento),digits=2, type=c("n", "row",
"column", "total"),style="long", row.vars=NULL,
col.vars=NULL,percentages=TRUE,addmargins=TRUE))
with(datos,ctab(table(surv,tratamiento),digits=2, type=c("n", "row", "column",
"total"),style="long", row.vars=NULL,
col.vars=NULL,percentages=TRUE,addmargins=TRUE))

summary(datos$edat)
sd(datos$edat,na.rm = TRUE)
with(datos,tapply(edat,tratamiento,summary))
with(datos,tapply(edat,tratamiento,sd,na.rm = TRUE))
summary(datos$psa)
```

```
sd(datos$psa,na.rm = TRUE)
with(datos,tapply(psa,tratamiento,summary))
with(datos,tapply(psa,tratamiento,sd,na.rm = TRUE))
summary(datos$gleas_tt)
sd(datos$gleas_tt,na.rm = TRUE)
with(datos,tapply(gleas_tt,tratamiento,summary))
with(datos,tapply(gleas_tt,tratamiento,sd,na.rm = TRUE))


##ANÀLISI DE LA SUPERVIVÈNCIA##
################################

datos$tratamiento<-as.factor(datos$tratamiento)
#datos$gedat_4c<-as.factor(datos$gedat_4c)
datos$gedat_3c<-as.factor(datos$gedat_3c)
datos$grisc<-as.factor(datos$grisc)
datos$adjuvant<-as.factor(datos$adjuvant)
datos$neoadjuvant<-as.factor(datos$neoadjuvant)


#Corba de supervivència mitjançant survival
prost<-with(datos,Surv(time,surv),sclae=365.25)
(svfdt<- survfit(prost ~1, type="fh2"))
par(font.main=2,font.lab=2,font.axis=2,las=1)
par(mfrow=c(1,1),font.main=1)
par(font.main=2,font.lab=2,font.axis=2,las=1)
plot(svfdt,lwd=3,xlab="Tieme, y",ylab="Survival,
%",col=c(colors()[574]),yscale=100,xscale=365.25,mark.time=FALSE)
#summary(svfdt)
summary(svfdt,scale=365.25,times=c(0,365.25,730.5,1095.75,1461,1826.25,2191.5,
2550,2922))

#Corba de supervivència mitjançant rms
datos$time2<-datos$time/365.25
#Grafiquem la funció de supervivencia
nprost<- npsurv(formula = Surv(time2,surv== 1) ~ 1, data = datos)
levels(datos$tratamiento)<-c('PR','RT','BT')
survplot(fit=nprost,xlab="Temps
(anys)",ylab="Supervivència",time.inc=1,col=c(colors()[574])
,conf="bands",conf.int=.95,n.risk=TRUE,adj.n.risk=0)


attach(datos)

## SUPERVIVÈNCIA EN FUNCIÓ DE TRACTAMENT
(surv_ttp<-survfit(prost~tratamiento))
summary(surv_ttp)

#Corba de supervivència mitjançant survival
par(mfrow=c(1,1),font.main=1)
par(font.main=2,font.lab=2,font.axis=2,las=1)
plot(survfit(prost~tratamiento),conf.int=F,lwd=3,col=c(colors()[574],colors()[
376],colors()[461]),lty=1,xlab="Time (years)",ylab="Survival
(%)",yscale=100,xscale=365.25,mark.time=FALSE)
legend('bottomleft',inset=0,c('PR','RT','BT'),col=c(colors()[574],colors()[376
],colors()[461]),lty=1,xjust=1)
text(6, 0, "Log-Rank p=0.132", cex = .8)
```

```
timev<-summary(survfit(prost~tratamiento),scale=365.25)$time
summary(survfit(prost~tratamiento),scale=365.25,times=c(0,365.25,730.5,1095.75
,1461,1826.25,2191.5,2556.75))
summary(survfit(prost~tratamiento),scale=365.25,times=c(0,365.25,730.5,1095.75
,1461,1826.25,2191.5,2550))

#Corba de supervivència mitjançant rms
par(mfrow=c(1,2),font.main=1)
nprost<- npsurv(formula = Surv(time2,surv== 1) ~ tratamiento, data = datos)
survplot(fit=nprost,xlab="Temps
(anys)",ylab="Supervivència",time.inc=1,col=c(colors()[574],colors()[376],colo
rs()[461])
,conf="none",n.risk=TRUE,adj.n.risk=0,levels.only=TRUE)
survplot(fit=nprost,xlab="Temps
(anys)",ylab="Supervivència",ylim=c(0.5,1),time.inc=1,col=c(colors()[574],colo
rs()[376],colors()[461])
,conf="none",n.risk=TRUE,adj.n.risk=0,levels.only=TRUE)

#Log rank (rho=0)
survdiff(prost~tratamiento,rho=0)
#Peto (rho=1)
survdiff(prost~tratamiento,rho=1)

#Utilitzem la funció FHtest
FHtestrcc(Surv(time,surv)~tratamiento,data=datos,rho=0,lambda=0)
FHtestrcc(Surv(time,surv)~tratamiento,data=datos,rho=1,lambda=0)
FHtestrcc(Surv(time,surv)~tratamiento,data=datos,rho=1,lambda=1)


