

Quelques exercices...

Exercice 1.

1. En quoi consiste le phénomène de troncature aléatoire gauche? Comment le modélise-t-on? Comment exprime-t-on les observations? Donner un exemple concret de jeu de données tronqué à gauche.
2. En quoi consiste le phénomène de troncature aléatoire droite? Comment le modélise-t-on? Comment exprime-t-on les observations? Donner un exemple concret de jeu de données tronqué à gauche.
3. Donner la définition et la modélisation mathématique d'un jeu de données doublement censurés (i.e. comportant des observations censurées à gauche et des observations censurées à droite). Donner un exemple concret de telles données.
4. En quoi consiste le phénomène de censure aléatoire droite? Comment le modélise-t-on? Comment exprime-t-on les observations? Donner un exemple concret de telles données.
5. En quoi consiste le phénomène de censure aléatoire gauche? Comment le modélise-t-on? Comment exprime-t-on les observations? Donner un exemple concret de telles données.
6. En quoi consiste le phénomène de censure aléatoire par intervalle? Comment le modélise-t-on? Comment exprime-t-on les observations? Donner un exemple concret de telles données.
7. En quoi consiste le phénomène de risques concurrents? Comment le modélise-t-on? Comment exprime-t-on les observations? Donner un exemple concret de telles données.

Exercice 2.

1. Au début de l'épidémie de SIDA, une étude est réalisée afin d'étudier le temps entre l'infection par le VIH et l'apparition des symptômes. Parmi les patients suivis dans le centre investigateur, 100 personnes séropositives asymptomatiques ont été recrutés au 1er mars 1986 pour une étude prospective de 10 ans (suivi longitudinal). Pour certains patients, la date de l'infection n'est pas connue: seule la date du premier test de séropositivité positif est connue. Au cours de l'étude, 2 patients décèdent lors d'un accident et 23 sont encore asymptomatiques. Par ailleurs, lorsque l'étude se termine et que les données sont analysées, 9 patients ont été perdus de vue tandis que 19 sont encore asymptomatiques. Décrire le type d'observation que l'on enregistre au cours de l'étude.
2. On réalise une étude sur la durée de survie des personnes hospitalisées pour infarctus du myocarde. La durée de survie est comptée à partir de la date de l'infarctus. Seules les personnes survivant à leur séjour à l'hôpital seront incluses dans l'étude. Certaines sont encore en vie lorsqu'on clôt l'étude et que l'on analyse les données. Décrire le type d'observation que l'on enregistre au cours de l'étude.

3. Le carcinome hépatocellulaire (CHC) est un cancer primitif du foie qui survient généralement sur une maladie hépatique déjà existante (cirrhose, hépatite B ou C chronique). Un médecin s'intéresse à la durée entre cette infection hépatique primaire et le diagnostic de CHC. Pour cela, une étude est organisée: 100 personnes souffrant d'une atteinte hépatique depuis plus de deux ans (mais non atteintes de cancer à leur entrée dans l'étude) sont randomisées. A la fin de l'étude, 30 personnes ne sont pas atteintes par le CHC, 2 personnes ont été perdues de vue, 12 sont décédées d'une maladie hépatique mais sans être atteintes de cancer, les autres personnes ont été diagnostiquées pendant l'étude comme atteintes du CHC. Donner le nom et la définition du(des) mécanisme(s) empêchant ici de réaliser systématiquement l'observation de manière exacte. Expliquer comment modéliser mathématiquement le(s) mécanisme(s) en question. Quelle(s) fonction(s) d'intérêt est-il pertinent de considérer ici et pourquoi?

Exercice 3.

Une étude est réalisée sur l'hypertension artérielle chez les personnes en surpoids. Pour cela, sont recrutées des personnes dont le poids atteint entre 110% et 165% du poids désirable (défini selon des critères médicaux) et souffrant d'hypertension artérielle. L'efficacité de 4 régimes sur l'hypertension est étudiée. Le principe d'action de ces 4 régimes consiste à diminuer le surpoids et le taux de sodium sanguin. Les objectifs de traitement visent une pression systolique inférieure à 140 mm Hg et une pression diastolique inférieure à 90 mm Hg. Pour les personnes atteintes de diabète, les valeurs cibles sont encore plus basses (inférieures respectivement à 130 et à 80 mm Hg). Les patients suivent quotidiennement leur tension. On s'intéresse à la durée entre le début du régime et le fait d'atteindre les valeurs cibles de pressions systolique et diastolique. A la fin de l'étude, certains patients ont été perdus de vue, d'autres n'ont jamais atteint la pression artérielle souhaitée. Par ailleurs, certains patients ont omis plusieurs jours durant de suivre leur tension.

1. Donner le nom et la définition du(des) mécanisme(s) empêchant ici de réaliser systématiquement l'observation de manière exacte. Expliquer comment modéliser mathématiquement le(s) mécanisme(s) en question.
2. Que se passe-t-il si l'on inclut seulement les patients ayant suivi leur régime depuis au moins un mois sans avoir encore atteint la pression artérielle recommandée? Même question si l'on inclut seulement les patients de plus de 50 ans.
3. Comment comparer l'influence des quatre régimes sur la loi de la durée d'intérêt introduite précédemment si l'on inclut à la fois des patients diabétiques et des patients non-diabétiques dans l'étude?

Exercice 4.

On réalise une étude sur 30 personnes consommant régulièrement de la drogue par injection intraveineuse. Les personnes sont suivies pendant 6 ans à compter du 1er février 1990 jusqu'au 1er février 1996 à raison d'une visite planifiée tous les 2 mois. A chaque visite un test diagnostique (par dosage sanguin) est effectué afin de déterminer si le patient est infecté ou non par le VHC (Virus de l'Hépatite C). Dans la table 1 sont présentées les informations obtenues pour 5 patients pris au hasard. On s'intéresse à la variable aléatoire X représentant la durée entre le début de la consommation de drogue et l'infection par le VHC. Comme la période d'incubation

n° de patient	âge lors de la 1 ^{ère} injection	âge lors de l'entrée dans l'étude	date du dernier test négatif	date du premier test positif
1	15 ans	19 ans	–	1er février 1993
2	19 ans	21 ans 6 mois	1er février 1995	1er avril 1995
3	31 ans 6 mois	35 ans	1er février 1996	–
4	18 ans 6 mois	20 ans	1er février 1991	1er février 1995
5	≥ 12 ans	27 ans	1er août 1993	1er octobre 1995

Table 1 – Informations sur la contamination par le VHC pour 5 patients pris au hasard

t'_i	3	5	6	8	12	15	17
r_i	20	19	17	12	9	5	2
m_i	1	1	3	2	3	1	1

Table 2 – Observations recueillies pour 20 individus

du VHC est en moyenne de 2 mois, lorsque seulement 2 mois séparent la date du dernier test négatif de la date du premier test positif, on considère que l'infection a eu lieu à la date du dernier test négatif effectué.

Pour chacune des 5 personnes ci-dessus, dire si l'observation de la durée entre le début de la consommation de drogues et l'infection par le VHC est réalisée de manière exacte ou censurée auquel cas on précisera le type de censure impliquée. Pour chacune des 5 personnes ci-dessus, déterminer la valeur de la durée observée en pratique, notée T , ou la fourchette de valeurs observées selon les cas.

Exercice 5.

Considérons les observations $(t_i, d_i)_{i=1, \dots, 20}$ obtenues sous un schéma de censure aléatoire droite indépendante. Notons $t'_1 < t'_2 < \dots < t'_{n'}$ (avec $n' = 7$) les observations non censurées ordonnées et distinctes parmi (t_1, \dots, t_{20}) et introduisons r_j le nombre de sujets à risque à l'instant t'_j ainsi que m_j le nombre de morts en t'_j . Retrouver les observations $(t_i, d_i)_{i=1, \dots, 20}$ à partir des données du tableau 2.

Exercice 6.

Sur 10 patients atteints de cancer des bronches, on a observé les durées de survie suivantes exprimées en mois: 1 3 4⁺ 5 7⁺ 8 9 10⁺ 11 13⁺. Les données suivies du signe ⁺ correspondent à des patients qui ont été perdus de vue à la date fournie.

1. On s'intéresse à la durée de survie au cancer du poumon modélisée par une variable aléatoire X . Sous quelle(s) condition(s) peut-on estimer la fonction de survie d'une variable aléatoire X lorsqu'on dispose d'un échantillon censuré recueilli à partir de n patients indépendants?
2. Exprimer l'estimateur en question à partir du nombre de morts M_i à l'instant T'_i et du nombre de sujets à risque R_i à l'instant T'_i en notant $T'_1 < T'_2 < \dots < T'_{n'}$ les durées observées non censurées ordonnées et distinctes.
3. Déterminer l'estimateur de Kaplan-Meier de la fonction de survie.

Traitement	Durée de rémission
6-MP	6,6,6,6 ⁺ ,7,9 ⁺ ,10,10 ⁺ ,11 ⁺ ,13,16,17 ⁺ ,19 ⁺ ,20 ⁺ ,22,23,25 ⁺ ,32 ⁺ ,32 ⁺ ,34 ⁺ ,35 ⁺
Placebo	1,1,2,2,3,4,4,5,5,8,8,8,8,11,11,12,12,15,17,22,23

Table 3 – Essai comparant l’effet de la 6-MP au placebo sur la durée de rémission dans le cas de la leucémie

n° de patient	observation
1	décès dû au cancer du pancréas au bout de 6 mois
2	décès dû au cancer du pancréas au bout de 6 mois
3	décès accidentel au bout de 5 mois
4	décès dû au cancer du pancréas au bout de 7 mois
5	décès dû au cancer du pancréas au bout de 1 mois
6	décès dû au cancer du pancréas au bout de 8 mois
7	décès dû au cancer du pancréas au bout de 8 mois
8	décès dû au cancer du pancréas au bout de 8 mois
9	rien ne se produit pendant les 18 mois de l’étude
10	décès dû au cancer du pancréas au bout de 3 mois
11	rien ne se produit pendant les 18 mois de l’étude
12	décès dû au cancer du pancréas au bout de 9 mois
13	perdu de vue après 7.5 mois de suivi
14	décès dû au cancer du pancréas au bout de 4.5 mois
15	décès dû au cancer du pancréas au bout de 2 mois

Table 4 – Durées de survie au cancer du pancréas

Exercice 7.

Les données fournies dans le tableau 3 ont été décrites par Freireich (1963). Elles ont été obtenues lors d’un essai thérapeutique ayant pour but de comparer les durées de rémission (exprimées en semaines) de sujets atteints de leucémie selon qu’ils ont reçu de la 6-mercaptopurine (notée 6-MP) ou un placebo. Les données suivies du signe ⁺ correspondent à des patients qui ont été perdus de vue à la date fournie. L’essai a été réalisé en double aveugle, ce qui signifie que ni le patient, ni le médecin ne sont informés de l’attribution du traitement ou du placebo. Déterminer l’estimateur de Kaplan-Meier de la fonction de survie dans chacun des deux groupes de traitement. Que constatez-vous?

Exercice 8.

On réalise une étude sur 18 mois portant sur 15 patients atteints de cancer du pancréas afin d’évaluer leur survie à cette pathologie.

1. Déterminer la valeur des observations réalisées à partir du tableau 4.
2. Sous quelle(s) condition(s) peut-on estimer la fonction de hasard cumulé d’une variable aléatoire X lorsqu’on dispose d’un échantillon censuré recueilli à partir de n patients indépendants?
3. Exprimer l’estimateur en question à partir du nombre de morts M_i à l’instant T'_i et du

nombre de sujets à risque R_i à l'instant T'_i en notant $T'_1 < T'_2 < \dots < T'_n$, les durées observées non censurées ordonnées et distinctes.

4. Calculer l'estimateur de Nelson-Aalen de la fonction de hasard cumulé de la durée de survie au cancer du pancréas.

Exercice 9.

1. On s'intéresse à la durée de survie au cancer du pancréas. On dispose pour cela d'un jeu de données de 225 durées éventuellement aléatoirement censurées à droite de manière indépendante, obtenues dans deux groupes de traitement, notés A et B. Sont présentés respectivement en figures 1 et 2 des intervalles de confiance et des bandes de confiance au niveau de confiance 95% pour la fonction de survie associée à la durée d'intérêt dans les deux groupes de traitement. Comment interpréter chacun des graphiques? En particulier, quelle est la différence entre des intervalles de confiance et des bandes de confiance, tous deux au niveau de confiance 95%, pour la fonction de survie?
2. On réalise une étude sur 22 personnes atteintes de cancer du pancréas. Un traitement de chimiothérapie leur est attribué aléatoirement (en double aveugle) parmi deux choix possibles de molécules notés A et B. Les données sont présentées dans le tableau 5. Sauf mention explicite du contraire, le décès est consécutif au cancer du pancréas. Mettre en oeuvre le test du log-rank afin de comparer l'efficacité des deux traitements.
3. Que changer à l'analyse si l'étude est multicentrique?
4. Ecrire l'équation du modèle de Cox que l'on pourrait ajuster dans le cas d'une étude multicentrique.

Exercice 10.

La leucémie aigüe est un cancer des cellules de la moëlle osseuse. La leucémie est traitée par une chimiothérapie intensive qui nécessite généralement une hospitalisation assez longue. Le but du traitement est de détruire les cellules anormales, mais les cellules saines, notamment les cellules de la moëlle osseuse, du cuir chevelu et du tube digestif, y sont aussi sensibles. Après une cure intensive, le patient ne peut plus renouveler seul les cellules de son sang et de son système immunitaire : on dit que le malade est en phase d'aplasie. Pour les formes ayant le pronostic le plus sévère, le patient reçoit une greffe de cellules souches hématopoïétiques (\approx sanguines) d'un donneur sain compatible. Simultanément, le patient démarre un traitement immunosuppresseur à vie. Le patient passe alors environ un mois en chambre stérile, le temps qu'il sorte de la phase de neutropénie (faible taux sanguin de globules blancs) consécutive à la greffe. Une complication propre à ce type de greffe est la réaction du greffon contre l'hôte (RGH), sorte de maladie de l'immunité qui peut prendre un aspect aigü précoce ou plus retardé et chronique, se manifestant par des signes cutanés, digestifs, hépatiques et bronchiques.

On réalise une étude pendant 6 ans sur le devenir de 20 patients greffés (l'entrée dans l'étude est comptée à la date de la fin de la période de neutropénie) auquel on attribue aléatoirement (en double aveugle) un protocole de soins. L'étude démarre le 1^{er} juillet 1990. La période de recrutement est d'un an.

Les données obtenues sont présentées dans la table 15. Sauf mention explicite du contraire,

n° de patient	observation	traitement
1	décès au bout de 6 mois	A
2	perte de suivi au bout de 8 mois	A
3	décès au bout de 6 mois	A
4	décès au bout de 5 mois	A
5	décès au bout de 2 mois	A
6	décès au bout de 6 mois	A
7	perte de suivi au bout de 6 mois	A
8	décès au bout de 8 mois	A
9	décès au bout de 6 mois	A
10	décès au bout de 5 mois	A
11	décès au bout de 6 mois	A
12	décès au bout de 3 mois	B
13	décès au bout de 4 mois	B
14	décès au bout de 6 mois	B
15	décès au bout de 6 mois	B
16	décès au bout de 3 mois	B
17	décès au bout de 6 mois	B
18	décès au bout de 6 mois	B
19	décès au bout de 3 mois	B
20	décès au bout de 4 mois	B
21	décès accidentel au bout de 8 mois	B
22	décès au bout d'un an	B

Table 5 – Observations réalisées pour les 22 patients atteints de cancer du pancréas

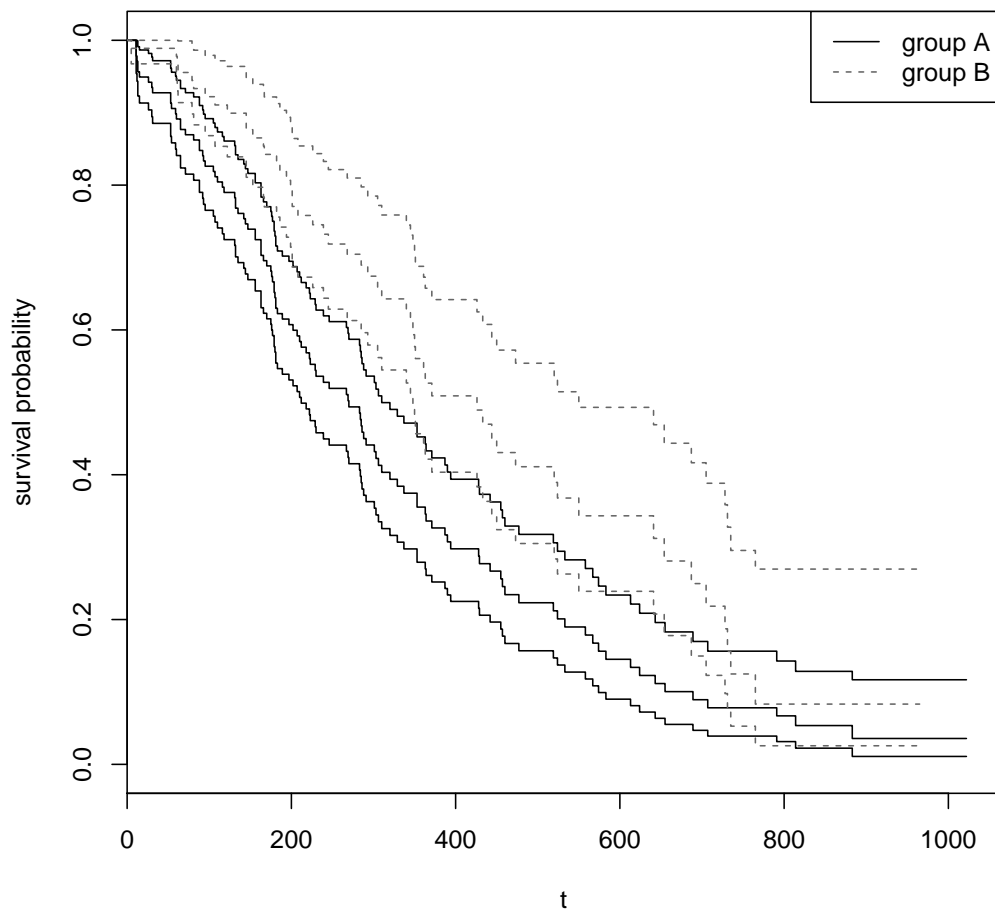


Figure 1 – Tracé des intervalles de confiance à 95% de niveau de confiance

le décès est lié à la maladie ou à son traitement. Pour chacune des questions suivantes, vous introduirez les variables aléatoires nécessaires à la formalisation de l'énoncé et justifierez la modélisation des phénomènes évoqués. Vous justifierez également l'emploi des estimateurs choisis.