## SUPERVIVÈNCIA EN FUNCIÓ DEL GRUP D'EDAT

datosedat3c_1<- subset(datos, gedat_3c==1)
detach(datos)
attach(datosedat3c_1)

proste3_1<-with(datosedat3c_1,Surv(time,surv),scale=365.25)

(surve3_1<-survfit(proste3_1~tratamiento,data=datosedat3c_1))
summary(surve3_1)

#Tratamiento, gedat_3c=1
(surve3_1<-survfit(proste3_1~tratamiento,data=datosedat3c_1))
summary(surve3_1)
summary(survfit(proste3_1~tratamiento),times=c(0,365.25,730.5,1095.75,1461,182
6.25,2191.5,2556.75))

par(mfrow=c(1,3),font.main=1)
nproste1<- npsurv(formula = Surv(time2,surv== 1) ~ tratamiento, data =
datosedat3c_1)
survplot(fit=nproste1,xlab="Temps
(anys)",ylab="Supervivència",time.inc=1,col=c(colors()[574],colors()[376],colo
rs()[461])
,conf="none",n.risk=TRUE,adj.n.risk=0,levels.only=TRUE)
title('Edat < 65 anys')

#Comaprem les corbes
FHtestrcc(Surv(time,surv)~tratamiento,data=datosedat3c_1,rho=0,lambda=0)
FHtestrcc(Surv(time,surv)~tratamiento,data=datosedat3c_1,rho=1,lambda=0)
```

```
FHtestrcc(Surv(time,surv)~tratamiento,data=datosedat3c_1,rho=1,lambda=1)

#Tratamiento, gedat_3c=2
datosedat3c_2<- subset(datos, gedat_3c==2)

detach(datosedat3c_1)
attach(datosedat3c_2)
with(datosedat3c_2,ctab(table(gedat_3c)))
proste3_2<-with(datosedat3c_2,Surv(time,surv),scale=365.25)

(surve3_2<-survfit(proste3_2~tratamiento,data=datosedat3c_2))
summary(surve3_2)
summary(survfit(proste3_2~tratamiento),times=c(0,365.25,730.5,1095.75,1461,182
6.25,2191.5,2556.75))

nproste2<- npsurv(formula = Surv(time2,surv== 1) ~ tratamiento, data =
datosedat3c_2)
survplot(fit=nproste2,xlab="Temps
(anys)",ylab="Supervivència",time.inc=1,col=c(colors()[574],colors()[376],colo
rs()[461])
,conf="none",n.risk=TRUE,adj.n.risk=0,levels.only=TRUE)
title('Edat [65,70) anys')

#Comaprem les corbes
FHtestrcc(Surv(time,surv)~tratamiento,data=datosedat3c_2,rho=0,lambda=0)
FHtestrcc(Surv(time,surv)~tratamiento,data=datosedat3c_2,rho=1,lambda=0)
FHtestrcc(Surv(time,surv)~tratamiento,data=datosedat3c_2,rho=1,lambda=1)

#Tratamiento, gedat_3c=3
datosedat3c_3<- subset(datos, gedat_3c==3)
detach(datosedat3c_2)
attach(datosedat3c_3)
proste3_3<-with(datosedat3c_3,Surv(time,surv),scale=365.25)

(surve3_3<-survfit(proste3_3~tratamiento,data=datosedat3c_3))
summary(surve3_3)
summary(survfit(proste3_3~tratamiento),times=c(0,365.25,730.5,1095.75,1461,182
6.25,2191.5,2556.75))

nproste3<- npsurv(formula = Surv(time2,surv== 1) ~ tratamiento, data =
datosedat3c_3)
survplot(fit=nproste3,xlab="Temps
(anys)",ylab="Supervivència",time.inc=1,col=c(colors()[574],colors()[376],colo
rs()[461])
,conf="none",n.risk=TRUE,adj.n.risk=0,levels.only=TRUE)
title('Edat >= 70 anys')

#Comaprem les corbes
FHtestrcc(Surv(time,surv)~tratamiento,data=datosedat3c_3,rho=0,lambda=0)
FHtestrcc(Surv(time,surv)~tratamiento,data=datosedat3c_3,rho=1,lambda=0)
FHtestrcc(Surv(time,surv)~tratamiento,data=datosedat3c_3,rho=1,lambda=1)


## SUPERVIVÈNCIA EN FUNCIÓ DEL T STAGE

#Tratamiento, T=1
datosT1_1<- subset(datos, T==1)
detach(datosedat3c_3)
```

```
attach(datosT1_1)
with(datosT1_1,ctab(table(T)))
prosT1_1<-with(datosT1_1,Surv(time,surv),scale=365.25)

(survT1_1<-survfit(prosT1_1~tratamiento,data=datosT1_1))
summary(survT1_1)

par(mfrow=c(1,2),font.main=1)
nprostt1<- npsurv(formula = Surv(time2,surv== 1) ~ tratamiento, data =
datosT1_1)
survplot(fit=nprostt1,xlab="Temps
(anys)",ylab="Supervivència",time.inc=1,col=c(colors()[574],colors()[376],colo
rs()[461])
,conf="none",n.risk=TRUE,adj.n.risk=0,levels.only=TRUE)
title('T1')

#Comaprem les corbes
FHtestrcc(Surv(time,surv)~tratamiento,data=datosT1_1,rho=0,lambda=0)
FHtestrcc(Surv(time,surv)~tratamiento,data=datosT1_1,rho=1,lambda=0)
FHtestrcc(Surv(time,surv)~tratamiento,data=datosT1_1,rho=1,lambda=1)


#Tratamiento, T=2

datosT2_1<- subset(datos, T==2)
detach(datosT1_1)
attach(datosT2_1)
with(datosT2_1,ctab(table(T)))
prosT2_1<-with(datosT2_1,Surv(time,surv),scale=365.25)

(survT2_1<-survfit(prosT2_1~tratamiento,data=datosT2_1))
summary(survT2_1)
summary(survfit(prosT2_1~tratamiento),times=c(0,365.25,730.5,1095.75,1461,1826
.25,2191.5,2556.75))

nprostt2<- npsurv(formula = Surv(time2,surv== 1) ~ tratamiento, data =
datosT2_1)
survplot(fit=nprostt2,xlab="Temps
(anys)",ylab="Supervivència",time.inc=1,col=c(colors()[574],colors()[376],colo
rs()[461])
,conf="none",n.risk=TRUE,adj.n.risk=0,levels.only=TRUE)
title('T2')

FHtestrcc(Surv(time,surv)~tratamiento,data=datosT2_1,rho=0,lambda=0)
FHtestrcc(Surv(time,surv)~tratamiento,data=datosT2_1,rho=1,lambda=0)
FHtestrcc(Surv(time,surv)~tratamiento,data=datosT2_1,rho=1,lambda=1)


## SUPERVIVÈNCIA EN FUNCIÓ DEL GRUP DE RISC

#RISC=0#
datosrisc<- subset(datos, grisc==0)
detach(datosT2_1)
attach(datosrisc)
with(datosrisc,ctab(table(grisc)))
prost_g0<-with(datosrisc,Surv(time,surv),scale=365.25)

(surv_r0<-survfit(prost_g0~tratamiento,data=datosrisc))
```

```
summary(surv_r0)
summary(survfit(prost_g0~tratamiento),times=c(0,365.25,730.5,1095.75,1461,1826
.25,2191.5,2556.75))

nprostg1<- npsurv(formula = Surv(time2,surv== 1) ~ tratamiento, data =
datosrisc)
survplot(fit=nprostg1,xlab="Temps
(anys)",ylab="Supervivència",time.inc=1,col=c(colors()[574],colors()[376],colo
rs()[461])
,conf="none",n.risk=TRUE,adj.n.risk=0,levels.only=TRUE)
title('Risc baix')

FHtestrcc(Surv(time,surv)~tratamiento,data=datosrisc,rho=0,lambda=0)
FHtestrcc(Surv(time,surv)~tratamiento,data=datosrisc,rho=1,lambda=0)
FHtestrcc(Surv(time,surv)~tratamiento,data=datosrisc,rho=1,lambda=1)

#RISC=1
datosrisc1<- subset(datos, grisc==1)
detach(datosrisc)
attach(datosrisc1)
with(datosrisc1,ctab(table(grisc)))
prost_g1<-with(datosrisc1,Surv(time,surv),scale=365.25)
(surv_r1<-survfit(prost_g1~tratamiento,data= datosrisc1))
summary(surv_r1)
summary(survfit(prost_g1~tratamiento),times=c(0,365.25,730.5,1095.75,1461,1826
.25,2191.5,2556.75))

nprostg2<- npsurv(formula = Surv(time2,surv== 1) ~ tratamiento, data =
datosrisc1)
survplot(fit=nprostg2,xlab="Temps
(anys)",ylab="Supervivència",time.inc=1,col=c(colors()[574],colors()[376],colo
rs()[461])
,conf="none",n.risk=TRUE,adj.n.risk=0,levels.only=TRUE)
title('Risc moderat')

FHtestrcc(Surv(time,surv)~tratamiento,data=datosrisc1,rho=0,lambda=0)
FHtestrcc(Surv(time,surv)~tratamiento,data=datosrisc1,rho=1,lambda=0)
FHtestrcc(Surv(time,surv)~tratamiento,data=datosrisc1,rho=1,lambda=1)


## SUPERVIVÈNCIA EN FUNCIÓ DEL TRACTAMENT HORMONAL PREVI
#Tractament, neoadjuvant=0

datosNeo0_1<- subset(datos, neoadjuvant==0)

detach(datosrisc1)
attach(datosNeo0_1)
prosNeo0_1<-with(datosNeo0_1,Surv(time,surv),scale=365.25)

(survNeo0_1<-survfit(prosNeo0_1~tratamiento,data=datosNeo0_1))
summary(survNeo0_1)
summary(survfit(prosNeo0_1~tratamiento),times=c(0,365.25,730.5,1095.75,1461,18
26.25,2191.5,2556.75))

nprostn0<- npsurv(formula = Surv(time2,surv== 1) ~ tratamiento, data =
datosNeo0_1)
```