1. On s'intéresse à la durée entre l'entrée dans l'étude et la survenue d'un premier épisode de maladie (fatal ou non-fatal). Quelle(s) fonction(s) considérez-vous? Estimer la(les) fonction(s) d'intérêt précédemment introduite(s).
2. On s'intéresse à la durée entre l'entrée dans l'étude et la rechute sans qu'il y ait eu auparavant réaction du greffon contre l'hôte (aigüe ou chronique). Quelle(s) fonction(s) considérez-vous? Estimer la(les) fonction(s) d'intérêt précédemment introduite(s).
3. On s'intéresse à la durée entre l'entrée dans l'étude et la rechute (qu'il y ait eu ou non auparavant réaction du greffon contre l'hôte). Quelle(s) fonction(s) considérez-vous? Estimer la(les) fonction(s) d'intérêt précédemment introduite(s).
4. On s'intéresse à la durée entre l'entrée dans l'étude et le décès, ceci quel que soit la

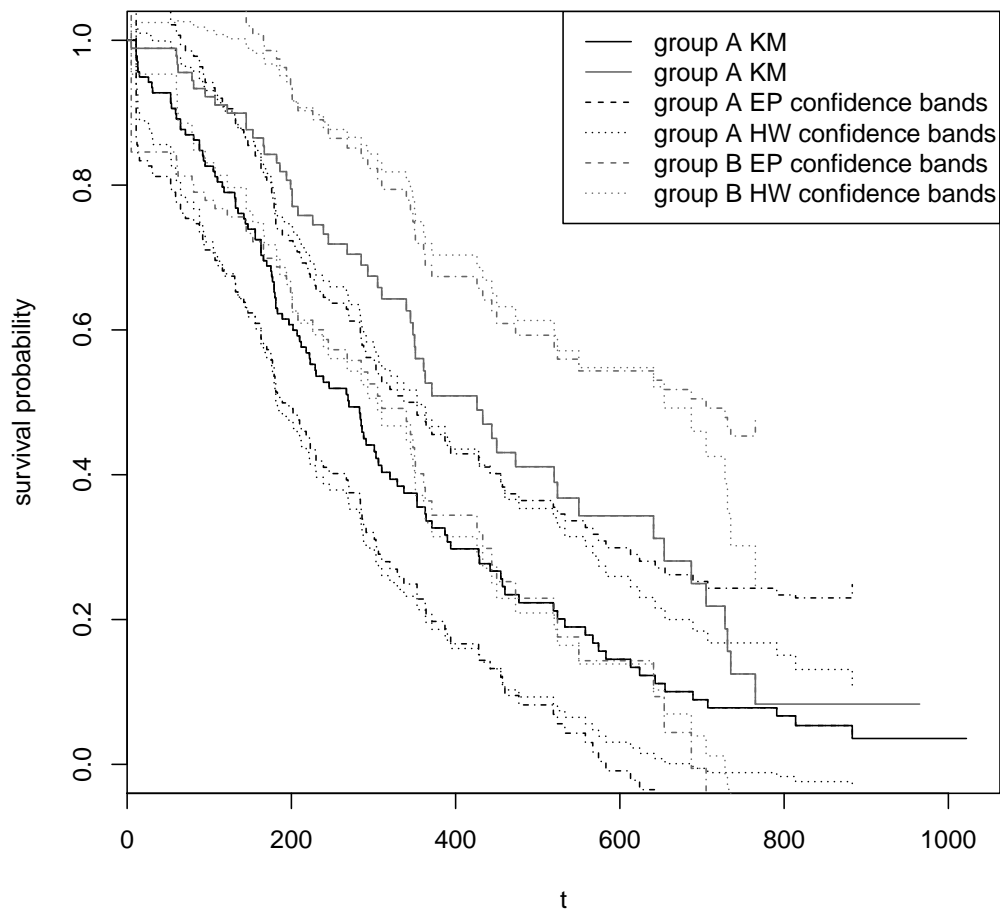


Figure 2 – Tracé des bandes de confiance à 95% de niveau de confiance

trajectoire médicale du patient. Quelle(s) fonction(s) considérez-vous? Estimer la(les) fonction(s) d'intérêt précédemment introduite(s).

- Quelle(s) analyse(s) peut-on proposer pour prendre en compte les covariables présentées dans la table 15?

Exercice 11.

Cet exercice est librement inspiré de Beyersman et al. (2007) A competing risks analysis of bloodstream infection after stem-cell transplantation using subdistribution hazards and cause-specific hazards paru dans *Statistics in Medicine*, 26: 5360-5369.

La leucémie aigue est un cancer des cellules de la moëlle osseuse. La leucémie est traitée par une chimiothérapie intensive qui nécessite généralement une hospitalisation assez longue. Le but est de détruire les cellules anormales, mais les cellules saines, notamment les cellules de la moelle osseuse, du cuir chevelu et du tube digestif, y sont aussi sensibles. Après une cure intensive, le patient, en particulier, ne peut plus renouveler seul les cellules de son sang et de son système immunitaire : on dit que le malade est alors en phase d'aplasie. Pour les formes

n° de patient	observation
1	septicémie au bout de 2 semaines
2	encore en phase de neutropénie (encore en vie et sans septicémie) au bout de 6 semaines
3	décès dû à des complications liées au traitement ou à la maladie au bout de 4 semaines
4	décès dû à des complications liées au traitement ou à la maladie au bout de 3 semaines
5	fin de la période de neutropénie (encore en vie et sans septicémie) au bout de 6 semaines
6	fin de la période de neutropénie (encore en vie et sans septicémie) au bout de 6 semaines
7	décès dû à des complications liées au traitement ou à la maladie au bout de 4 semaines
8	encore en phase de neutropénie (encore en vie et sans septicémie) au bout de 5 semaines
9	septicémie au bout de 14 jours
10	décès dû à des complications liées au traitement ou à la maladie au bout de 3 semaines
11	septicémie au bout de 5 semaines
12	septicémie au bout de 5 semaines
13	fin de la période de neutropénie (encore en vie et sans septicémie) au bout de 6 semaines
14	fin de la période de neutropénie (encore en vie et sans septicémie) au bout de 7 semaines
15	septicémie au bout de 5 semaines

Table 6 – 15 plus petites durées observées

ayant le pronostic le plus sévère, une greffe de moëlle osseuse est réalisée. On réalise une étude pendant 3 mois chez 50 patients greffés en phase de neutropénie (l'entrée dans l'étude est comptée à la date de la greffe). Les données recueillies pour les 15 patients présentant les 15 plus petites durées observées sont représentées dans le tableau 6.

1. Introduire les variables aléatoires permettant de modéliser les données recueillies dans le tableau 6.
2. On s'intéresse à la durée sans épisode de maladie décomptée à partir de la date de la greffe. Quelle(s) fonction(s) considérez-vous?
3. A partir des 15 plus petites observations fournies dans le tableau 6, estimer la(les) fonction(s) d'intérêt précédemment introduite(s).
4. On s'intéresse au risque de septicémie pendant la phase de neutropénie. Quelle(s) fonction(s) considérez-vous?
5. A partir des 15 plus petites observations fournies dans le tableau 6, estimer la(les) fonction(s) d'intérêt précédemment introduite(s).
6. On s'intéresse au risque de décès pendant la phase de neutropénie. Quelle(s) fonction(s) considérez-vous?
7. A partir des 15 plus petites observations fournies dans le tableau 6, estimer la(les) fonction(s) d'intérêt précédemment introduite(s).

Exercice 12.

La transplantation rénale est réalisée chez les personnes en insuffisance rénale terminale ne

pouvant survivre sans des séances de dialyse qui sont longues, pénibles et fréquentes. Les rejets peuvent être de deux types possibles: aigu ou chronique auquel cas on parle de maladie du greffon contre l'hôte notée GVHD dans la littérature médicale pour le terme anglais *Graft Versus Host Disease*. Le rejet aigu entraîne, outre une fièvre importante, le dysfonctionnement du greffon puis la destruction précoce de l'organe si aucun traitement (typiquement à base d'immunosuppresseurs) n'est entrepris. Le rejet chronique ou GVHD survient en général tardivement (plusieurs mois à plusieurs années) et il s'agit d'une dégénérescence lente de l'organe. Dans tous les cas, ces rejets mettent en péril la fonctionnalité du greffon. Un schéma explicatif simplifié se trouve sur la figure 3. Ici, deux groupes de patients, notés A et B, expérimentent deux traitements post-opératoires différents au cours d'une étude qui dure 10 ans. Les patients sont inclus pendant les trois premières années de l'étude à l'issue de leur première transplantation. Les observations sont présentées dans les tableaux 7 et 8. Toutes les durées sont décomptées à partir de la première opération.

1. Soient S_A et S_B la fonction de survie dans chacun des deux groupes A et B, la durée de survie étant décomptée à partir de la première opération. Déterminer l'estimateur de S_A et S_B . Donner l'allure de chacune des deux courbes.
2. Soit F_A la fonction de répartition du décès dans le groupe A et soit $t_0 > 0$ un instant quelconque. Donner la construction d'un intervalle de confiance pour $F_A(t_0)$ au niveau de confiance $1 - \alpha = 95\%$. Exprimer le résultat en fonction des différents nombres de sujets à risque et des différents nombres de décès dans le groupe A.
3. On suppose ici que λ_A et λ_B , les fonctions de hasard instantané associées au décès respectivement dans les groupes A et B, sont proportionnelles. On souhaite tester l'hypothèse $H_0: \lambda_A = \lambda_B$ contre $H_1: \lambda_A \neq \lambda_B$. Mettre en oeuvre le test du log-rank. On utilisera le niveau de risque de première espèce de $\alpha = 5\%$.
4. Recommencer dans le cas de la démonstration de l'équivalence des traitements.
5. Recommencer dans le cas de la démonstration de la non-infériorité du traitement A.
6. Comment faire pour améliorer la qualité de l'analyse à des fins prédictives lorsqu'on dispose, pour tous les patients, d'informations supplémentaires telles que sexe, âge, poids, score de santé de Karnofsky évalué avant la première transplantation (100% correspond à une santé parfaite, puis, plus le score diminue, plus l'état de santé du patient est dégradé, 0% correspond à l'état terminal)? Proposer un modèle pour le taux de hasard instantané associé au décès. En déduire un modèle pour la fonction de survie.
7. On souhaite refaire une étude avec deux protocoles post-opératoires améliorés. Soient à nouveau λ_A et λ_B les fonctions de hasard instantané associées au décès dans chacun des deux groupes. Afin de maximiser la puissance du test de $H_0: \lambda_A = \lambda_B$ contre $H_1: \lambda_A \neq \lambda_B$. Sous l'hypothèse de proportionnalité des fonctions de hasard λ_A et λ_B , déterminer la taille de l'échantillon requise en fonction de θ_R défini (pour $t > 0$ quelconque) par:

$$\theta_R = \log \left(\frac{\lambda_A(t)}{\lambda_B(t)} \right),$$

en fonction de $p_A =$ proportion de patients randomisés dans le groupe A et en fonction de $\psi =$ probabilité de mourir pendant l'étude. On utilisera le niveau de risque de première espèce de $\alpha = 5\%$ et le niveau de puissance de $1 - \beta = 90\%$. On précisera les hypothèses nécessaires au calcul.

8. Recommencer dans le cas de la démonstration de l'équivalence des traitements.
9. Recommencer dans le cas de la démonstration de la non-infériorité du traitement A.
10. Soient $\Lambda_A^{(1)}$ et $\Lambda_B^{(1)}$ les fonctions de hasard cumulé brutes associées à un premier rejet aigu dans chacun des deux groupes. Déterminer l'estimateur de $\Lambda_A^{(1)}$ et $\Lambda_B^{(1)}$. Donner l'allure de chacune des deux courbes.
11. Que pourrait-on proposer pour exploiter au mieux ces données?

n° de patient	observations
1	GVHD du 6 ^{ème} au 10 ^{ème} mois, 1 ^{er} rejet aigu au boût de 10 mois, décès au boût de 12 mois
2	1 ^{er} rejet aigu au boût de 4 mois, décès au boût de 8 mois
3	décès au bout de 10 ans
4	décès accidentel au boût de 5 ans
5	GVHD du 24 ^{ème} au 30 ^{ème} mois, décès au boût de 8 ans
6	1 ^{er} rejet aigu au boût de 6 mois, décès au boût de 2 ans
7	1 ^{er} rejet aigu au boût de 1 mois, nouvelle greffe au boût de 6 mois, encore vivant à la fin de l'étude, ce qui correspond à 9 ans
8	1 ^{er} rejet aigu au boût de 4 mois, décès au boût de 2 ans
9	1 ^{er} rejet aigu au boût de 3 mois, nouvelle greffe au boût de 1 an, décès au boût de 7 ans
10	encore vivant à la fin de l'étude, ce qui correspond à 8 ans
11	décès au boût de 4 ans
12	décès au boût de 5 ans
13	décès au boût de 8 mois
14	décès au boût de 9 ans
15	changement de centre de soins au bout de 2 ans
16	1 ^{er} rejet aigu au bout de 10 mois, décès au boût de 4 ans
17	encore vivant à la fin de l'étude, ce qui correspond à 9 ans
18	encore vivant à la fin de l'étude, ce qui correspond à 7 ans
19	GVHD du 66 ^{ème} au 70 ^{ème} mois, décès au boût de 8 ans
20	GVHD du 44 ^{ème} au 50 ^{ème} mois puis du 76 ^{ème} au 82 ^{ème} mois, décès au boût de 9 ans

Table 7 – Observations réalisées pour les patients du groupe A

Exercice 13.

On réalise une étude sur le tabagisme. Trente patients fumeurs depuis 20 ans sont inclus dans l'étude. Tous les patients rentrent dans l'étude au même moment. Les données recueillies sur 10 ans figurent dans le tableau 13. Proposer une analyse de ces données.

Exercice 14.

1. Ecrire l'équation du modèle de Cox en incluant 4 covariables dont 2 sont quantitatives Z_1 et Z_2 et dont 2 sont qualitatives Z_3 et Z_4 à respectivement 2 et 3 modalités, en supposant qu'il existe une interaction entre Z_1 et Z_3 . Recommencer en stratifiant par rapport à une variable Z_5 qualitative à 4 modalités.

n° de patient	observations
1	GVHD du 30 ^{ème} au 36 ^{ème} mois, décès au boût de 4 ans
2	déménagement au bôt de 5 ans
3	GVHD du 6 ^{ème} au 8 ^{ème} mois, décès au boût de 5 ans
4	encore vivant à la fin de l'étude, ce qui correspond à 8 ans
5	encore vivant à la fin de l'étude, ce qui correspond à 9 ans
6	1 ^{er} rejet aïgu au boût de 6 mois, décès au boût de 2 ans
7	1 ^{er} rejet aïgu au boût de 4 mois, décès au boût de 3 ans
8	GVHD du 66 ^{ème} au 100 ^{ème} mois, décès au boût de 9 ans
9	GVHD du 10 ^{ème} au 14 ^{ème} mois, 1 ^{er} rejet aïgu au boût de 14 mois, décès au boût de 2 ans
10	GVHD du 36 ^{ème} au 50 ^{ème} mois, décès au boût de 6 ans
11	décès au boût de 3 ans
12	perdu de vu au bôt de 12 mois
13	décès au boût de 6 ans
14	décès au boût de 9 ans
15	1 ^{er} rejet aïgu au boût de 3 mois, nouvelle greffe au boût de 6 mois, GVHD du 64 ^{ème} au 68 ^{ème} mois, décès au boût de 8 ans
16	1 ^{er} rejet aïgu au boût de 1 mois, décès au boût de 2 ans
17	GVHD du 66 ^{ème} au 70 ^{ème} mois, perdu de vue à 6 ans
18	1 ^{er} rejet aïgu au boût de 10 mois, décès au boût de 4 ans
19	1 ^{er} rejet aïgu au boût de 14 mois, décès au boût de 4 ans
20	1 ^{er} rejet aïgu au boût de 14 mois, décès au boût de 5 ans

Table 8 – Observations réalisées pour les patients du groupe B

2. Ecrire les hypothèses nulle et alternative lors du test de l'impact de Z_1 sur le taux de hasard.
3. Un lymphome est un cancer du système lymphatique. Il existe différentes formes de la maladie, notamment on distingue les lymphomes hodgkiniens et les lymphomes non-hodgkiniens. Un des traitements possibles consiste en une chimiothérapie suivie d'une greffe de moëlle osseuse qui peut être allogénique ou autologue. Afin d'étudier comment la survie est influencée par le type de greffe, on réalise une étude incluant 10 patients sur lesquels on relève également le temps écoulé entre le diagnostique et la transplantation (noté temps t-g) ainsi que l'indice de Karnofsky (noté entre 0 et 100 par le médecin où 100 = santé parfaite et 0 = mort) qui traduit l'état de santé prétransplantatoire. On obtient les données suivantes présentées dans le tableau 9. Déterminer un estimateur de la survie au lymphome.
4. En supposant l'absence d'interaction entre les différentes variables explicatives, proposer un modèle de Cox adapté à notre étude.