```
survplot(fit=nprostn0,xlab="Temps
(anys)",ylab="Supervivència",time.inc=1,col=c(colors()[574],colors()[376],colo
rs()[461])
,conf="none",n.risk=TRUE,adj.n.risk=0,levels.only=TRUE)
title('Sense tractament hormonal previ')

FHtestrcc(Surv(time,surv)~tratamiento,data=datosNeo0_1,rho=0,lambda=0)
FHtestrcc(Surv(time,surv)~tratamiento,data=datosNeo0_1,rho=1,lambda=0)
FHtestrcc(Surv(time,surv)~tratamiento,data=datosNeo0_1,rho=1,lambda=1)

#Tractament neoadjuvant=1

datosNeo1_1<- subset(datos, neoadjuvant==1)
detach(datosNeo0_1)
attach(datosNeo1_1)
prosNeo1_1<-with(datosNeo1_1,Surv(time,surv),scale=365.25)
(survNeo1_1<-survfit(prosNeo1_1~tratamiento,data=datosNeo1_1))
summary(survNeo1_1)
summary(survfit(prosNeo1_1~tratamiento),times=c(0,365.25,730.5,1095.75,1461,18
26.25,2191.5,2556.75))

nprostn1<- npsurv(formula = Surv(time2,surv== 1) ~ tratamiento, data =
datosNeo1_1)
survplot(fit=nprostn1,xlab="Temps
(anys)",ylab="Supervivència",time.inc=1,col=c(colors()[574],colors()[376],colo
rs()[461])
,conf="none",n.risk=TRUE,adj.n.risk=0,levels.only=TRUE)
title('Tractament hormonal previ')

FHtestrcc(Surv(time,surv)~tratamiento,data=datosNeo1_1,rho=0,lambda=0)
FHtestrcc(Surv(time,surv)~tratamiento,data=datosNeo1_1,rho=1,lambda=0)
FHtestrcc(Surv(time,surv)~tratamiento,data=datosNeo1_1,rho=1,lambda=1)


## SUPERVIVÈNCIA EN FUNCIÓ DEL TRACTAMENT HORMONAL ADJUVANT

#Tractament adjuvant=0

datosAdj0_1<- subset(datos, adjuvant==0)
detach(datosNeo1_1)
attach(datosAdj0_1)
prosAdj0_1<-with(datosAdj0_1,Surv(time,surv),scale=365.25)

(survAdj0_1<-survfit(prosAdj0_1~tratamiento,data=datosAdj0_1))
summary(survAdj0_1)
summary(survfit(prosAdj0_1~tratamiento),times=c(0,365.25,730.5,1095.75,1461,18
26.25,2191.5,2556.75))

nprosta0<- npsurv(formula = Surv(time2,surv== 1) ~ tratamiento, data =
datosAdj0_1)
survplot(fit=nprosta0,xlab="Temps
(anys)",ylab="Supervivència",time.inc=1,col=c(colors()[574],colors()[376],colo
rs()[461])
,conf="none",n.risk=TRUE,adj.n.risk=0,levels.only=TRUE)
title('Sense tractament adjuvant')

FHtestrcc(Surv(time,surv)~tratamiento,data=datosAdj0_1,rho=0,lambda=0)
FHtestrcc(Surv(time,surv)~tratamiento,data=datosAdj0_1,rho=1,lambda=0)
```

```
FHtestrcc(Surv(time,surv)~tratamiento,data=datosAdj0_1,rho=1,lambda=1)

#Tratcament adjuvant=1
datosAdj1_1<- subset(datos, adjuvant==1)

detach(datosAdj0_1)
attach(datosAdj1_1)
prosAdj1_1<-with(datosAdj1_1,Surv(time,surv),scale=365.25)
(survAdj1_1<-survfit(prosAdj1_1~tratamiento,data=datosAdj1_1))
summary(survAdj1_1)
summary(survfit(prosAdj1_1~tratamiento),times=c(0,365.25,730.5,1095.75,1461,18
26.25,2191.5,2556.75))

nprosta1<- npsurv(formula = Surv(time2,surv== 1) ~ tratamiento, data =
datosAdj0_1)
survplot(fit=nprosta1,xlab="Temps
(anys)",ylab="Supervivència",time.inc=1,col=c(colors()[574],colors()[376],colo
rs()[461])
,conf="none",n.risk=TRUE,adj.n.risk=0,levels.only=TRUE)
title('Tractament adjuvant')

FHtestrcc(Surv(time,surv)~tratamiento,data=datosAdj1_1,rho=0,lambda=0)
FHtestrcc(Surv(time,surv)~tratamiento,data=datosAdj1_1,rho=1,lambda=0)
FHtestrcc(Surv(time,surv)~tratamiento,data=datosAdj1_1,rho=1,lambda=1)


## SUPERVIVÈNCIA EN FUNCIÓ DEL TRACTAMENT HORMONAL DE RESCAT

#Tratcament horm_rescat=0
datosRes0_1<- subset(datos, horm_rescat==0)
detach(datosAdj1_1)
attach(datosRes0_1)
prosRes0_1<-with(datosRes0_1,Surv(time,surv),scale=365.25)
(survRes0_1<-survfit(prosRes0_1~tratamiento,data=datosRes0_1))
summary(survRes0_1)
summary(survfit(prosRes0_1~tratamiento),times=c(0,365.25,730.5,1095.75,1461,18
26.25,2191.5,2556.75))

nprostr0<- npsurv(formula = Surv(time2,surv== 1) ~ tratamiento, data =
datosRes0_1)
survplot(fit=nprostr0,xlab="Temps
(anys)",ylab="Supervivència",time.inc=1,col=c(colors()[574],colors()[376],colo
rs()[461])
,conf="none",n.risk=TRUE,adj.n.risk=0,levels.only=TRUE)
title('Sense tractament de rescat')

FHtestrcc(Surv(time,surv)~tratamiento,data=datosRes0_1,rho=0,lambda=0)
FHtestrcc(Surv(time,surv)~tratamiento,data=datosRes0_1,rho=1,lambda=0)
FHtestrcc(Surv(time,surv)~tratamiento,data=datosRes0_1,rho=1,lambda=1)

#Tratamiento, horm_rescat=1
datosRes1_1<- subset(datos, horm_rescat==1)
detach(datosRes0_1)
attach(datosRes1_1)
with(datosRes1_1,ctab(table(horm_rescat)))
prosRes1_1<-with(datosRes1_1,Surv(time,surv),scale=365.25)
(survRes1_1<-survfit(prosRes1_1~tratamiento,data=datosRes1_1))
summary(survRes1_1)
```

```
summary(survfit(prosRes1_1~tratamiento),times=c(0,365.25,730.5,1095.75,1461,18
26.25,2191.5,2556.75))

nprostr1<- npsurv(formula = Surv(time2,surv== 1) ~ tratamiento, data =
datosRes1_1)
survplot(fit=nprostr1,xlab="Temps
(anys)",ylab="Supervivència",time.inc=1,col=c(colors()[574],colors()[376],colo
rs()[461])
,conf="none",n.risk=TRUE,adj.n.risk=0,levels.only=TRUE)
title('Tractament de rescat')

FHtestrcc(Surv(time,surv)~tratamiento,data=datosRes1_1,rho=0,lambda=0)
FHtestrcc(Surv(time,surv)~tratamiento,data=datosRes1_1,rho=1,lambda=0)
FHtestrcc(Surv(time,surv)~tratamiento,data=datosRes1_1,rho=1,lambda=1)

##MODEL DE COX ##
##################

#Correlació entre les variables d'interés
detach(datosRes1_1)
attach(datos)

#Edat
cor.test(datos$edat, datos$psa, alternative = "two.sided",method = "pearson")
plot(datos$psa~datos$edat)

aov1 = aov(edat~adjuvant,datos)
summary(aov1)
boxplot(datos$edat~datos$adjuvant)
etasq(aov1)
aov1 = aov(edat~gleas_tt,datos)
summary(aov1)
boxplot(datos$edat~datos$gleas_tt)
etasq(aov1)
aov1 = aov(edat~T,datos)
summary(aov1)
boxplot(datos$edat~datos$T)
etasq(aov1)

#PSA
aov1 = aov(psa~adjuvant,datos)
summary(aov1)
boxplot(datos$psa~datos$adjuvant)
etasq(aov1)
aov1 = aov(psa~gleas_tt,datos)
summary(aov1)
boxplot(datos$psa~datos$gleas_tt)
etasq(aov1)
aov1 = aov(psa~T,datos)
summary(aov1)
boxplot(datos$psa~datos$T)
etasq(aov1)

#ADJ
tabla = ftable(as.factor(datos$adjuvant), datos$gleas_tt)
assocstats(tabla)

tabla = ftable(as.factor(datos$adjuvant), datos$T)
```

```
assocstats(tabla)

#Gleas T
tabla = ftable(as.factor(datos$gleas_tt), datos$T)
assocstats(tabla)

#Construim el model de Cox
dades <- subset(datos, complete.cases(edat))
prost<-with(dades, Surv(time,surv),sclae=365.25)
(cox5<-coxph(prost ~ tratamiento+edat+psa+adjuvant+gleas_tt+T, dades))
summary(cox5)

## Hazard ratios
ci.lin(cox5)
round(ci.lin(cox5, Exp=T), 3)

## Estimació de la funció de risc basal
coxph.detail(cox5)
coxph.detail(cox5)$hazard

#Representació gràfica
windows(width=8)
par(font=2, font.lab=4, font.axis=2, las=1, mar=c(5, 5, 4, 2))
with(coxph.detail(cox5), plot(hazard~time, xlab='Days',
     ylab=expression(hat(lambda)(t)), pch=16))
with(coxph.detail(cox5), lines(lowess(hazard~time), type='l', lwd=3))
with(coxph.detail(cox5), lines(lowess(hazard~time, f=1/3), type='l', lwd=3,
col=2))
title('Estimació de la funció hazard')