Exercice 15.

1. Ecrire l'équation du modèle de Cox en incluant 4 covariables dont 2 sont quantitatives Z_1 et Z_2 et dont 2 sont qualitatives Z_3 et Z_4 à respectivement 2 et 4 modalités, en

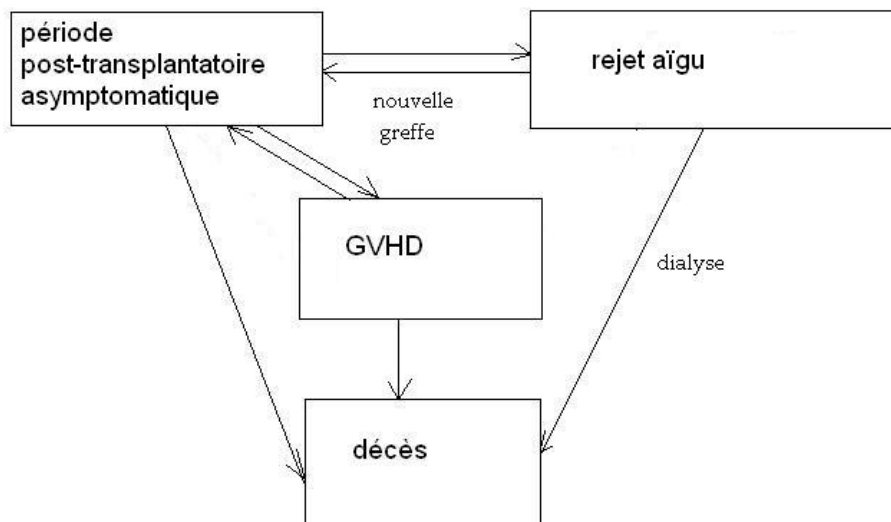


Figure 3 – Schéma simplifié des suites post-transplantatoires

supposant qu'il existe une interaction entre Z_1 et Z_2 d'une part et Z_1 et Z_3 d'autre part. Recommencer en stratifiant par rapport à une variable Z_5 qualitative à 3 modalités.

2. Ecrire les hypothèses nulle et alternative lors du test de l'impact de Z_4 sur le taux de hasard.
3. Un mélanome est un cancer de la peau. Le traitement comprend notamment l'ablation de la tumeur. Afin d'étudier comment la survie est influencée par l'âge, l'épaisseur de la tumeur et la présence d'ulcération, on réalise une étude incluant 10 patients sur lesquels on relève les données présentées dans le tableau 10 (statut: '0' code la présence de censure et '1' code l'absence de censure). Déterminer un estimateur de la survie au mélanome.
4. En supposant l'absence d'interaction entre les différentes variables explicatives, proposer un modèle de Cox adapté à notre étude.

Exercice 16.

On réalise une étude sur 10 ans portant sur 30 patients séropositifs i.e. infectés par le VIH (Virus de l'Immunodéficience Humaine) et co-infectés par la tuberculose. L'infection tuberculeuse est ici toujours postérieure à l'infection par le VIH. On souhaite étudier la durée entre l'infection par le VIH et le décès, durée qui est représentée par une variable aléatoire notée X . Les informations obtenues pour les 30 patients sont présentées en table 12. On note C la variable aléatoire représentant la durée au bout de laquelle la censure se produit. On suppose dans toute la suite que C est indépendante de X . On note \mathcal{C} la variable aléatoire à valeurs dans $\{1, \dots, \mathcal{J}\}$ représentant la cause du décès. On pose $T = \min(X, C)$ et $J = \mathcal{C} I(X \leq C)$.

1. Quel nom porte ce modèle?

n° patient	durée de survie	présence de censure	type de lymphome	type de greffe	temps t-g	score
1	10	0	hodgkinien	autologue	3	50
2	9	1	non-hodgkinien	allogénique	4	40
3	3	1	hodgkinien	allogénique	1	30
4	12	0	non-hodgkinien	allogénique	2	50
5	6	1	non-hodgkinien	autologue	1	40
6	10	1	hodgkinien	autologue	1	45
7	15	1	non-hodgkinien	allogénique	2	60
8	24	1	hodgkinien	allogénique	4	85
9	4	1	hodgkinien	autologue	4	30
10	18	0	hodgkinien	autologue	3	80

Table 9 – Données recueillies dans l'étude sur 10 patients atteints d'un lymphome

n° patient	durée de survie	statut	sexe	âge	épaisseur de la tumeur	présence d'ulcérations
1	10	0	M	44	3.1	0
2	9	1	M	50	4.1	0
3	3	1	F	39	5.1	1
4	12	0	F	51	2.1	1
5	6	1	F	46	2.5	1
6	10	1	M	60	1.9	0
7	15	1	M	48	2.0	0
8	24	1	M	52	1.2	0
9	4	1	F	56	4.1	1
10	18	0	M	35	1.8	0

Table 10 – Données recueillies dans l'étude sur 10 patients atteints d'un melanome

2. Que vaut \mathcal{J} ici? Que représentent T et J ?
3. Calculer les réalisations de T , J et $D = I(X \leq C)$ pour chacune des 30 personnes ci-dessus.
4. On pose pour $t \geq 0$ et pour j dans $\{1, \dots, \mathcal{J}\}$

$$\begin{aligned}
 F(t) &= \mathbb{P}[X \leq t] \text{ avec } F^-(t) = \mathbb{P}[X < t], \\
 \Lambda(t) &= \int_0^t \frac{dF}{1 - F^-}, \\
 F^{(j)}(t) &= \mathbb{P}[X \leq t, \mathcal{C} = j], \\
 \Lambda^{(j)}(t) &= \int_0^t \frac{dF^{(j)}}{1 - F^-}.
 \end{aligned}$$

Que représentent $F(t)$, $\Lambda(t)$, $F^{(j)}(t)$ et $\Lambda^{(j)}(t)$?

5. On veut estimer Λ et $\Lambda^{(j)}$ pour j dans $\{1, \dots, \mathcal{J}\}$. On utilise pour cela les estimateurs rappelés ci-dessous pour $t \geq 0$ et pour j dans $\{1, \dots, \mathcal{J}\}$

$$\begin{aligned}
 \hat{\Lambda}_n(t) &= \sum_{s \leq t} \frac{\Delta H_n^{(1)}(s)}{1 - H_n(s^-)} \\
 \hat{\Lambda}_n^{(j)}(t) &= \sum_{s \leq t} \frac{\Delta H_n^{(1,j)}(s)}{1 - H_n(s^-)}
 \end{aligned}$$

avec

$$\begin{aligned}
 H_n(t) &= \frac{1}{n} \sum_{i=1}^n I(T_i \leq t) \text{ et donc } H_n(t^-) = \frac{1}{n} \sum_{i=1}^n I(T_i < t), \\
 H_n^{(1)}(t) &= \frac{1}{n} \sum_{i=1}^n I(T_i \leq t, D_i = 1), \\
 H_n^{(1,j)}(t) &= \frac{1}{n} \sum_{i=1}^n I(T_i \leq t, J_i = j).
 \end{aligned}$$

Que représentent $H_n(t)$, $H_n^{(1)}(t)$ et $H_n^{(1,j)}(t)$?

6. Pour $t \geq 0$ et pour j dans $\{1, \dots, \mathcal{J}\}$, exprimer $1 - H_n(t^-)$, $\Delta H_n^{(1)}(t)$ et $\Delta H_n^{(1,j)}(t)$ puis $\hat{\Lambda}_n(t)$ et $\hat{\Lambda}_n^{(j)}(t)$ en fonction des quantités suivantes:
 - $T'_1 < T'_2 < \dots < T'_{n'}$: observations non censurées distinctes et ordonnées
 - M_i : nombre de morts en T'_i
 - R_i : nombre de sujets à risque en T'_i
 - $T'^{(j)}_1 < T'^{(j)}_2 < \dots < T'^{(j)}_{n'_j}$: observations non censurées de cause j distinctes et ordonnées
 - $M_i^{(j)}$: nombre de morts en $T'^{(j)}_i$ dues à la cause j
 - $R_i^{(j)}$: nombre de sujets à risque en $T'^{(j)}_i$
7. A partir des 30 observations fournies, calculer les quantités T'_i , M_i , R_i et $T'^{(j)}_i$, $M_i^{(j)}$ et $R_i^{(j)}$ pour $j = 1, \dots, \mathcal{J}$

8. Pour $t \leq 4$ ans, calculer $\hat{\Lambda}_n(t)$ et $\hat{\Lambda}_n^{(j)}(t)$ pour $j = 1, \dots, \mathcal{J}$.
9. Supposons que l'on dispose des données suivantes pour chacun des patients:
- $Z_1 =$ âge à la date de l'infection par le VIH
 - $Z_2 =$ âge à la date de l'infection par la tuberculose
 - $Z_3 =$ assiduité au traitement (score entre 0 et 10 évalué après réponse à un questionnaire médical à la fin de l'étude)
 - $Z_4 = \begin{cases} 1 & \text{en cas de localisation pulmonaire de la tuberculose} \\ 2 & \text{en cas de localisation extra-pulmonaire de la tuberculose} \\ 3 & \text{en cas de localisation mixte de la tuberculose} \end{cases}$
 - $Z_5 = \begin{cases} 1 & \text{en cas de SIDA déclaré (séropositivité avec symptômes)} \\ 0 & \text{en cas de séropositivité sans symptômes (phase de latence)} \end{cases}$
 - $Z_6 = \begin{cases} 1 & \text{hémophilie} \\ 2 & \text{consommation de drogue par voie intraveineuse} \\ 3 & \text{pratiques sexuelles à risque} \\ 4 & \text{consommation de drogue par voie intraveineuse et de pratiques sexuelles à risque} \\ 5 & \text{en l'absence des facteurs de risque précédents} \end{cases}$

Ecrire l'équation du modèle de Cox en incluant les covariables Z_1, Z_2, Z_3, Z_4 et Z_5 , en admettant qu'il existe une interaction uniquement entre Z_3 et Z_5 . Recommencer en stratifiant par rapport à Z_6 .

10. On suppose qu'un deuxième groupe de 35 patients a été suivi selon le même protocole. Ces patients reçoivent un traitement noté B différent du traitement noté A reçu par le premier groupe de patients. On veut comparer la survie dans ces deux groupes. Quel test utiliseriez-vous?

Exercice 17.

L'exercice suivant est inspiré des travaux de Fyles et al. sur le cancer du sein, publiés en 2004 dans le *New England Journal of Medicine*. Deux traitements (notés ici A et B) sont comparés lors d'un essai clinique multicentrique randomisé incluant des patientes de plus de 50 ans diagnostiquées avec un cancer primitif du sein localisé (tumeur inférieure à 5cm, sans envahissement ganglionnaire, sans métastase à distance). L'étude débute le 1er décembre 1992 et est close le 31 mars 2003, les patientes ayant été recrutées entre le 1er décembre 1992 et le 30 juin 2000. Lors du diagnostic, sont relevées un certain nombre de caractéristiques des patientes à savoir:

- âge au moment du diagnostic
- taille de la tumeur au moment du diagnostic
- pratique d'exercice physique
- consommation d'alcool

- prise de traitements hormonaux
- centre de soins (C1: Princess Margaret Hospital, C2: Women’s College Campus of the Sunnybrook in Toronto, C3: Women’s College Health Sciences of the Sunnybrook in Toronto, C4: British Columbia Center Agency in Vancouver, C5: British Columbia Center Agency in Victoria)

Les évènements observés pendant l’étude liés au cancer sont les suivants: rechute locale, atteinte des ganglions axillaires (au niveau de l’aisselle), rechute distante (métastases pulmonaires ou lymphatiques par exemple). Sauf mention explicite du contraire, les décès observés sont liés au cancer.

Les données obtenues pour le groupe A sont présentées dans la table 11. Pour chacune des questions suivantes, vous introduirez les variables aléatoires nécessaires à la formalisation de l’énoncé et justifierez la modélisation des phénomènes évoqués. Vous justifierez également l’emploi des estimateurs choisis.

1. Dans un premier temps, on s’intéresse à la durée sans aucun épisode lié à la maladie (ni rechute locale, ni atteinte des ganglions axillaires, ni rechute à distance, ni décès) depuis l’entrée dans l’étude. Estimer la fonction de hasard cumulé et la fonction de survie associées à cette durée à partir des observations fournies dans le tableau 11.
2. On s’intéresse ensuite à la durée de survie des patientes comptée depuis l’entrée dans l’étude. Estimer la fonction de survie à partir des observations fournies dans le tableau 11.
3. On s’intéresse maintenant à la durée entre l’entrée dans l’étude et la survenue d’une première rechute locale. Quelle(s) fonction(s) est-il pertinent de considérer pour cela? Estimer la (les) à partir des observations fournies dans le tableau 11.
4. On souhaite maintenant évaluer l’impact des différents facteurs de risque potentiels relevés au moment du diagnostic sur l’évolution du pronostic vital au cours du temps. Proposer un modèle adapté à ce travail.
5. Commenter l’extrait suivant de l’article de Fyles et al. (2004):

A noninferiority design was used to test the hypothesis that for women 50 years age or older who had [localized breast cancer, treatment A is equivalent to treatment B]. The treatment groups were considered equivalent if the upper limit of the 95 percent confidence interval of the difference of disease-free survival at five years was no more than 7 percent. We estimated that a sample of 800 women was required with a one-sided alpha of 0.05 and a statistical power of 90 percent or greater, assuming a 5 percent loss-to-follow-up.

Exercice 18.

Des figures 4 à 7, est présentée une brève analyse d’un jeu de données réalisée avec le code ci-dessous au moyen du logiciel R. Ce jeu de données contient les durées de survie de 250 patients atteints d’un cancer du poumon. Certaines des durées sont aléatoirement censurées à droite par la fin de l’étude ou par un décès sans lien avec le cancer du poumon. Sont également recueillies les covariables suivantes: l’âge au diagnostic (`age`) exprimé en années, un score de santé (noté

sur 100 par le médecin après examen du patient), le stade du cancer au diagnostic (**stade**) à évaluer par le médecin (stade I, II, III, ou IV, avec stade I = peu avancé, stade IV = très avancé) et la consommation de tabac pendant les 10 dernières années (**tabac**: 1 = jamais, 2 = occasionnellement, 3 = régulièrement, 4= intensivement).

1. Ecrire l'équation du modèle de Cox ajustés ici. Interpréter les paramètres introduits.
2. Quelles sont les hypothèses sous-jacentes posées par le modèle de Cox?
3. Quelles techniques de vérification de ces hypothèses connaissez-vous?
4. Que pensez-vous de la qualité de l'ajustement du modèle présenté aux données?
5. Quelles conclusions pouvez-vous tirer des résultats présentés?

```
> mycox<- coxph(mysurv~age+score+stade+tabac)
> summary(mycox)
Call:
coxph(formula = mysurv ~ age + score + stade + tabac)

n= 250, number of events= 205

              coef exp(coef)  se(coef)      z Pr(>|z|)
age          1.159e+00 3.188e+00 8.915e-02 13.006 < 2e-16 ***
score       -5.714e+00 3.298e-03 4.396e-01 -12.998 < 2e-16 ***
stade2       1.889e+00 6.616e+00 4.820e-01  3.920 8.85e-05 ***
stade3       2.936e+00 1.884e+01 4.736e-01  6.199 5.67e-10 ***
stade4       8.721e+00 6.130e+03 8.146e-01 10.706 < 2e-16 ***
tabac2       2.721e+00 1.520e+01 3.395e-01  8.016 1.11e-15 ***
tabac3       3.791e+00 4.428e+01 4.178e-01  9.074 < 2e-16 ***
tabac4       1.031e+01 3.015e+04 8.871e-01 11.626 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
age          3.188e+00  3.137e-01 2.677e+00 3.797e+00
score        3.298e-03  3.032e+02 1.393e-03 7.807e-03
stade2       6.616e+00  1.512e-01 2.572e+00 1.702e+01
stade3       1.884e+01  5.308e-02 7.446e+00 4.766e+01
stade4       6.130e+03  1.631e-04 1.242e+03 3.026e+04
tabac2       1.520e+01  6.578e-02 7.815e+00 2.957e+01
tabac3       4.428e+01  2.258e-02 1.953e+01 1.004e+02
tabac4       3.015e+04  3.317e-05 5.299e+03 1.716e+05

Concordance= 0.994 (se = 0.023 )
Rsquare= 0.997 (max possible= 0.999 )
Likelihood ratio test= 1475 on 8 df, p=0
Wald test              = 170.7 on 8 df, p=0
```

Score (logrank) test = 462.1 on 8 df, p=0

```
> mytest=cox.zph(mycox)
> print(mytest)
      rho  chisq  p
age    -0.04375 0.42362 0.515
score   0.05150 0.59871 0.439
stade2  0.00968 0.01928 0.890
stade3  0.02576 0.12194 0.727
stade4 -0.03757 0.29261 0.589
tabac2  0.00531 0.00552 0.941
tabac3 -0.03248 0.23942 0.625
tabac4 -0.04334 0.47348 0.491
GLOBAL      NA 3.98631 0.858
```

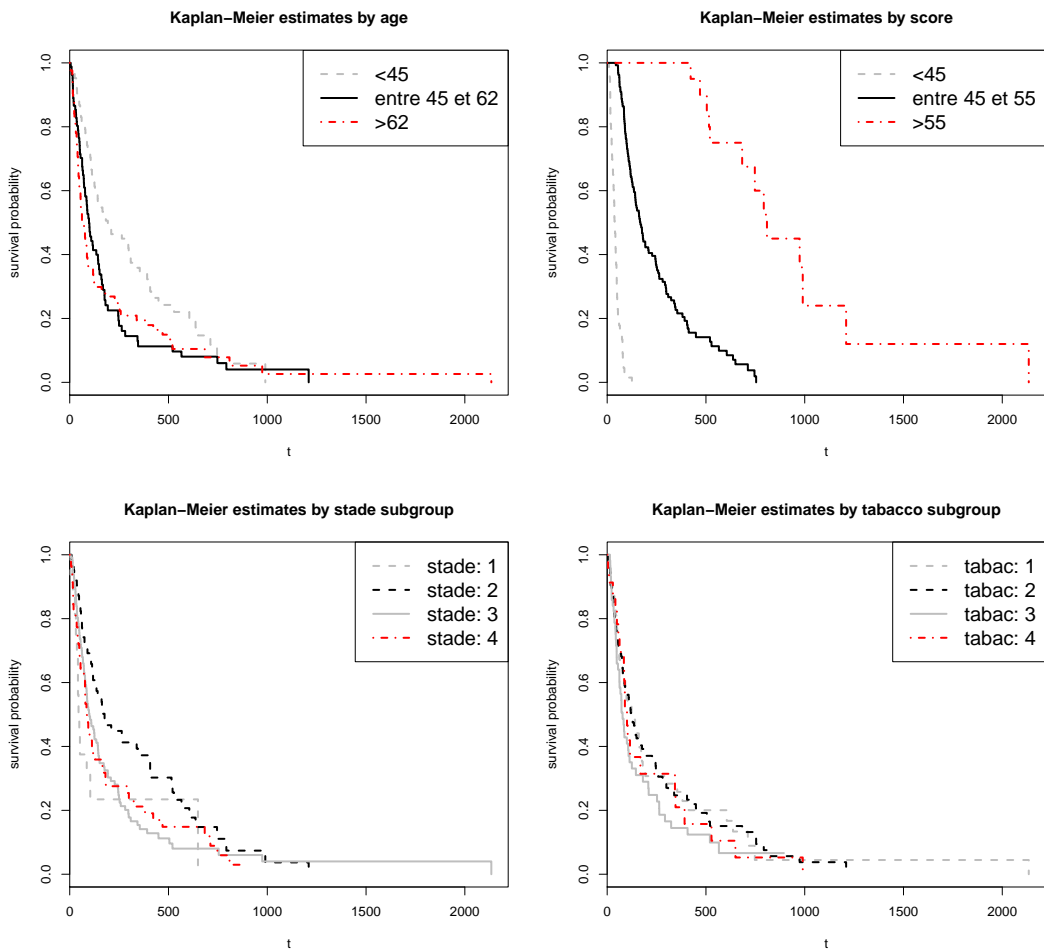


Figure 4 – Tracé des courbes de survie estimées par sous-groupes, cancer du poumon

Exercice 19.