## Model basat en la predicció
survfit(cox5)
plot(survfit(cox5), lwd=2, col=3)

newdata <- data.frame(tratamiento=factor(rep(c('1','2','3'), 2)), edat=67.3,
psa=7.45,

adjuvant=factor(rep(c('0','1'),3)),gleas_tt=5.8,T=factor(rep(c('1','2'),3)))
newdata
survfit(cox5, newdata=newdata)
summary(survfit(cox5, newdata=newdata))

windows(width=8)
par(font=2, font.lab=4, font.axis=2, las=1)
plot(survfit(cox5, newdata=newdata), col=1:4, lwd=3, xlab='Dies')
title('Model basat en prediccions de la funció de supervivència')

graphics.off()

## Bondat d'ajust del model de COX
residuals(cox5)

resids1 <- residuals(update(cox5, ~.-edat))
resids2 <- residuals(update(cox5, ~.-psa))
resids3 <- residuals(update(cox5, ~.-gleas_tt))

windows(width=12, height=5)
par(mfrow=c(1, 3), font=2, font.lab=2, font.axis=2, las=1)
```

```
plot(resids1~dades$edat, xlab='Edat', ylab='Residus', pch=19)
abline(h=0, lwd=2, lty=2)
lines(lowess(dades[-c(1:6), 'edat'], resids1[-c(1:6)]), lwd=2)
plot(resids2~dades$psa, xlab='PSA', ylab='Residus', pch=19)
abline(h=0, lwd=2, lty=2)
lines(lowess(dades[, 'psa'], resids2), lwd=2)
plot(resids3~dades$gleas_tt, xlab='Gleason score', ylab='Residus', pch=19)
abline(h=0, lwd=2, lty=2)
lines(lowess(dades[, 'gleas_tt'], resids3), lwd=2)

## Residus de la deviance: ajust global
windows(width=8)
par(font=2, font.lab=4, font.axis=2, las=1)
plot(residuals(cox5, type='dev')~predict(cox5), xlab='Predictor lineal',
     ylab='Residus basats en la deviancia', pch=16)

## Residus dfbeta: comprovem la influéncia de les observacions de forma
individual
dfbet <- residuals(cox5, type='dfbeta')
dim(dfbet)
windows(width=12, height=7)
par(mfrow=c(2, 3), font=2, font.lab=2, font.axis=2, las=1)
for (i in 1:6){
  plot(dfbet[, i], pch=16, ylab=names(coef(cox5))[i])
  axis(1, at = seq(5, 45, 5))
}
par(mfrow=c(1, 1), font=2, font.lab=2, font.axis=2, las=1)
plot(dfbet[, 7], pch=16, ylab=names(coef(cox5))[7])
  axis(1, at = seq(5, 45, 5))

#Tractament adjuvant tenim 2 observacions influents
round(dfbet[,5],2)
#Pacients 108 i 535
head(datos, 108)
head(datos, 535)

#Mirem els pacients que reben tto adjuvant quina mitja de temps tenen

datos0 = subset(datos, adjuvant==0)
datos1 = subset(datos, adjuvant==1)
summary(datos0$time)
summary(datos1$time)

## Podem assumir proporcionalitat de riscos?
cox.zph(cox5)
plot(cox.zph(cox5))
windows(width=12, height=7)
par(mfrow=c(3, 3), font=2, font.lab=4, font.axis=2, las=1)
plot(cox.zph(cox5))


###Calculem el model sense les 2 observacions influents

dadesb <- subset(dades,Ident!=109236)
dades <- subset(dadesb,Ident!=109003)
prost<-with(dades, Surv(time,surv),sclae=365.25)
(cox5<-coxph(prost ~ tratamiento+edat+psa+adjuvant+gleas_tt+T, dades))
summary(cox5)
```

```
## hazard ratios
ci.lin(cox5)
round(ci.lin(cox5, Exp=T), 3)

## Estimació de la funció de risc basal
coxph.detail(cox5)
coxph.detail(cox5)$hazard

# Representació gràfica
windows(width=8)
par(font=2, font.lab=4, font.axis=2, las=1, mar=c(5, 5, 4, 2))
with(coxph.detail(cox5), plot(hazard~time, xlab='Days',
     ylab=expression(hat(lambda)(t)), pch=16))
with(coxph.detail(cox5), lines(lowess(hazard~time), type='l', lwd=3))
with(coxph.detail(cox5), lines(lowess(hazard~time, f=1/3), type='l', lwd=3,
col=2))
title('Estimation of the hazard function')

## Model basat en la predició
survfit(cox5)
plot(survfit(cox5), lwd=2, col=3)

newdata <- data.frame(tratamiento=factor(rep(c('1','2','3'), 2)), edat=67.3,
psa=7.45,

adjuvant=factor(rep(c('0','1'),3)),gleas_tt=5.8,T=factor(rep(c('1','2'),3)))
newdata
survfit(cox5, newdata=newdata)
summary(survfit(cox5, newdata=newdata))

windows(width=8)
par(font=2, font.lab=4, font.axis=2, las=1)
plot(survfit(cox5, newdata=newdata), col=1:4, lwd=3, xlab='Dies')
title('Model basat en prediccions de la funció de supervivència')

graphics.off()