Une étude est réalisée pendant 3 ans sur 40 patients atteints de cancer primitif du poumon

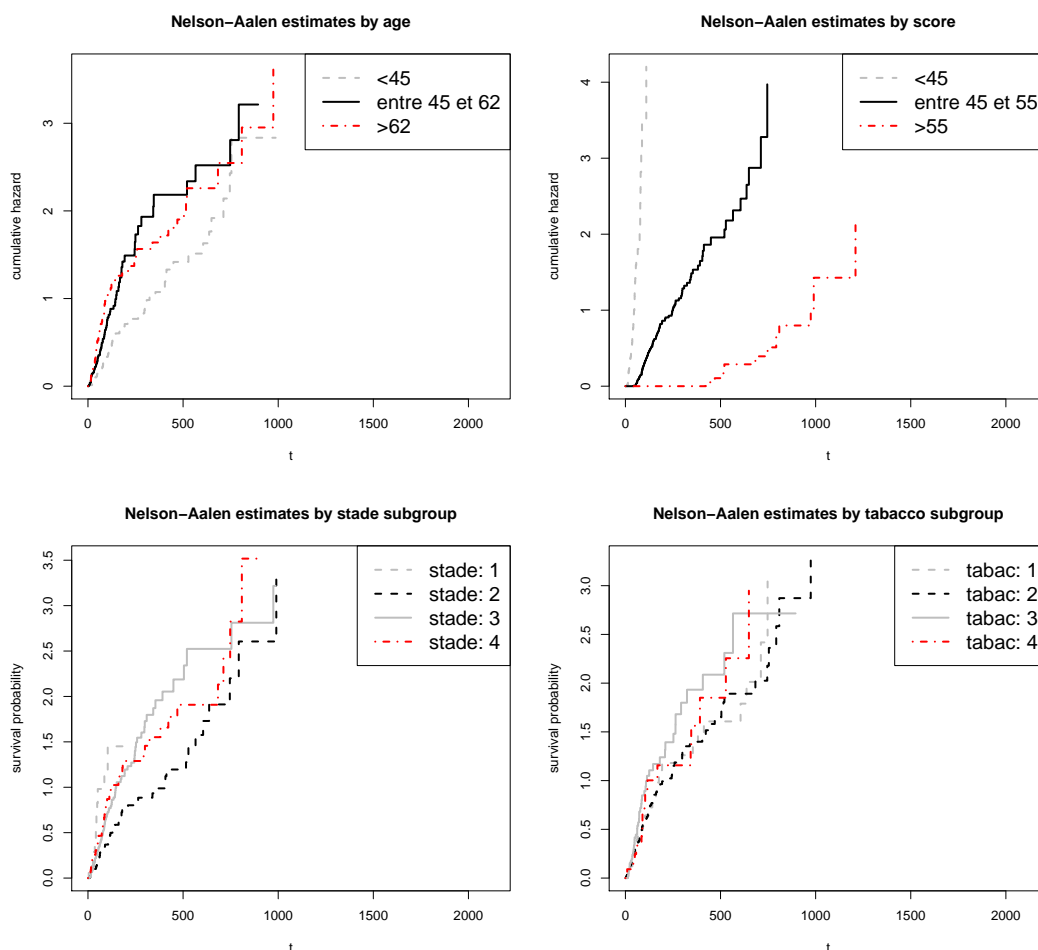


Figure 5 – Tracé des courbes de hasard cumulé estimées par sous-groupes, cancer du poumon

inopérable. L'entrée des patients dans l'étude intervient entre le 1er février 1990 et le 31 janvier 1992, et coïncide avec le moment où le diagnostic de cancer primitif du poumon est posé. A chaque patient, en plus d'un traitement de chimiothérapie attribué systématiquement, est attribué un traitement supplémentaire choisi aléatoirement (en double aveugle) parmi radiothérapie et immunothérapie. Lors du diagnostic, sont relevées un certain nombre de caractéristiques des patients à savoir:

- âge au moment du diagnostic (en années)
- sexe (H= homme, F= femme)
- IMC (= indice de masse corporelle)
- stade du cancer au moment du diagnostic (I = localisé au thorax, II = localisé au thorax avec atteinte ganglionnaire à proximité, III = localement avancé)
- statut tabagique (non fumeur, fumeur occasionnel, fumeur régulier, fumeur excessif)
- type histologique de la tumeur (carcinome épidermoïde, adénocarcinome, à grandes cellules, à petites cellules).

La progression de la maladie est schématiquement représentée sur la figure 8. On dispose des données recueillies sur 40 personnes en tables 18 et 19. Sauf mention explicite du contraire, le décès est consécutif au cancer primitif du poumon.

Vous veillerez à introduire à chaque fois que nécessaire les variables aléatoires permettant de modéliser les phénomènes étudiés et les données correspondantes.

1. On s'intéresse à la durée sans progression de la maladie (ni rechute locale, ni métastases, ni décès liés au cancer) à compter de la date du diagnostic de cancer du poumon dans le groupe "immunothérapie". A quelle(s) fonction(s) vous intéressez-vous? Estimer la(les) fonction(s) précédemment introduite(s).
2. On s'intéresse à la durée jusqu'à une 1^{ère} rechute locale dans le groupe "immunothérapie" à compter de la date du diagnostic de cancer du poumon. A quelle(s) fonction(s) vous intéressez-vous? Estimer la(les) fonction(s) précédemment introduite(s).
3. On s'intéresse à la durée de survie au cancer primitif du poumon dans le groupe "immunothérapie" à compter de la date du diagnostic. A quelle(s) fonction(s) vous intéressez-vous? Estimer la(les) fonction(s) précédemment introduite(s).
4. On s'intéresse à la durée entre le diagnostic de cancer du poumon et la survenue d'un 1^{er} évènement non-fatal lié à l'évolution du cancer primitif du poumon dans le groupe "immunothérapie". A quelle(s) fonction(s) vous intéressez-vous? Estimer la(les) fonction(s) précédemment introduite(s).
5. On s'intéresse à la durée entre le diagnostic de cancer du poumon et la survenue de métastases dans le groupe "immunothérapie". A quelle(s) fonction(s) vous intéressez-vous? Estimer la(les) fonction(s) précédemment introduite(s).
6. Avant le début de l'étude, l'investigateur annonce dans le protocole qu'il compte employer le test du log-rank au niveau de risque de 1^{ère} espèce α pour comparer les deux traitements. Détailler l'hypothèse nulle et l'hypothèse alternative dans le cadre d'un essai en supériorité et dans le cadre d'un essai en non-infériorité.
7. Dans le cadre d'un essai en supériorité au niveau de risque de 1^{ère} espèce α , effectuer le calcul de puissance que doit alors produire l'investigateur dans le protocole.
8. Mettre en oeuvre le test du log-rank au niveau de risque de 1^{ère} espèce 5% afin de comparer l'efficacité des deux traitements dans le cadre d'un essai en supériorité.
9. On souhaite évaluer l'impact des différents facteurs de risque relevés au moment du diagnostic sur l'évolution du pronostic vital au cours du temps. Proposer un modèle adapté à ce travail. Quelles techniques de vérification des hypothèses sous-jacentes à ce modèle connaissez-vous?
10. On souhaite évaluer l'impact des différents facteurs de risque relevés au moment du diagnostic sur l'évolution au cours du temps du risque de 1^{ère} rechute locale. Proposer un modèle adapté à ce travail.

Exercice 20.

Des figures 9 à 16, est présentée une brève analyse d'un jeu de données nommé **MeLanoma** réalisée

avec le code ci-dessous au moyen du logiciel R. Ce jeu de données contient les durées de survie au mélanome (cancer de la peau) de 205 patients diagnostiqués atteints de mélanome et traités chirurgicalement. Certaines des durées sont aléatoirement censurées à droite par la fin de l'étude ou par un décès sans lien avec le mélanome. L'objet R "Melanoma.surv" contient les durées réellement observées ainsi que les indicatrices de censure des 205 patients ie $(T_i, D_i)_{i=1, \dots, 205}$. Sont également recueillies les covariables suivantes: le genre (**sex**), l'âge (**age**) exprimé en années, l'année d'opération (**year**), l'épaisseur de la tumeur ôtée (**thickness**) exprimée en mm et la présence d'ulcération à la surface de la tumeur (**ulcer**: 1 = présence, 0 = absence).

1. Ecrire l'équation des modèles de Cox ajustés ici. Interpréter les paramètres introduits.
2. Quelles sont les hypothèses sous-jacentes posées par le modèle de Cox?
3. Quelles techniques de vérification de ces hypothèses connaissez-vous?
4. Que pensez-vous de la qualité de l'ajustement du modèle présenté aux données? Justifier.
5. Quelles conclusions pouvez-vous tirer des résultats obtenus?

```
> cox.Melanoma.sex=coxph(Melanoma.surv~sex,data=Melanoma)
> summary(cox.Melanoma.sex)
Call:
coxph(formula = Melanoma.surv ~ sex, data = Melanoma)

n= 205, number of events= 57

      coef exp(coef) se(coef)      z Pr(>|z|)
sex 0.6622   1.9390   0.2651  2.498   0.0125 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
sex      1.939      0.5157      1.153      3.26

Concordance= 0.59 (se = 0.033 )
Rsquare= 0.03 (max possible= 0.937 )
Likelihood ratio test= 6.15 on 1 df,  p=0.01314
Wald test              = 6.24 on 1 df,  p=0.01251
Score (logrank) test = 6.47 on 1 df,  p=0.01098
```

```
> cox.Melanoma.age=coxph(Melanoma.surv~age,data=Melanoma)
> summary(cox.Melanoma.age)
Call:
coxph(formula = Melanoma.surv ~ age, data = Melanoma)

n= 205, number of events= 57
```

```

      coef exp(coef) se(coef)      z Pr(>|z|)
age 0.019220  1.019406 0.008769 2.192  0.0284 *

```

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

```

      exp(coef) exp(-coef) lower .95 upper .95
age      1.019      0.981      1.002      1.037

```

```

Concordance= 0.572 (se = 0.04 )
Rsquare= 0.024 (max possible= 0.937 )
Likelihood ratio test= 5 on 1 df,  p=0.02534
Wald test              = 4.8 on 1 df,  p=0.02839
Score (logrank) test = 4.83 on 1 df,  p=0.02796

```

```

> cox.Melanoma.year=coxph(Melanoma.surv~year,data=Melanoma)
> summary(cox.Melanoma.year)
Call:
coxph(formula = Melanoma.surv ~ year, data = Melanoma)

```

n= 205, number of events= 57

```

      coef exp(coef) se(coef)      z Pr(>|z|)
year -0.06867  0.93364 0.05287 -1.299  0.194

```

```

      exp(coef) exp(-coef) lower .95 upper .95
year      0.9336      1.071      0.8417      1.036

```

```

Concordance= 0.573 (se = 0.04 )
Rsquare= 0.008 (max possible= 0.937 )
Likelihood ratio test= 1.65 on 1 df,  p=0.1996
Wald test              = 1.69 on 1 df,  p=0.194
Score (logrank) test = 1.7 on 1 df,  p=0.1929

```

```

> cox.Melanoma.thickness=coxph(Melanoma.surv~thickness,data=Melanoma)
> summary(cox.Melanoma.thickness)
Call:
coxph(formula = Melanoma.surv ~ thickness, data = Melanoma)

```

n= 205, number of events= 57

```

      coef exp(coef) se(coef)      z Pr(>|z|)
thickness 0.16024  1.17380 0.03126 5.126 2.96e-07 ***

```

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

	exp(coef)	exp(-coef)	lower .95	upper .95
thickness	1.174	0.8519	1.104	1.248

Concordance= 0.741 (se = 0.04)
Rsquare= 0.089 (max possible= 0.937)
Likelihood ratio test= 19.19 on 1 df, p=1.186e-05
Wald test = 26.27 on 1 df, p=2.964e-07
Score (logrank) test = 28.7 on 1 df, p=8.432e-08

```
> cox.Melanoma.ulcer=coxph(Melanoma.surv~ulcer,data=Melanoma)  
> summary(cox.Melanoma.ulcer)
```

Call:
coxph(formula = Melanoma.surv ~ ulcer, data = Melanoma)

n= 205, number of events= 57

	coef	exp(coef)	se(coef)	z	Pr(> z)
ulcer	1.4717	4.3567	0.2954	4.982	6.29e-07 ***

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

	exp(coef)	exp(-coef)	lower .95	upper .95
ulcer	4.357	0.2295	2.442	7.773

Concordance= 0.689 (se = 0.034)
Rsquare= 0.13 (max possible= 0.937)
Likelihood ratio test= 28.44 on 1 df, p=9.68e-08
Wald test = 24.82 on 1 df, p=6.29e-07
Score (logrank) test = 29.56 on 1 df, p=5.413e-08

```
> cox.Melanoma.cov=coxph(Melanoma.surv~sex+age+year+thickness+ulcer,data=Melanoma)  
> summary(cox.Melanoma.cov)
```

Call:
coxph(formula = Melanoma.surv ~ sex + age + year + thickness +
ulcer, data = Melanoma)

n= 205, number of events= 57

	coef	exp(coef)	se(coef)	z	Pr(> z)
sex	0.448121	1.565368	0.266861	1.679	0.093107 .
age	0.016805	1.016947	0.008578	1.959	0.050094 .
year	-0.102566	0.902518	0.061007	-1.681	0.092719 .


```

thickness 0.100312 1.105516 0.038212 2.625 0.008660 **
ulcer      1.194555 3.302087 0.309254 3.863 0.000112 ***

```

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

	exp(coef)	exp(-coef)	lower .95	upper .95
sex	1.5654	0.6388	0.9278	2.641
age	1.0169	0.9833	1.0000	1.034
year	0.9025	1.1080	0.8008	1.017
thickness	1.1055	0.9046	1.0257	1.191
ulcer	3.3021	0.3028	1.8012	6.054

Concordance= 0.757 (se = 0.04)

Rsquare= 0.195 (max possible= 0.937)

Likelihood ratio test= 44.4 on 5 df, p=1.922e-08

Wald test = 40.89 on 5 df, p=9.88e-08

Score (logrank) test = 48.14 on 5 df, p=3.328e-09

```
> cox.Melanoma.cov=coxph(Melanoma.surv~sex+age+year+ulcer+age:year,data=Melanoma)
```

```
> summary(cox.Melanoma.cov)
```

Call:

```
coxph(formula = Melanoma.surv ~ sex + age + year + ulcer + age:year,
      data = Melanoma)
```

n= 205, number of events= 57

	coef	exp(coef)	se(coef)	z	Pr(> z)
sex	5.317e-01	1.702e+00	2.684e-01	1.981	0.0476 *
age	1.258e+01	2.916e+05	6.847e+00	1.838	0.0661 .
year	2.108e-01	1.235e+00	1.892e-01	1.114	0.2653
ulcer	1.438e+00	4.213e+00	2.995e-01	4.801	1.58e-06 ***
age:year	-6.383e-03	9.936e-01	3.478e-03	-1.835	0.0665 .

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

	exp(coef)	exp(-coef)	lower .95	upper .95
sex	1.702e+00	5.876e-01	1.0057	2.880e+00
age	2.916e+05	3.429e-06	0.4330	1.964e+11
year	1.235e+00	8.100e-01	0.8521	1.789e+00
ulcer	4.213e+00	2.374e-01	2.3422	7.578e+00
age:year	9.936e-01	1.006e+00	0.9869	1.000e+00

Concordance= 0.745 (se = 0.04)

Rsquare= 0.185 (max possible= 0.937)

Likelihood ratio test= 42 on 5 df, p=5.886e-08

Wald test = 32.32 on 5 df, p=5.13e-06

Score (logrank) test = 44.64 on 5 df, p=1.718e-08

```
> cox.Melanoma.cov2=coxph(Melanoma.surv~thickness+ulcer,data=Melanoma)
> summary(cox.Melanoma.cov2)
```

Call:

```
coxph(formula = Melanoma.surv ~ thickness + ulcer, data = Melanoma)
```

n= 205, number of events= 57

	coef	exp(coef)	se(coef)	z	Pr(> z)	
thickness	0.1140	1.1208	0.0361	3.158	0.00159	**
ulcer	1.2180	3.3805	0.3091	3.941	8.12e-05	***

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

	exp(coef)	exp(-coef)	lower .95	upper .95
thickness	1.121	0.8922	1.044	1.203
ulcer	3.380	0.2958	1.845	6.195

Concordance= 0.765 (se = 0.04)

Rsquare= 0.163 (max possible= 0.937)

Likelihood ratio test= 36.44 on 2 df, p=1.224e-08

Wald test = 35.87 on 2 df, p=1.624e-08

Score (logrank) test = 42.23 on 2 df, p=6.753e-10

```
> test.Melanoma=cox.zph(cox.Melanoma.cov)
```

```
> print(test.Melanoma)
```

	rho	chisq	p
sex	-0.1626	1.5606	0.2116
age	0.0325	0.0498	0.8234
year	0.0648	0.1922	0.6611
ulcer	-0.2398	3.2469	0.0716
age:year	-0.0322	0.0490	0.8248
GLOBAL	NA	7.0652	0.2158

```
> test.Melanoma=cox.zph(cox.Melanoma.cov2)
```

```
> print(test.Melanoma)
```

	rho	chisq	p
thickness	-0.231	2.45	0.1172
ulcer	-0.187	1.98	0.1598
GLOBAL	NA	6.26	0.0438

Exercice 21.

Ci-dessous est présentée une brève analyse du jeu de données `veteran` disponible dans le package `survival` du logiciel R. Ce jeu de données a été décrit dans le manuel paru en 1980, écrit par D. Kalbfleisch et R.L. Prentice et intitulé *The Statistical Analysis of Failure Time Data*. Il contient les durées de survie de 137 patients atteints d'un cancer du poumon inopérable recueillies dans le cadre d'un essai randomisé de comparaison de deux traitements. Certaines des durées sont aléatoirement censurées à droite par la fin de l'étude ou par un décès sans lien avec le cancer du poumon. Sont également recueillies les covariables suivantes:

- `trt`: type de traitement reçu codé par 1 (traitement standard) ou 2 (nouvelle chimiothérapie),
- `celltype`: type histologique de la tumeur codé par 1=squamous, 2=smallcell, 3=adeno, 4=large,
- `karno`: indice de Karnofsky. Il s'agit d'un score entre 0 et 100 (100=normal, absence de maladie, 0= mort) évalué par le médecin après examen du patient au moment de la randomisation,
- `diagtime`: nombre de mois écoulé entre le diagnostique de cancer et l'entrée dans l'étude,
- `age`: âge exprimé en années,
- `prior`: code 0 pour les patients ayant déjà reçu un traitement pour leur cancer, code 10 pour les patients n'ayant encore jamais reçu de traitement.

1. En quoi consiste le phénomène de censure aléatoire droite? Comment le modélise-t-on? Comment exprime-t-on les observations en fonction des durées d'intérêt?
2. Ecrire l'équation du modèle de Cox ajusté ici. Interpréter les paramètres introduits.
3. Quelles sont les hypothèses sous-jacentes posées par le modèle?
4. Quelles techniques de vérification de ces hypothèses connaissez-vous?
5. Que pensez-vous de la qualité de l'ajustement du modèle présenté aux données?
6. Quelles conclusions pouvez-vous tirer des résultats présentés?

```
> data(veteran)
> attach(veteran)
> str(veteran)
'data.frame':  137 obs. of  8 variables:
 $ trt      : num  1 1 1 1 1 1 1 1 1 1 ...
 $ celltype: Factor w/ 4 levels "squamous","smallcell",...: 1 1 1 1 1 1 1 1 1 1 ...
 $ time     : num  72 411 228 126 118 10 82 110 314 100 ...
 $ status   : num  1 1 1 1 1 1 1 1 1 0 ...
 $ karno    : num  60 70 60 60 70 20 40 80 50 70 ...
 $ diagtime: num  7 5 3 9 11 5 10 29 18 6 ...
 $ age      : num  69 64 38 63 65 49 69 68 43 70 ...
```

```
$ prior : num 0 10 0 10 10 0 10 0 0 0 ...
```

```
> head(veteran)
```

	trt	celltype	time	status	karno	diagtime	age	prior
1	1	squamous	72	1	60	7	69	0
2	1	squamous	411	1	70	5	64	10
3	1	squamous	228	1	60	3	38	0
4	1	squamous	126	1	60	9	63	10
5	1	squamous	118	1	70	11	65	10
6	1	squamous	10	1	20	5	49	0

```
> summary(veteran)
```

trt		celltype		time		status	
Min.	:1.000	squamous	:35	Min.	: 1.0	Min.	:0.0000
1st Qu.	:1.000	smallcell	:48	1st Qu.	: 25.0	1st Qu.	:1.0000
Median	:1.000	adeno	:27	Median	: 80.0	Median	:1.0000
Mean	:1.496	large	:27	Mean	:121.6	Mean	:0.9343
3rd Qu.	:2.000			3rd Qu.	:144.0	3rd Qu.	:1.0000
Max.	:2.000			Max.	:999.0	Max.	:1.0000

karno		diagtime		age		prior	
Min.	:10.00	Min.	: 1.000	Min.	:34.00	Min.	: 0.00
1st Qu.	:40.00	1st Qu.	: 3.000	1st Qu.	:51.00	1st Qu.	: 0.00
Median	:60.00	Median	: 5.000	Median	:62.00	Median	: 0.00
Mean	:58.57	Mean	: 8.774	Mean	:58.31	Mean	: 2.92
3rd Qu.	:75.00	3rd Qu.	:11.000	3rd Qu.	:66.00	3rd Qu.	:10.00
Max.	:99.00	Max.	:87.000	Max.	:81.00	Max.	:10.00

```
> survdiff(formula=veteran~trt,data=veteran)
```

```
Call:
```

```
survdiff(formula = veteran ~ trt, data = veteran)
```

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
trt=1	69	64	64.5	0.00388	0.00823
trt=2	68	64	63.5	0.00394	0.00823

Chisq= 0 on 1 degrees of freedom, p= 0.928

```
> survdiff(formula=veteran~trt,data=lung,rho=1)
```

```
Call:
```

```
survdiff(formula = lung ~ trt, data = lung, rho = 1)
```

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
trt=1	69	32.2	35.4	0.279	0.871
trt=2	68	35.2	32.1	0.308	0.871

Chisq= 0.9 on 1 degrees of freedom, p= 0.351

```
> cox.vet1=coxph(veteran~factor(trt)+celltype+ karno+ diagtime+ age
+factor(10-prior),data=veteran)
```

```
> summary(cox.vet1)
```

```
Call:
```

```
coxph(formula = vet.surv ~ factor(trt) + celltype + karno + diagtime +
      age + factor(10 - prior), data = veteran)
```

n= 137, number of events= 128

	coef	exp(coef)	se(coef)	z	Pr(> z)	
factor(trt)2	2.946e-01	1.343e+00	2.075e-01	1.419	0.15577	
celltypesmallcell	8.616e-01	2.367e+00	2.753e-01	3.130	0.00175	**
celltypeadeno	1.196e+00	3.307e+00	3.009e-01	3.975	7.05e-05	***
celltypelarge	4.013e-01	1.494e+00	2.827e-01	1.420	0.15574	
karno	-3.282e-02	9.677e-01	5.508e-03	-5.958	2.55e-09	***
diagtime	8.132e-05	1.000e+00	9.136e-03	0.009	0.99290	
age	-8.706e-03	9.913e-01	9.300e-03	-0.936	0.34920	
factor(10 - prior)10	-7.159e-02	9.309e-01	2.323e-01	-0.308	0.75794	

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

	exp(coef)	exp(-coef)	lower .95	upper .95
factor(trt)2	1.3426	0.7448	0.8939	2.0166
celltypesmallcell	2.3669	0.4225	1.3799	4.0597
celltypeadeno	3.3071	0.3024	1.8336	5.9647
celltypelarge	1.4938	0.6695	0.8583	2.5996
karno	0.9677	1.0334	0.9573	0.9782
diagtime	1.0001	0.9999	0.9823	1.0182
age	0.9913	1.0087	0.9734	1.0096
factor(10 - prior)10	0.9309	1.0742	0.5904	1.4677

Concordance= 0.736 (se = 0.03)

Rsquare= 0.364 (max possible= 0.999)

Likelihood ratio test= 62.1 on 8 df, p=1.799e-10

Wald test = 62.37 on 8 df, p=1.596e-10

Score (logrank) test = 66.74 on 8 df, p=2.186e-11

Exercice 22.

La rétinopathie diabétique est une atteinte de la rétine survenant dans le cadre d'un diabète, pouvant évoluer vers un décollement rétinien synonyme de perte de la vision, en passant par l'oedème maculaire (la macula est le centre de la vision) ou par l'hémorragie rétinienne. L'évolution de la maladie suit schématiquement l'évolution simplifiée présentée en figure 28. Une étude de 4 ans est réalisée sur 20 patients atteints de diabète et dont les vaisseaux sanguins ont été révélés comme étant altérés dans la zone de la rétine à la suite d'un examen (fond d'oeil complété par une angiographie par fluorescence). L'entrée dans l'étude se fait au moment du diagnostic. L'étude démarre à la date du 1er mars 2000. La période de recrutement des patients est de 6 mois. Les données recueillies sont présentées dans la table 14.

1. On s'intéresse à la durée entre l'entrée dans l'étude et la survenue d'un premier évènement affectant la vision. Introduire les variables aléatoires permettant de modéliser les données correspondantes dans la table 14. A quelle(s) fonction(s) vous intéressez-vous? A partir des observations fournies dans la table 14, estimer la(les) fonction(s) précédemment introduite(s).

2. On s'intéresse à la durée entre l'entrée dans l'étude et la perte de vision. Introduire les variables aléatoires permettant de modéliser les données correspondantes dans la table 14. A quelle(s) fonction(s) vous intéressez-vous? A partir des observations fournies dans la table 14, estimer la(les) fonction(s) précédemment introduite(s).
3. On s'intéresse à la durée entre l'entrée dans l'étude et la survenue d'un premier oedème maculaire. A quelle(s) fonction(s) vous intéressez-vous? A partir des observations fournies dans la table 14, estimer la(les) fonction(s) d'intérêt précédemment introduite(s).
4. On recueille les informations suivantes chez chaque patient: âge, IMC (indice de masse corporelle), tension artérielle systolique, type de diabète (I ou II). Comment utiliser ces informations dans le cadre de chacune des trois questions précédentes? Proposer un modèle pour chaque cas et interpréter.

Exercice 23.

Le cancer de la vessie correspond à la multiplication anormale de cellules de la paroi de la vessie. La paroi est constituée de plusieurs couches. Les tumeurs se forment généralement dans la couche superficielle qui tapisse l'intérieur de la vessie (stade I). Elles envahissent ensuite les couches musculaires qui entourent l'organe (stade II). L'évolution se poursuit par un envahissement ganglionnaire (stade III) puis par des métastases (stade IV). Ci-dessous est présentée une brève analyse d'un jeu de données, effectuée au moyen du logiciel R. Il contient les durées de survie de 316 patients atteints d'un cancer de la vessie. Certaines des durées sont aléatoirement censurées à droite par la fin de l'étude ou par un décès sans lien avec le cancer de la vessie. La durée enregistrée correspond donc à la durée entre l'entrée dans l'étude (qui coïncide avec la date du diagnostic) et le décès dû au cancer de la vessie ou la censure. Sont recueillies les covariables suivantes:

- **age**: âge exprimé en années,
- **sex**: 1 = homme, 2 = femme,
- **stade**: stade du cancer au moment du diagnostic,
- **tabac**: 1 = non-fumeur, 2 = fumeur occasionnel, 3 = fumeur régulier, 4 = fumeur intensif,
- **trt**: type de traitement reçu codé par 1 = ablation de la tumeur, 2 = ablation de la vessie et des ganglions du bassin, 3 = chimiothérapie, 4 = radiothérapie.

1. Ecrire l'équation du modèle de Cox ajusté ici.
2. Quelles sont les hypothèses sous-jacentes au modèle?
3. Quelles techniques de vérification de ces hypothèses connaissez-vous?
4. Que pensez-vous de la qualité de l'ajustement du modèle présenté aux données?
5. Interpréter les paramètres introduits. Donner leur estimation. Quelles conclusions pouvez-vous tirer des résultats présentés?

```

> library(survival)
Le chargement a nécessité le package : splines
> library(MASS)
>
> mysurv<- Surv(t,event=D)
> summary(mysurv)
      time          status
Min.   : 0.5203   Min.    :0.0000
1st Qu.: 9.7196   1st Qu.:0.0000
Median :11.9225   Median :1.0000
Mean   :11.5825   Mean    :0.6835
3rd Qu.:13.9712   3rd Qu.:1.0000
Max.   :20.4397   Max.    :1.0000
>
> g<-3
> quantile(age,probs=seq(0,1,1/g))
      0% 33.33333% 66.66667%      100%
40.11330 54.61825 67.39133 79.70297
> age.factor<-cut(age,breaks=quantile(age,probs=seq(0,1,1/g)),include.lowest=T)
> fit.byage=survfit(mysurv~age.factor,conf.type="none")
> x11()
> plot(fit.byage, fun="cumhaz",xlab="t",ylab="cumulative hazard",
+ main="Nelson-Aalen estimates by age", conf.int=TRUE,
+ col=c("gray","black","red"),lty=c(2,1,4))
> legend("topleft",col=c("gray","black","red"),lty=c(2,1,4),
+ legend=c("<55","entre 55 et 67",">67"))
>
> myfit.sex<- survfit(mysurv~sex)
> x11()
> plot(myfit.sex,fun="cumhaz",xlab="t",ylab="survival probability",
+ main="Nelson-Aalen estimates by sex subgroup",
+ conf.int=F,col=c("gray","black"),lty=c(2,2))
> legend("topleft",col=c("gray","black"),lty=c(2,2),legend=c("sex: 1","sex: 2"))
>
> myfit.stade<- survfit(mysurv~stade)
> x11()
> plot(myfit.stade,fun="cumhaz",xlab="t",ylab="survival probability",
+ main="Nelson-Aalen estimates by stade subgroup",
+ conf.int=F,col=c("gray","black","gray","red"),lty=c(2,2,1,4))
> legend("topleft",col=c("gray","black","gray","red"),lty=c(2,2,1,4),
+ legend=c("stade: 1","stade: 2","stade: 3","stade: 4"))
>
> myfit.tabac<- survfit(mysurv~tabac)
> x11()
> plot(myfit.tabac,fun="cumhaz",xlab="t",ylab="survival probability",
+ main="Nelson-Aalen estimates by tobacco subgroup",
+ conf.int=F,col=c("gray","black","gray","red"),lty=c(2,2,1,4))
> legend("topleft",col=c("gray","black","gray","red"),lty=c(2,2,1,4),

```

```

+ legend=c("tabac: 1","tabac: 2","tabac: 3","tabac: 4"))
>
> myfit.trt<- survfit(mysurv~trt)
> x11()
> plot(myfit.trt,fun="cumhaz",xlab="t",ylab="survival probability",
+ main="Nelson-Aalen estimates by treatment subgroup",
+ conf.int=F,col=c("gray","black","gray","red"),lty=c(2,2,1,4))
> legend("topleft",col=c("gray","black","gray","red"),lty=c(2,2,1,4),
+ legend=c("trt: 1","trt: 2","trt: 3","trt: 4"))
>
> mycox<- coxph(mysurv~age+sex+tabac+stade*trt)
> summary(mycox)
Call:
coxph(formula = mysurv ~ age + sex + tabac + stade * trt)

```

n= 316, number of events= 216

	coef	exp(coef)	se(coef)	z	Pr(> z)	
age	-0.05641	0.94515	0.00737	-7.653	1.95e-14	***
sex2	-0.27244	0.76152	0.14666	-1.858	0.063212	.
tabac2	0.73688	2.08941	0.20509	3.593	0.000327	***
tabac3	2.02535	7.57879	0.22823	8.874	< 2e-16	***
tabac4	4.61731	101.22125	0.41892	11.022	< 2e-16	***
stade2	0.76895	2.15749	0.46492	1.654	0.098144	.
stade3	1.57652	4.83809	0.47793	3.299	0.000972	***
stade4	5.88373	359.14625	0.88562	6.644	3.06e-11	***
trt2	-0.39535	0.67344	0.52399	-0.754	0.450550	
trt3	-0.60869	0.54407	0.49291	-1.235	0.216875	
trt4	-0.07235	0.93020	0.49877	-0.145	0.884664	
stade2:trt2	0.39652	1.48664	0.61288	0.647	0.517646	
stade3:trt2	0.85074	2.34138	0.62308	1.365	0.172136	
stade4:trt2	1.78762	5.97520	1.10821	1.613	0.106730	
stade2:trt3	0.61003	1.84048	0.55784	1.094	0.274148	
stade3:trt3	-0.94486	0.38873	0.74110	-1.275	0.202329	
stade4:trt3	-0.92442	0.39676	1.03967	-0.889	0.373924	
stade2:trt4	-0.14310	0.86667	0.59884	-0.239	0.811135	
stade3:trt4	-2.93952	0.05289	0.63164	-4.654	3.26e-06	***
stade4:trt4	-2.09960	0.12251	1.13977	-1.842	0.065456	.