## Bondat d'ajust del model de COX sense les observacions influents
residuals(cox5)

resids1 <- residuals(update(cox5, ~.-edat))
resids2 <- residuals(update(cox5, ~.-psa))
resids3 <- residuals(update(cox5, ~.-gleas_tt))

windows(width=12, height=5)
par(mfrow=c(1, 3), font=2, font.lab=2, font.axis=2, las=1)
plot(resids1~dades$edat, xlab='Edat', ylab='Residus', pch=19)
abline(h=0, lwd=2, lty=2)
lines(lowess(dades[-c(1:6), 'edat'], resids1[-c(1:6)]), lwd=2)
plot(resids2~dades$psa, xlab='PSA', ylab='Residus', pch=19)
abline(h=0, lwd=2, lty=2)
lines(lowess(dades[, 'psa'], resids2), lwd=2)
plot(resids3~dades$gleas_tt, xlab='Gleason score', ylab='Residus', pch=19)
abline(h=0, lwd=2, lty=2)
lines(lowess(dades[, 'gleas_tt'], resids3), lwd=2)

## Residus de la deviance: ajust global
```

```
windows(width=8)
par(font=2, font.lab=4, font.axis=2, las=1)
plot(residuals(cox5, type='dev')~predict(cox5), xlab='Predictor lineal',
     ylab='Residus basats en la Deviancia', pch=16)

## Residus dfbeta: comprovem la influéncia de les observacions de forma
individual
dfbet <- residuals(cox5, type='dfbeta')
dim(dfbet)
windows(width=12, height=7)
par(mfrow=c(2, 3), font=2, font.lab=2, font.axis=2, las=1)
for (i in 1:6){
  plot(dfbet[, i], pch=16, ylab=names(coef(cox5))[i])
  axis(1, at = seq(5, 45, 5))
}
par(mfrow=c(1, 1), font=2, font.lab=2, font.axis=2, las=1)
plot(dfbet[, 7], pch=16, ylab=names(coef(cox5))[7])
  axis(1, at = seq(5, 45, 5))