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

	exp(coef)	exp(-coef)	lower .95	upper .95
age	0.94515	1.058029	0.93160	0.9589
sex2	0.76152	1.313171	0.57127	1.0151
tabac2	2.08941	0.478603	1.39783	3.1232
tabac3	7.57879	0.131947	4.84543	11.8541
tabac4	101.22125	0.009879	44.53339	230.0688
stade2	2.15749	0.463501	0.86737	5.3665

stade3	4.83809	0.206693	1.89607	12.3451
stade4	359.14625	0.002784	63.30400	2037.5651
trt2	0.67344	1.484906	0.24115	1.8807
trt3	0.54407	1.838013	0.20706	1.4296
trt4	0.93020	1.075033	0.34997	2.4725
stade2:trt2	1.48664	0.672657	0.44722	4.9419
stade3:trt2	2.34138	0.427098	0.69040	7.9404
stade4:trt2	5.97520	0.167358	0.68083	52.4401
stade2:trt3	1.84048	0.543336	0.61674	5.4924
stade3:trt3	0.38874	2.572446	0.09096	1.6614
stade4:trt3	0.39676	2.520412	0.05171	3.0444
stade2:trt4	0.86667	1.153844	0.26799	2.8027
stade3:trt4	0.05289	18.906789	0.01534	0.1824
stade4:trt4	0.12251	8.162931	0.01312	1.1437

Concordance= 0.827 (se = 0.023)
 Rsquare= 0.635 (max possible= 0.998)
 Likelihood ratio test= 318.4 on 20 df, p=0
 Wald test = 265.9 on 20 df, p=0
 Score (logrank) test = 683.8 on 20 df, p=0

```
> drop1(mycox,test="Chisq")
Single term deletions
```

Model:

```
mysurv ~ age + sex + tabac + stade * trt
      Df   AIC   LRT Pr(>Chi)
<none> 1684.3
age     1 1744.3 62.030 3.382e-15 ***
sex     1 1685.7  3.451  0.06322 .
tabac   3 1815.4 137.153 < 2.2e-16 ***
stade:trt 9 1734.5 68.200 3.428e-11 ***
---
```

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

```
> mycox<- coxph(mysurv~age+tabac+stade*trt)
> drop1(mycox,test="Chisq")
Single term deletions
```

Model:

```
mysurv ~ age + tabac + stade * trt
      Df   AIC   LRT Pr(>Chi)
<none> 1685.7
age     1 1744.2 60.432 7.618e-15 ***
tabac   3 1817.1 137.410 < 2.2e-16 ***
stade:trt 9 1733.9 66.131 8.689e-11 ***
---
```

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

```
> summary(mycox)
```

Call:

coxph(formula = mysurv ~ age + tabac + stade * trt)

n= 316, number of events= 216

	coef	exp(coef)	se(coef)	z	Pr(> z)	
age	-0.05492	0.94656	0.00723	-7.595	3.06e-14	***
tabac2	0.76482	2.14861	0.20472	3.736	0.000187	***
tabac3	2.03152	7.62570	0.22768	8.923	< 2e-16	***
tabac4	4.60906	100.38984	0.41887	11.004	< 2e-16	***
stade2	0.76727	2.15387	0.46521	1.649	0.099089	.
stade3	1.58490	4.87881	0.47738	3.320	0.000900	***
stade4	5.76970	320.44292	0.88178	6.543	6.02e-11	***
trt2	-0.32182	0.72483	0.52180	-0.617	0.537398	
trt3	-0.54805	0.57808	0.49174	-1.115	0.265060	
trt4	-0.10543	0.89994	0.49787	-0.212	0.832289	
stade2:trt2	0.35515	1.42639	0.61250	0.580	0.562026	
stade3:trt2	0.79980	2.22509	0.62189	1.286	0.198415	
stade4:trt2	1.72767	5.62752	1.11478	1.550	0.121194	
stade2:trt3	0.59074	1.80532	0.55751	1.060	0.289322	
stade3:trt3	-1.03317	0.35588	0.73960	-1.397	0.162431	
stade4:trt3	-0.74111	0.47658	1.03503	-0.716	0.473970	
stade2:trt4	-0.09014	0.91381	0.59820	-0.151	0.880230	
stade3:trt4	-2.86617	0.05692	0.62843	-4.561	5.10e-06	***
stade4:trt4	-2.05931	0.12754	1.13731	-1.811	0.070191	.

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

	exp(coef)	exp(-coef)	lower .95	upper .95
age	0.94656	1.056454	0.93324	0.9601
tabac2	2.14861	0.465418	1.43845	3.2094
tabac3	7.62570	0.131135	4.88064	11.9147
tabac4	100.38984	0.009961	44.17198	228.1564
stade2	2.15387	0.464280	0.86543	5.3605
stade3	4.87881	0.204968	1.91413	12.4353
stade4	320.44292	0.003121	56.90833	1804.3697
trt2	0.72483	1.379634	0.26067	2.0155
trt3	0.57808	1.729877	0.22050	1.5155
trt4	0.89994	1.111191	0.33918	2.3878
stade2:trt2	1.42639	0.701069	0.42942	4.7381
stade3:trt2	2.22509	0.449420	0.65765	7.5283
stade4:trt2	5.62752	0.177698	0.63301	50.0289
stade2:trt3	1.80532	0.553918	0.60535	5.3840
stade3:trt3	0.35588	2.809963	0.08351	1.5165
stade4:trt3	0.47658	2.098274	0.06268	3.6237
stade2:trt4	0.91381	1.094323	0.28292	2.9515
stade3:trt4	0.05692	17.569513	0.01661	0.1951
stade4:trt4	0.12754	7.840529	0.01373	1.1851

```

Concordance= 0.827 (se = 0.023 )
Rsquare= 0.631 (max possible= 0.998 )
Likelihood ratio test= 314.9 on 19 df, p=0
Wald test = 265.1 on 19 df, p=0
Score (logrank) test = 681.9 on 19 df, p=0
>
> mytest=cox.zph(mycox)
> print(mytest)

```

	rho	chisq	p
age	-0.04780	0.60724	0.436
tabac2	0.05094	0.57179	0.450
tabac3	0.04289	0.40920	0.522
tabac4	0.07179	1.25091	0.263
stade2	-0.02418	0.12877	0.720
stade3	0.02252	0.10566	0.745
stade4	0.01174	0.02851	0.866
trt2	-0.02280	0.11682	0.733
trt3	-0.02536	0.13744	0.711
trt4	-0.01224	0.03120	0.860
stade2:trt2	0.03545	0.28545	0.593
stade3:trt2	-0.01568	0.05574	0.813
stade4:trt2	0.01093	0.02947	0.864
stade2:trt3	0.03356	0.24148	0.623
stade3:trt3	0.02799	0.16442	0.685
stade4:trt3	0.01765	0.06757	0.795
stade2:trt4	0.04061	0.35277	0.553
stade3:trt4	-0.08314	1.53790	0.215
stade4:trt4	0.00516	0.00544	0.941
GLOBAL	NA	7.34802	0.992

```

>
> x11()
> par(mfrow=c(4,5))
> plot(mytest)
>
> res.mart=residuals(mycox,type="martingale")
> x11()
> par(mfrow=c(1,2))
> plot(age,res.mart,xlab="age",ylab="martingale residuals")
> lines(lowess(age,res.mart,iter=0))
> plot(t,res.mart,xlab="observed lifetime",ylab="martingale residuals")
> lines(lowess(t,res.mart,iter=0))
>
> res.dfbetas=residuals(mycox,type="dfbetas")
> x11()
> par(mfrow=c(4,5))
> plot(1:n,res.dfbetas[,1],xlab="subject",ylab="dfbetas residuals for age")
> lines(lowess(1:n,res.dfbetas[,1],iter=0))

```

```

> plot(1:n,res.dfbetas[,2],xlab="subject",ylab="dfbetas residuals for stade=2")
> lines(lowess(1:n,res.dfbetas[,2],iter=0))
> plot(1:n,res.dfbetas[,3],xlab="subject",ylab="dfbetas residuals for stade=3")
> lines(lowess(1:n,res.dfbetas[,3],iter=0))
> plot(1:n,res.dfbetas[,4],xlab="subject",ylab="dfbetas residuals for stade=4")
> lines(lowess(1:n,res.dfbetas[,4],iter=0))
> plot(1:n,res.dfbetas[,5],xlab="subject",ylab="dfbetas residuals for tobacco=2")
> lines(lowess(1:n,res.dfbetas[,5],iter=0))
> plot(1:n,res.dfbetas[,6],xlab="subject",ylab="dfbetas residuals for tobacco=3")
> lines(lowess(1:n,res.dfbetas[,6],iter=0))
> plot(1:n,res.dfbetas[,7],xlab="subject",ylab="dfbetas residuals for tobacco=4")
> lines(lowess(1:n,res.dfbetas[,7],iter=0))
> plot(1:n,res.dfbetas[,8],xlab="subject",ylab="dfbetas residuals for trt=2")
> lines(lowess(1:n,res.dfbetas[,8],iter=0))
> plot(1:n,res.dfbetas[,9],xlab="subject",ylab="dfbetas residuals for trt=3")
> lines(lowess(1:n,res.dfbetas[,9],iter=0))
> plot(1:n,res.dfbetas[,10],xlab="subject",ylab="dfbetas residuals for trt=4")
> lines(lowess(1:n,res.dfbetas[,10],iter=0))
> plot(1:n,res.dfbetas[,11],xlab="subject",ylab="dfbetas residuals for stade=2, trt=2")
> lines(lowess(1:n,res.dfbetas[,11],iter=0))
> plot(1:n,res.dfbetas[,12],xlab="subject",ylab="dfbetas residuals for stade=3, trt=2")
> lines(lowess(1:n,res.dfbetas[,12],iter=0))
> plot(1:n,res.dfbetas[,13],xlab="subject",ylab="dfbetas residuals for stade=4, trt=2")
> lines(lowess(1:n,res.dfbetas[,13],iter=0))
> plot(1:n,res.dfbetas[,14],xlab="subject",ylab="dfbetas residuals for stade=2, trt=3")
> lines(lowess(1:n,res.dfbetas[,14],iter=0))
> plot(1:n,res.dfbetas[,15],xlab="subject",ylab="dfbetas residuals for stade=3, trt=3")
> lines(lowess(1:n,res.dfbetas[,15],iter=0))
> plot(1:n,res.dfbetas[,16],xlab="subject",ylab="dfbetas residuals for stade=4, trt=3")
> lines(lowess(1:n,res.dfbetas[,16],iter=0))
> plot(1:n,res.dfbetas[,17],xlab="subject",ylab="dfbetas residuals for stade=2, trt=4")
> lines(lowess(1:n,res.dfbetas[,17],iter=0))
> plot(1:n,res.dfbetas[,18],xlab="subject",ylab="dfbetas residuals for stade=3, trt=4")
> lines(lowess(1:n,res.dfbetas[,18],iter=0))
> plot(1:n,res.dfbetas[,19],xlab="subject",ylab="dfbetas residuals for stade=4, trt=4")
> lines(lowess(1:n,res.dfbetas[,19],iter=0))

```

Exercice 24.

Le lymphome est un cancer du système lymphatique qui se développe au dépend des lymphocytes. Le lymphome non-hodgkinien (LNH) prend généralement naissance dans les ganglions lymphatiques. Le LNH peut se propager par le système lymphatique, et parfois par la circulation sanguine, à presque n'importe quel tissu ou organe du corps. Quand il se propage à un organe ou à du tissu situé à l'extérieur des ganglions lymphatiques, on parle de propagation extra-ganglionnaire. Une étude de 10 ans est réalisée sur 23 patients atteints de LNH. L'entrée dans l'étude se fait au moment du début du traitement. L'étude démarre à la date du 1er mars 2000. La période de recrutement des patients est de 1 an. Les données recueillies sont présentées dans la table 16. Sauf mention explicite du contraire, le décès est lié au LNH ou à son traitement.

1. En quoi consiste le phénomène de risques concurrents? Comment le modélise-t-on? Comment exprime-t-on les observations? Proposer un exemple concret de modèle de risques concurrents mutuellement indépendants et donner la(les) fonction(s) d'intérêt. Proposer un exemple concret de modèle de risques concurrents dépendants et donner la(les) fonction(s) d'intérêt.
2. On s'intéresse à la durée de rémission (entre l'entrée dans l'étude et la survenue d'un premier évènement, fatal ou non, lié au LNH). Introduire les variables aléatoires permettant de modéliser les données correspondantes dans la table 16. A quelle(s) fonction(s) vous intéressez-vous? A partir des observations fournies dans la table 16, estimer la(les) fonction(s) précédemment introduite(s).
3. On s'intéresse à la durée de survie au lymphome (entre l'entrée dans l'étude et le décès lié au lymphome ou à son traitement). Introduire les variables aléatoires permettant de modéliser les données correspondantes dans la table 16. A quelle(s) fonction(s) vous intéressez-vous? A partir des observations fournies dans la table 16, estimer la(les) fonction(s) précédemment introduite(s).
4. On s'intéresse à la durée entre l'entrée dans l'étude et la 1^{ère} survenue d'une récurrence ou de propagation extra-ganglionnaire. A quelle(s) fonction(s) vous intéressez-vous? A partir des observations fournies dans la table 16, estimer la(les) fonction(s) d'intérêt précédemment introduite(s).

Exercice 25.

La cirrhose du foie ou cirrhose hépatique est une maladie chronique du foie, généralement considérée comme irréversible, dans laquelle l'architecture hépatique se désorganise (le foie devient dur, augmente de volume et prend un aspect granuleux), ce qui s'accompagne d'une altération des fonctions hépatiques. Son évolution simplifiée est représentée schématiquement sur la figure 37. Une étude sur 20 patients atteints de cirrhose du foie est réalisée sur 15 ans. L'entrée des patients dans l'étude intervient entre le 1er février 1990 et le 1er février 1992, et coïncide avec le moment où le diagnostic de cirrhose du foie est posé. Différents facteurs de risque sont également relevés au moment du diagnostic. Les données recueillies sont présentées dans la table 17. Sauf mention explicite du contraire, le décès est lié à la cirrhose du foie. Vous veillerez à introduire les variables aléatoires permettant de modéliser les phénomènes étudiés et les données correspondantes dans la table 17.

patiente	observation	âge (en années)	taille de la tumeur (en cm)	pratique d'exercice physique	consommation d'alcool	prise de TH	centre de soins
1	rechute locale au bout de 16 mois rechute distante au bout de 18 mois décès au bout de 24 mois	54	4,1	occasionnellement	occasionnelle	oui	C1
2	rechute distante au bout de 16 mois décès au bout de 24 mois	62	4,2	jamais	régulière	oui	C1
3	rechute locale au bout de 2 ans rechute locale au bout de 3 ans décès au bout de 4 ans	51	2,1	régulièrement	occasionnelle	non	C1
4	rechute locale au bout de 3 ans rechute locale au bout de 4 ans	58	3,2	régulièrement	occasionnelle	non	C1
5	rechute locale au bout de 5 ans	53	1,2	régulièrement	occasionnelle	non	C2
6	rechute distante au bout de 12 mois rechute distante au bout de 18 mois	71	2,8	régulièrement	régulière	oui	C2
7	atteinte des ganglions axillaires au bout de 10 mois rechute distante au bout de 18 mois décès au bout de 6 ans	61	2,6	occasionnellement	jamais	non	C2
8	perdue de vue au bout de 6 ans	62	1,8	occasionnellement	occasionnelle	non	C3
9	rechute distante au bout de 14 mois rechute distante au bout de 18 mois	57	4,7	occasionnellement	excessive	oui	C3
10	rechute locale au bout de 3 ans	56	2,8	jamais	excessive	non	C3
11	rechute locale au bout de 4 mois encore en vie à la fin de l'étude entrée au 01/04/1995	54	1,0	régulièrement	occasionnelle	oui	C4
12	rechute distante au bout de 14 mois perdue de vue au bout de 2 ans	60	3,5	occasionnellement	excessive	non	C4
13	encore en vie à la fin de l'étude entrée au 01/03/1994	52	1,8	intensivement	occasionnelle	non	C4
14	rechute locale au bout de 7 mois atteinte des ganglions axillaires au bout de 1,5 an décès au bout de 2 ans	56	3,9	jamais	occasionnelle	oui	C5
15	atteinte des ganglions axillaires au bout de 10 mois perdue de vue au bout de 2 ans	69	2,8	jamais	jamais	oui	C5
16	perdue de vue au bout de 8 ans	57	1,4	intensivement	régulière	non	C5
17	encore en vie à la fin de l'étude (entrée au 01/06/1999)	58	0,9	régulièrement	régulière	non	C5

Table 11 – Observations réalisées pour les 17 patientes du groupe A

n° de patient	observation
1	décès dû au SIDA au bout de 6 ans
2	décès dû au SIDA au bout de 6 ans
3	décès dû à la tuberculose au bout de 5 ans
4	décès dû à la tuberculose au bout de 7 ans
5	décès dû au SIDA au bout de 6 ans
6	décès dû au SIDA au bout de 8.5 ans
7	décès dû au SIDA au bout de 8.5 ans
8	décès dû au SIDA au bout de 8.5 ans
9	décès dû à la tuberculose au bout de 12 mois
10	décès dû au SIDA au bout de 3 ans
11	rien ne se produit pendant les 10 années de l'étude
12	décès dû à la tuberculose au bout de 9 ans
13	perdu de vue après 7.5 années de suivi
14	décès dû au SIDA au bout de 4.5 ans
15	décès dû au SIDA au bout de 3 ans
16	décès naturel après 6 ans de suivi
17	décès dû à la tuberculose au bout de 18 mois
18	décès par traumatologie après 18 mois de suivi
19	décès dû au SIDA au bout de 4 ans
20	décès dû à la tuberculose au bout de 7 ans
21	décès dû au SIDA au bout de 4 ans
22	décès dû à la tuberculose au bout de 8 ans
23	décès dû au SIDA au bout de 9.5 ans
24	rien ne se produit pendant les 10 années de l'étude
25	décès dû au SIDA au bout de 7.5 ans
26	décès dû à la tuberculose au bout de 8 ans
27	décès dû au SIDA au bout de 9.5 ans
28	décès dû au SIDA au bout de 7.5 ans
29	perdu de vue après 5 années
30	décès dû à la tuberculose au bout de 24 mois

Table 12 – Observations réalisées pour les 30 patients séropositifs

identifiant	observation
1	diagnostic de cancer primitif des bronches 31 mois après le début du suivi
2	diagnostic de cancer primitif des bronches 56 mois après le début du suivi
3	rien se produit pendant les 10 années de l'étude
4	diagnostic de cardiopathie ischémique 14 mois après le début du suivi
5	diagnostic de cardiopathie ischémique 45 mois après le début du suivi
6	diagnostic de cardiopathie ischémique 45 mois après le début du suivi
7	diagnostic de cardiopathie ischémique 45 mois après le début du suivi
8	diagnostic de cardiopathie ischémique 38 mois après le début du suivi
9	plus de nouvelles après 6.5 années
10	rien se produit pendant les 10 années de l'étude
11	diagnostic de cancer primitif des bronches 21 mois après le début du suivi
12	décès par traumatologie 21 mois après le début du suivi
13	décès par accident ischémique 57 mois après le début du traitement
14	rien se produit pendant les 10 années de l'étude
15	diagnostic de cancer primitif des bronches 56 mois après le début du suivi
16	décès naturel après 5 années de suivi
17	diagnostic de cancer primitif des bronches 24 mois après le début du suivi
18	rien se produit pendant les 10 années de l'étude
19	diagnostic de cancer primitif des bronches 56 mois après le début du suivi
20	diagnostic de cancer primitif des bronches 56 mois après le début du suivi
21	déménagement après 8 années de suivi
22	diagnostic de cancer primitif des bronches 11 mois après le début du suivi
23	arrête de fumer 1 an après le début du suivi
24	diagnostic de cancer primitif des bronches 18 mois après le début du suivi
25	décès par accident ischémique 2.5 années après le début du suivi
26	diagnostic de cancer primitif des bronches 47 mois après le début du suivi
27	diagnostic de cardiopathie ischémique 76 mois après le début du suivi
28	diagnostic de cancer primitif des bronches 78 mois après le début du suivi
29	diagnostic de cardiopathie ischémique 30 mois après le début du suivi
30	diagnostic de cancer primitif des bronches 31 mois après le début du suivi

Table 13 – Données recueillies sur 30 fumeurs

1. On s'intéresse à la durée de survie des patients à compter de la date du diagnostic de cirrhose du foie. A quelle(s) fonction(s) vous intéressez-vous? A partir des observations fournies dans la table 17, estimer la(les) fonction(s) précédemment introduite(s).
2. On s'intéresse à la durée sans évènement lié à la cirrhose du foie (ni encéphalopathie hépatique, ni perte de fonctionnalité du foie, ni cancer primitif du foie, ni décès) comptée depuis le diagnostic de cirrhose du foie. A quelle(s) fonction(s) vous intéressez-vous? A partir des observations fournies dans la table 17, estimer la(les) fonction(s) précédemment introduite(s).
3. On s'intéresse à la durée entre le diagnostic de cirrhose du foie et le diagnostic de cancer primitif du foie. A quelle(s) fonction(s) vous intéressez-vous? A partir des observations fournies dans la table 17, estimer la(les) fonction(s) d'intérêt précédemment introduite(s).
4. On s'intéresse à la durée entre le diagnostic de cirrhose du foie et la survenue d'un évènement non-fatal lié à l'évolution de la cirrhose du foie. A quelle(s) fonction(s) vous intéressez-vous? A partir des observations fournies dans la table 17, estimer la(les) fonction(s) d'intérêt précédemment introduite(s).
5. On souhaite évaluer l'impact des différents facteurs de risque relevés au moment du diagnostic sur l'évolution du pronostic vital au cours du temps. Proposer un modèle adapté à ce travail. Quelles techniques de vérification des hypothèses sous-jacentes à ce modèle connaissez-vous?
6. On souhaite évaluer l'impact des différents facteurs de risque relevés au moment du diagnostic sur l'évolution au cours du temps du risque d'encéphalopathie hépatique. Proposer un modèle adapté à ce travail.

Exercice 26.

Ci-dessous est présentée une brève analyse (effectuée au moyen du logiciel R) de la survie chez 287 patients atteints d'un lymphome non-hodgkinien (LNH). Les patients rentrent dans l'étude lorsqu'ils débutent leur traitement. Certaines des durées de survie sont aléatoirement censurées à droite par la fin de l'étude ou par un décès sans lien avec la maladie ou son traitement. La durée enregistrée correspond donc à la durée entre le début du traitement et le décès dû au LNH ou la censure. Les prédicteurs suivants sont pris en considération:

- $\text{age} = \hat{\text{age}}$ (exprimé en années)
- $\text{type} = \begin{cases} 1 & \text{si le LNH est un lymphome à lymphocytes B,} \\ 2 & \text{si le LNH est lymphome à lymphocytes T.} \end{cases}$

- $\text{stade} = \begin{cases} 1 & \text{si le cancer est présent dans un seul groupe ganglionnaire ou dans un seul organe ou région à l'extérieur des ganglions lymphatiques (début de la maladie),} \\ 2 & \text{si le cancer est présent dans plusieurs groupes ganglionnaire situés du même côté du diaphragme (maladie localisée avancée),} \\ 3 & \text{si le cancer est présent dans les groupes ganglionnaires situés des deux côtés du diaphragme (maladie avancée),} \\ 4 & \text{si le cancer a atteint plusieurs organes en dehors des ganglions lymphatiques et de la rate, et s'il s'est étendu à un ou plusieurs organes tels que les os, la moelle osseuse, la peau ou le foie (extension de la maladie).} \end{cases}$
- $\text{grade} = \begin{cases} 1 & \text{si le LNH est indolent (de bas grade, ou d'évolution lente),} \\ 2 & \text{si le LNH est agressif (de haut grade, ou d'évolution rapide).} \end{cases}$
- $\text{sympt} = '0'$ en l'absence des symptômes: fièvre inexplicquée, des sueurs nocturnes importantes et une perte de poids inexplicquée, $= '1'$ en présence de ces symptômes
- $\text{trt} = \begin{cases} 1 & \text{si le patient reçoit une chimiothérapie,} \\ 2 & \text{si le patient reçoit une chimiothérapie et une radiothérapie,} \\ 3 & \text{le patient reçoit une chimiothérapie et une immunothérapie.} \end{cases}$

1. Ecrire l'équation du modèle ajusté ici.
2. Quelles sont les hypothèses sous-jacentes au modèle?
3. Quelles techniques de vérification de ces hypothèses connaissez-vous?
4. Que pensez-vous de la qualité de l'ajustement du modèle proposé aux données?
5. Interpréter les paramètres introduits. Donner leur estimation. Quelles conclusions pouvez-vous tirer des résultats présentés?

```
> library(survival)
Le chargement a nécessité le package : splines
> library(MASS)
>
> mysurv<- Surv(t,event=D)
> summary(mysurv)
      time      status
Min.   : 0.5154  Min.   :0.0000
1st Qu.: 8.3658  1st Qu.:0.0000
Median :11.0149  Median :1.0000
Mean   :10.5891  Mean   :0.6934
3rd Qu.:13.2715  3rd Qu.:1.0000
Max.   :19.7577  Max.   :1.0000
>
> g<-3
```

```

> quantile(age,probs=seq(0,1,1/g))
      0% 33.33333% 66.66667%      100%
15.13133 33.13473 53.30369 69.85322
> age.factor<-cut(age,breaks=quantile(age,probs=seq(0,1,1/g)),include.lowest=T)
> fit.byage=survfit(mysurv~age.factor,conf.type="none")
> x11()
> plot(fit.byage, fun="cumhaz",xlab="t",ylab="cumulative hazard",
+ main="Nelson-Aalen estimates by age",conf.int=F,
+ col=c("gray","black","red"),lty=c(2,1,4))
> legend("topleft",col=c("gray","black","red"),lty=c(2,1,4),
+ legend=c("entre 15 et 33","entre 33 et 53","entre 53 et 70"))
>
> myfit.type<- survfit(mysurv~type)
> x11()
> plot(myfit.type,fun="cumhaz",xlab="t",ylab="cumulative hazard",
+ main="Nelson-Aalen estimates by type subgroup",conf.int=F,
+ col=c("gray","black"),lty=c(2,2))
> legend("topleft",col=c("gray","black"),lty=c(2,2),legend=c("type: 1","type: 2"))
>
> myfit.stade<- survfit(mysurv~stade)
> x11()
> plot(myfit.stade,fun="cumhaz",xlab="t",ylab="cumulative hazard",
+ main="Nelson-Aalen estimates by stade subgroup",
+ conf.int=F,col=c("gray","black","gray","red"),lty=c(2,2,1,4))
> legend("topleft",col=c("gray","black","gray","red"),lty=c(2,2,1,4),
+ legend=c("stade: 1","stade: 2","stade: 3","stade: 4"))
>
> myfit.grade<- survfit(mysurv~grade)
> x11()
> plot(myfit.grade,fun="cumhaz",xlab="t",ylab="cumulative hazard",
+ main="Nelson-Aalen estimates by grade subgroup", conf.int=F,
+ col=c("gray","black"),lty=c(2,2))
> legend("topleft",col=c("gray","black"),lty=c(2,2),legend=c("type: 1","type: 2"))
>
> myfit.sympt<- survfit(mysurv~sympt)
> x11()
> plot(myfit.sympt,fun="cumhaz",xlab="t",ylab="cumulative hazard",
+ main="Nelson-Aalen estimates by symptome subgroup",conf.int=F,
+ col=c("gray","black"),lty=c(2,2))
> legend("topleft",col=c("gray","black"),lty=c(2,2),legend=c("type: 1","type: 2"))
>
> myfit.trt<- survfit(mysurv~trt)
> x11()
> plot(myfit.trt,fun="cumhaz",xlab="t",ylab="cumulative hazard",
+ main="Nelson-Aalen estimates by treatment subgroup",conf.int=F,
+ col=c("gray","black","gray","red"),lty=c(2,2,1,4))
> legend("topleft",col=c("gray","black","gray","red"),lty=c(2,2,1,4),
+ legend=c("trt: 1","trt: 2","trt: 3","trt: 4"))

```

```
>
>
> mycox<- coxph(mysurv~age+type+stade+grade+sympt+trt+trt:type+trt:stade)
> summary(mycox)
```

Call:

```
coxph(formula = mysurv ~ age + type + stade + grade + sympt +
      trt + trt:type + trt:stade)
```

n= 287, number of events= 199

	coef	exp(coef)	se(coef)	z	Pr(> z)	
age	-0.058139	0.943518	0.005829	-9.974	< 2e-16	***
type2	1.069334	2.913439	0.260212	4.109	3.97e-05	***
stade2	0.971600	2.642168	0.325908	2.981	0.00287	**
stade3	1.866095	6.463012	0.342167	5.454	4.93e-08	***
stade4	6.066722	431.264889	0.638631	9.500	< 2e-16	***
grade2	2.248717	9.475575	0.200124	11.237	< 2e-16	***
sympt1	1.055943	2.874684	0.167086	6.320	2.62e-10	***
trt2	0.010194	1.010246	0.406001	0.025	0.97997	
trt3	-0.214121	0.807250	0.413961	-0.517	0.60498	
type2:trt2	0.218477	1.244180	0.379081	0.576	0.56439	
type2:trt3	-3.333959	0.035652	0.418432	-7.968	1.67e-15	***
stade2:trt2	-0.082258	0.921034	0.480235	-0.171	0.86400	
stade3:trt2	-2.951426	0.052265	0.517419	-5.704	1.17e-08	***
stade4:trt2	-1.010478	0.364045	0.854779	-1.182	0.23715	
stade2:trt3	0.435910	1.546369	0.492199	0.886	0.37581	
stade3:trt3	0.999117	2.715883	0.516514	1.934	0.05307	.
stade4:trt3	0.988323	2.686724	0.727253	1.359	0.17415	

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

	exp(coef)	exp(-coef)	lower .95	upper .95
age	0.94352	1.059863	0.93280	9.544e-01
type2	2.91344	0.343237	1.74950	4.852e+00
stade2	2.64217	0.378477	1.39491	5.005e+00
stade3	6.46301	0.154727	3.30508	1.264e+01
stade4	431.26489	0.002319	123.35045	1.508e+03
grade2	9.47557	0.105534	6.40118	1.403e+01
sympt1	2.87468	0.347864	2.07189	3.989e+00
trt2	1.01025	0.989858	0.45587	2.239e+00
trt3	0.80725	1.238773	0.35863	1.817e+00
type2:trt2	1.24418	0.803742	0.59185	2.616e+00
type2:trt3	0.03565	28.049180	0.01570	8.096e-02
stade2:trt2	0.92103	1.085736	0.35933	2.361e+00
stade3:trt2	0.05227	19.133216	0.01896	1.441e-01
stade4:trt2	0.36404	2.746913	0.06817	1.944e+00
stade2:trt3	1.54637	0.646676	0.58932	4.058e+00
stade3:trt3	2.71588	0.368204	0.98686	7.474e+00

```
stade4:trt3 2.68672 0.372200 0.64593 1.118e+01
```

```
Concordance= 0.864 (se = 0.024 )  
Rsquare= 0.701 (max possible= 0.998 )  
Likelihood ratio test= 346.1 on 17 df, p=0  
Wald test = 250.3 on 17 df, p=0  
Score (logrank) test = 384.4 on 17 df, p=0
```

```
> drop1(mycox,test="Chisq")
```

```
Single term deletions
```

```
Model:
```

```
mysurv ~ age + type + stade + grade + sympt + trt + trt:type +  
trt:stade
```

	Df	AIC	LRT	Pr(>Chi)
<none>		1478.9		
age	1	1581.7	104.787	< 2.2e-16 ***
grade	1	1610.2	133.337	< 2.2e-16 ***
sympt	1	1516.2	39.339	3.562e-10 ***
type:trt	2	1555.7	80.801	< 2.2e-16 ***
stade:trt	6	1537.5	70.569	3.126e-13 ***

```
---
```

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

```
>
```

```
> mytest=cox.zph(mycox)
```

```
> print(mytest)
```

	rho	chisq	p
age	0.048153	0.510921	0.475
type2	-0.020514	0.094811	0.758
stade2	0.007726	0.011963	0.913
stade3	-0.005649	0.006762	0.934
stade4	-0.007326	0.013876	0.906
grade2	-0.061648	0.788570	0.375
sympt1	-0.059204	0.817947	0.366
trt2	0.023909	0.124367	0.724
trt3	-0.007130	0.011696	0.914
type2:trt2	0.000727	0.000114	0.991
type2:trt3	0.022004	0.097755	0.755
stade2:trt2	-0.069882	1.005280	0.316
stade3:trt2	-0.048548	0.487417	0.485
stade4:trt2	-0.009787	0.019321	0.889
stade2:trt3	-0.025102	0.131860	0.717
stade3:trt3	-0.032901	0.237530	0.626
stade4:trt3	0.004404	0.004012	0.949
GLOBAL	NA	5.178609	0.997

```
>
```

```
> x11()
```

```
> par(mfrow=c(3,6))
```

```

> plot(mytest)
>
> res.mart=residuals(mycox,type="martingale")
> x11()
> par(mfrow=c(1,2))
> plot(age,res.mart,xlab="age",ylab="martingale residuals")
> lines(lowess(age,res.mart,iter=0))
> plot(t,res.mart,xlab="observed lifetime",ylab="martingale residuals")
> lines(lowess(t,res.mart,iter=0))
>
> mydfbetas=residuals(mycox,type="dfbetas")
> x11()
> par(mfrow=c(3,6))
> plot(1:n,mydfbetas[,1],xlab="subject",ylab="dfbetas residuals for age")
> lines(lowess(1:n,mydfbetas[,1],iter=0))
> plot(1:n,mydfbetas[,2],xlab="subject",ylab="dfbetas residuals for type=2")
> lines(lowess(1:n,mydfbetas[,2],iter=0))
> plot(1:n,mydfbetas[,3],xlab="subject",ylab="dfbetas residuals for stade=2")
> lines(lowess(1:n,mydfbetas[,3],iter=0))
> plot(1:n,mydfbetas[,4],xlab="subject",ylab="dfbetas residuals for stade=3")
> lines(lowess(1:n,mydfbetas[,4],iter=0))
> plot(1:n,mydfbetas[,5],xlab="subject",ylab="dfbetas residuals for stade=4")
> lines(lowess(1:n,mydfbetas[,5],iter=0))
> plot(1:n,mydfbetas[,6],xlab="subject",ylab="dfbetas residuals for grade=2")
> lines(lowess(1:n,mydfbetas[,6],iter=0))
> plot(1:n,mydfbetas[,7],xlab="subject",ylab="dfbetas residuals for sympt=1")
> lines(lowess(1:n,mydfbetas[,7],iter=0))
> plot(1:n,mydfbetas[,8],xlab="subject",ylab="dfbetas residuals for trt=2")
> lines(lowess(1:n,mydfbetas[,8],iter=0))
> plot(1:n,mydfbetas[,9],xlab="subject",ylab="dfbetas residuals for trt=3")
> lines(lowess(1:n,mydfbetas[,9],iter=0))
> plot(1:n,mydfbetas[,10],xlab="subject",ylab="dfbetas residuals for type=2, trt=2")
> lines(lowess(1:n,mydfbetas[,10],iter=0))
> plot(1:n,mydfbetas[,11],xlab="subject",ylab="dfbetas residuals for type=3, trt=3")
> lines(lowess(1:n,mydfbetas[,11],iter=0))
> plot(1:n,mydfbetas[,12],xlab="subject",ylab="dfbetas residuals for stade=2, trt=2")
> lines(lowess(1:n,mydfbetas[,12],iter=0))
> plot(1:n,mydfbetas[,13],xlab="subject",ylab="dfbetas residuals for stade=3, trt=2")
> lines(lowess(1:n,mydfbetas[,13],iter=0))
> plot(1:n,mydfbetas[,14],xlab="subject",ylab="dfbetas residuals for stade=4, trt=2")
> lines(lowess(1:n,mydfbetas[,14],iter=0))
> plot(1:n,mydfbetas[,15],xlab="subject",ylab="dfbetas residuals for stade=2, trt=3")
> lines(lowess(1:n,mydfbetas[,15],iter=0))
> plot(1:n,mydfbetas[,16],xlab="subject",ylab="dfbetas residuals for stade=3, trt=3")
> lines(lowess(1:n,mydfbetas[,16],iter=0))
> plot(1:n,mydfbetas[,17],xlab="subject",ylab="dfbetas residuals for stade=4, trt=3")
> lines(lowess(1:n,mydfbetas[,17],iter=0))