## Podem assumir proporcionalitat de riscos?
cox.zph(cox5)
plot(cox.zph(cox5))
windows(width=12, height=7)
par(mfrow=c(3, 3), font=2, font.lab=4, font.axis=2, las=1)
plot(cox.zph(cox5))
```

Apèndix B. Corbes de supervivència addicionals

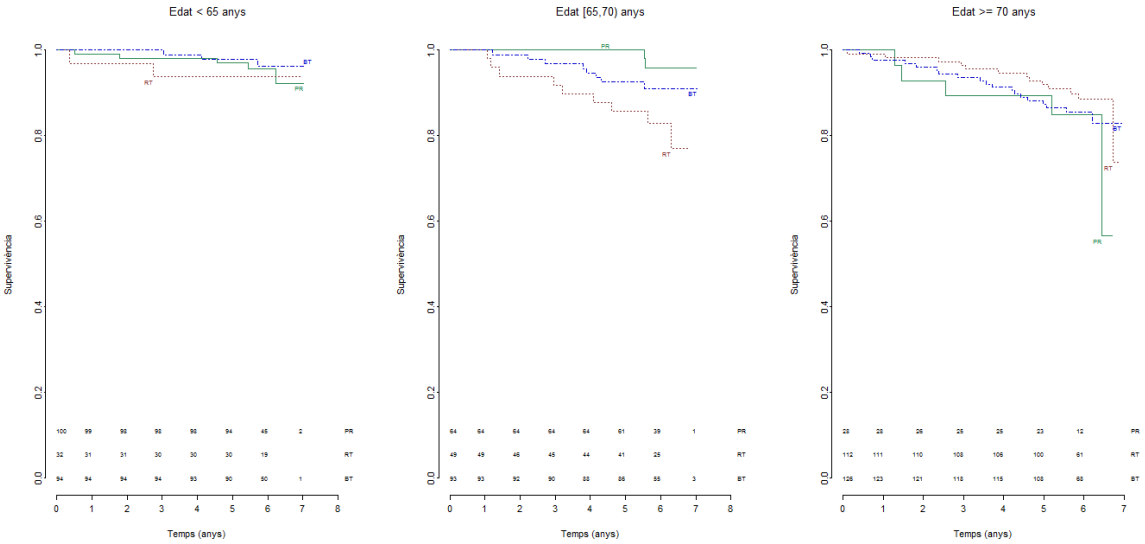

Figura 7.1 Corba de supervivència segons grup d'edat i grup de tractament

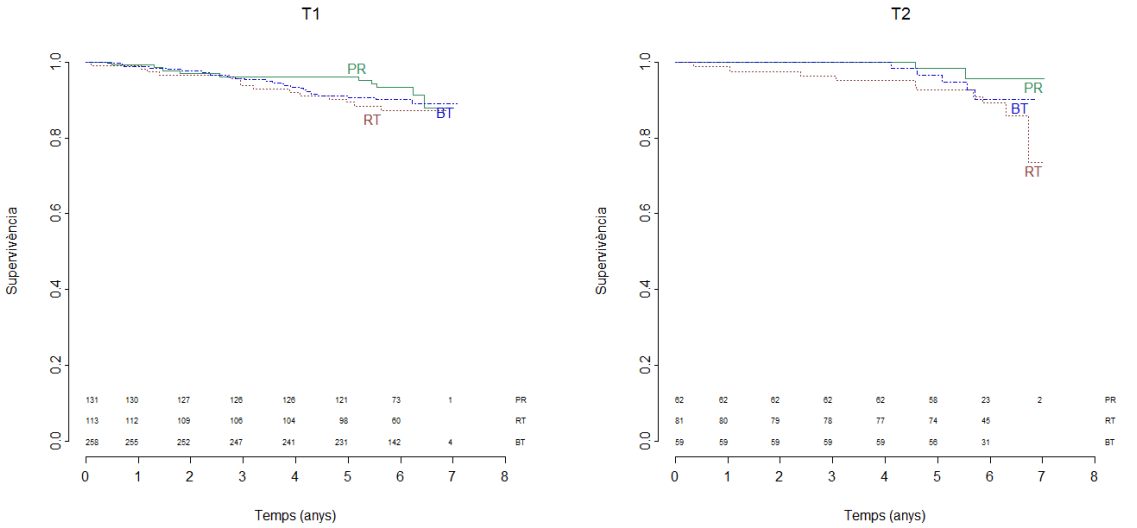

Figura 7.2 Corba de supervivència segons *clinical T stage* i grup de tractament

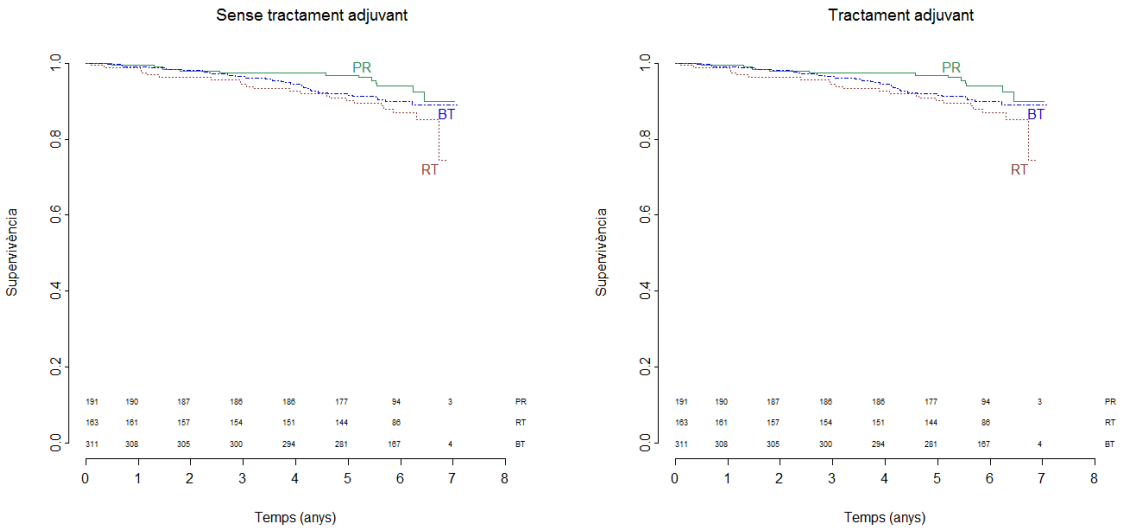

Figura 7.3 Corba de supervivència segons tractament hormonal actual i grup de tractament

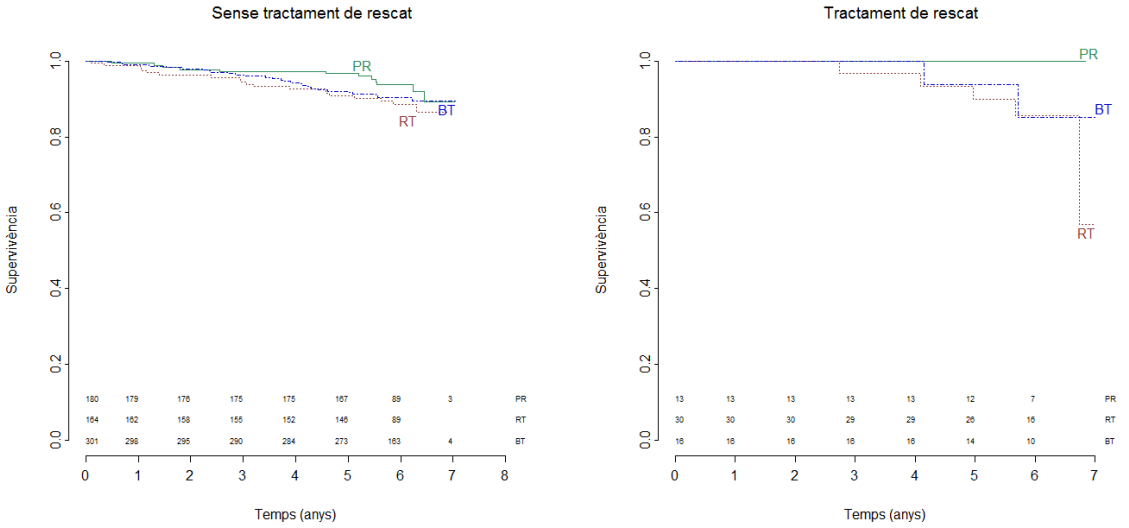

Figura 7.4 Corba de supervivència segons tractament hormonal de rescat i grup de tractament