```

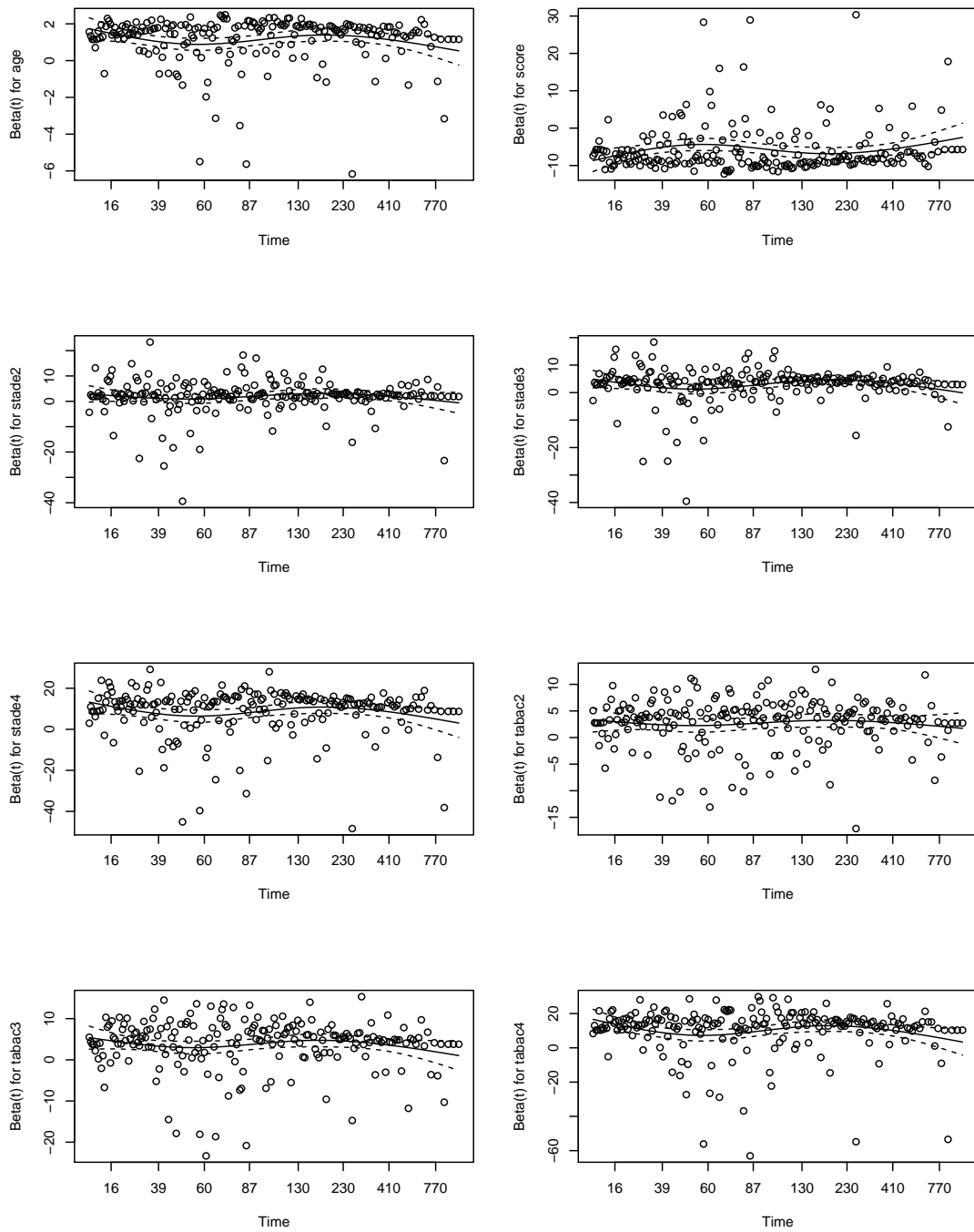


Figure 6 – Tracé des résidus de Schoenfeld en fonction du temps, cancer du poumon

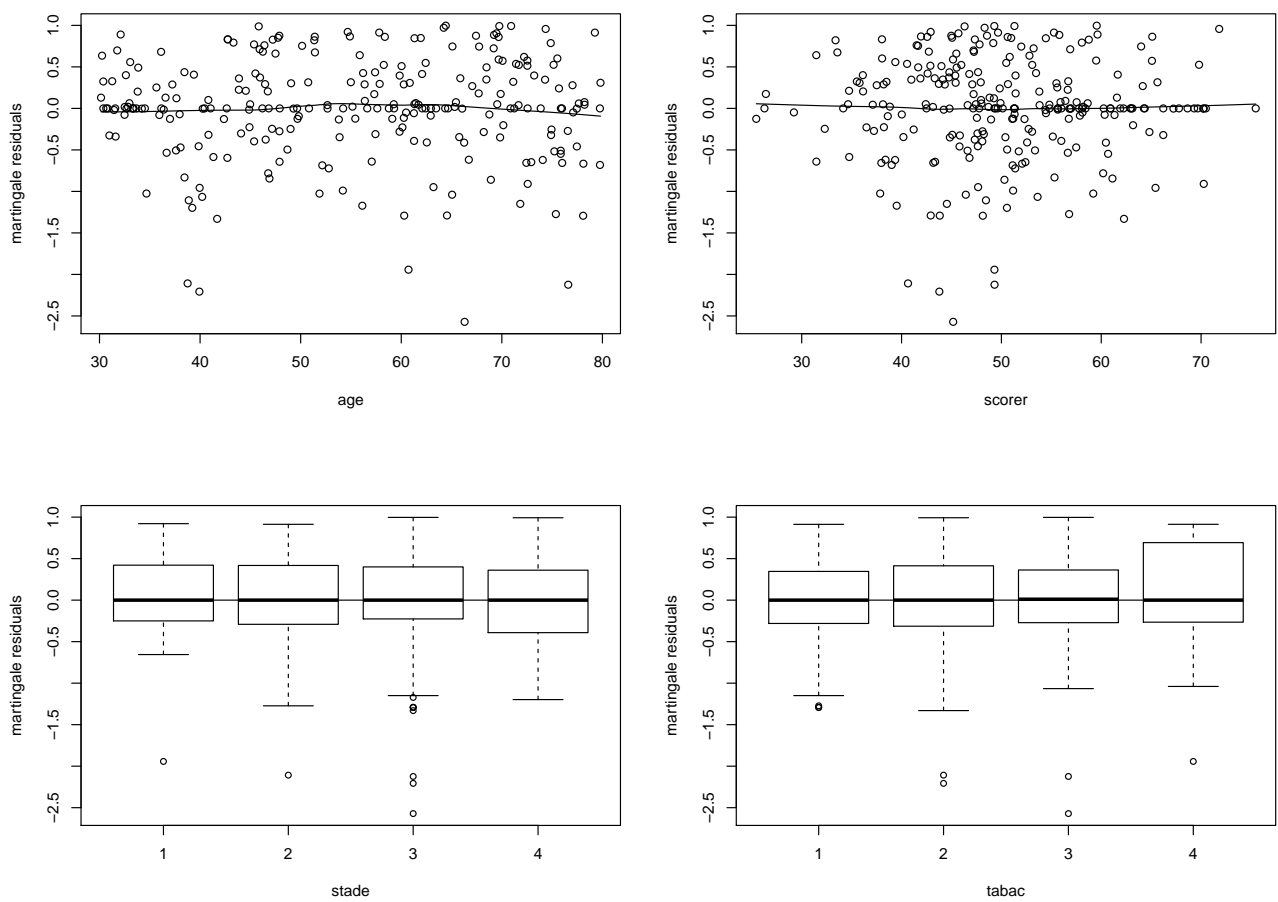


Figure 7 – Tracé des résidus martingale en fonction des prédicteurs, cancer du poumon

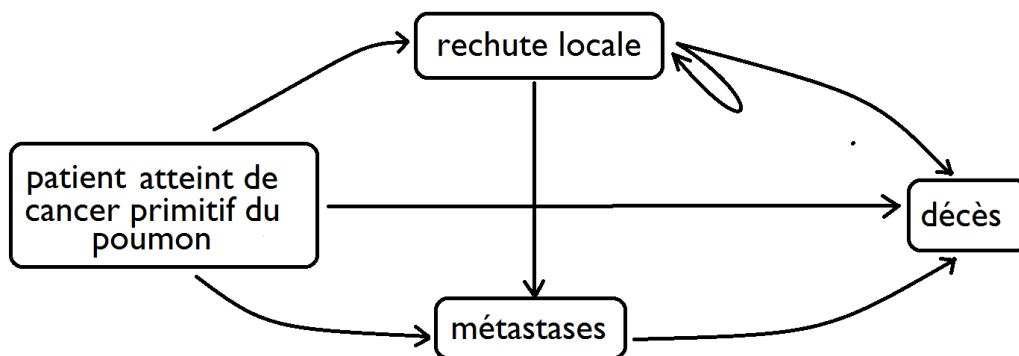


Figure 8 – Evolution simplifiée du cancer primitif du poumon

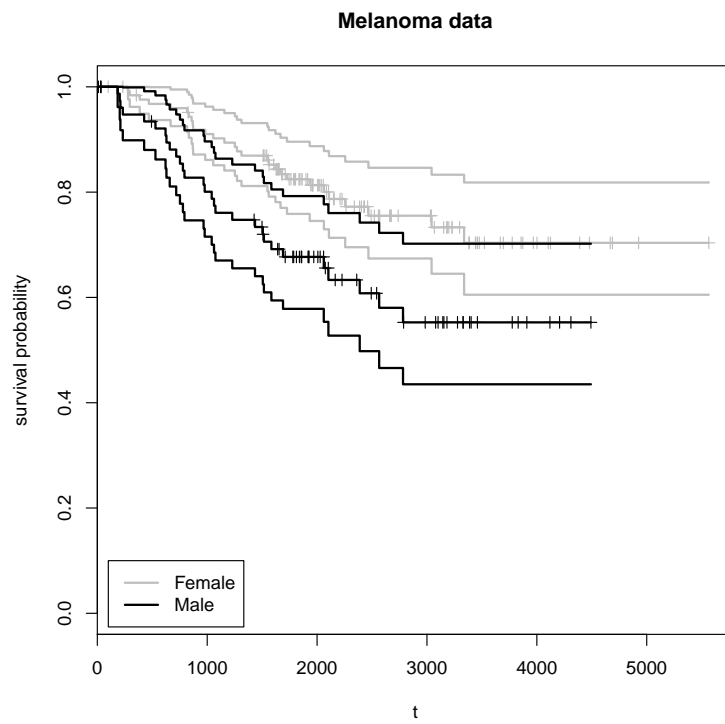


Figure 9 – Tracé des courbes de survie estimées par sexe et des intervalles de confiance ponctuels associés, jeu de données `Melanoma`

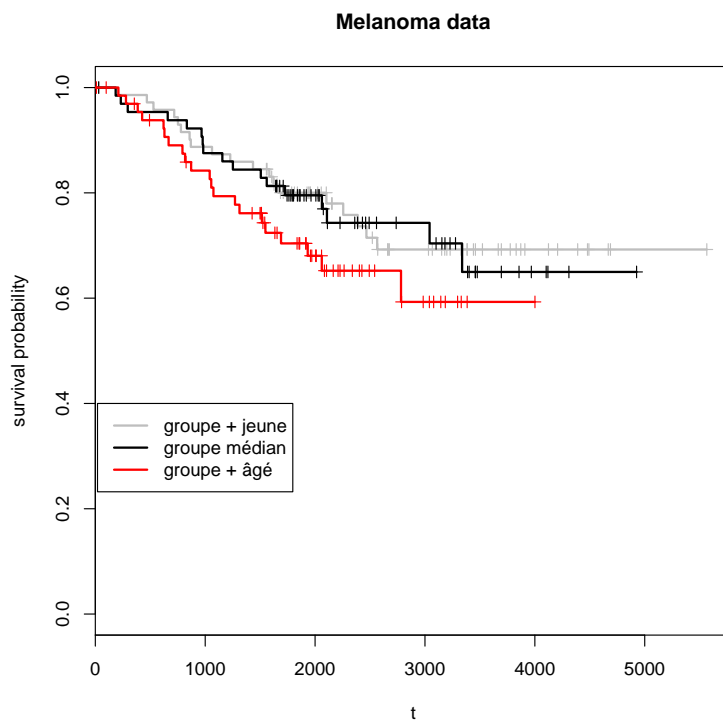


Figure 10 – Tracé des courbes de survie estimées par niveau d'âge (3 niveaux fournis par les quantiles d'ordre 1/3 et 2/3 des durées réellement observées à savoir: entre 4 et 45 ans, entre 45 et 60 ans et entre 60 et 95 ans), jeu de données Melanoma

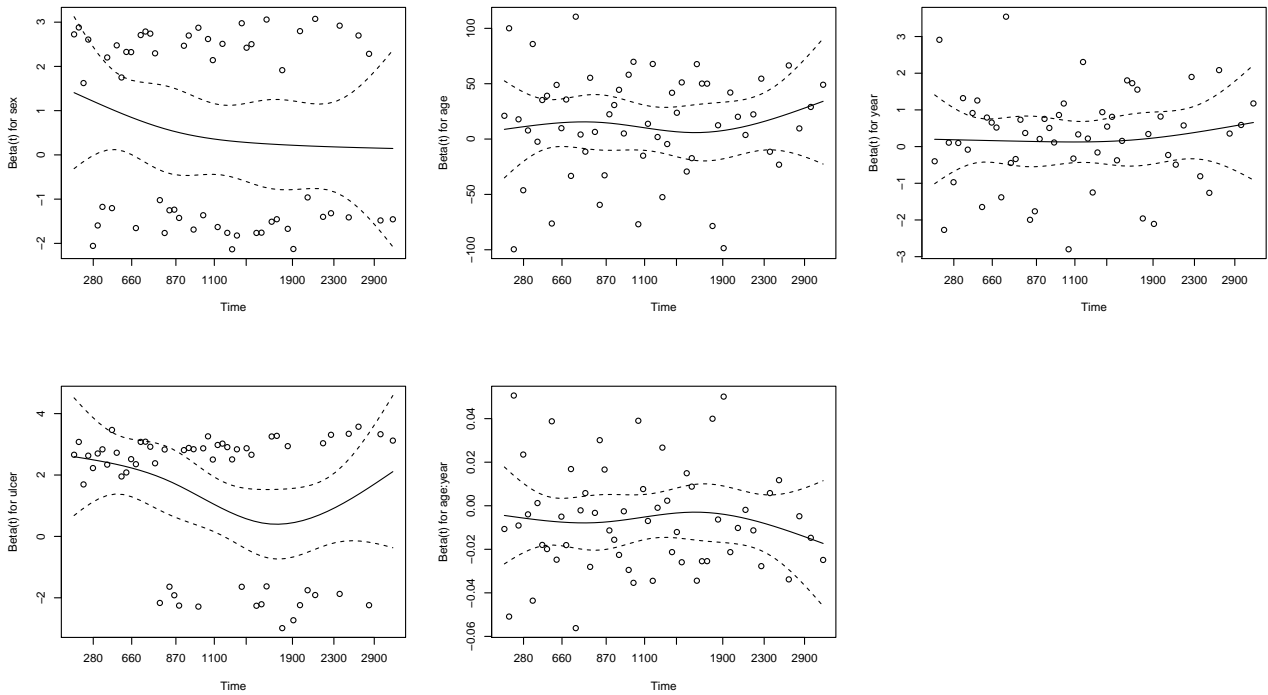


Figure 11 – Tracé des résidus de Schoenfeld en fonction du temps pour le modèle `cox.Melanoma.cov`

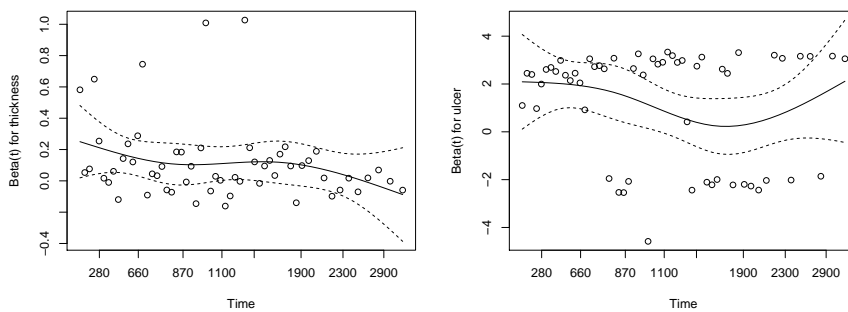


Figure 12 – Tracé des résidus de Schoenfeld en fonction du temps pour le modèle `cox.Melanoma.cov2`

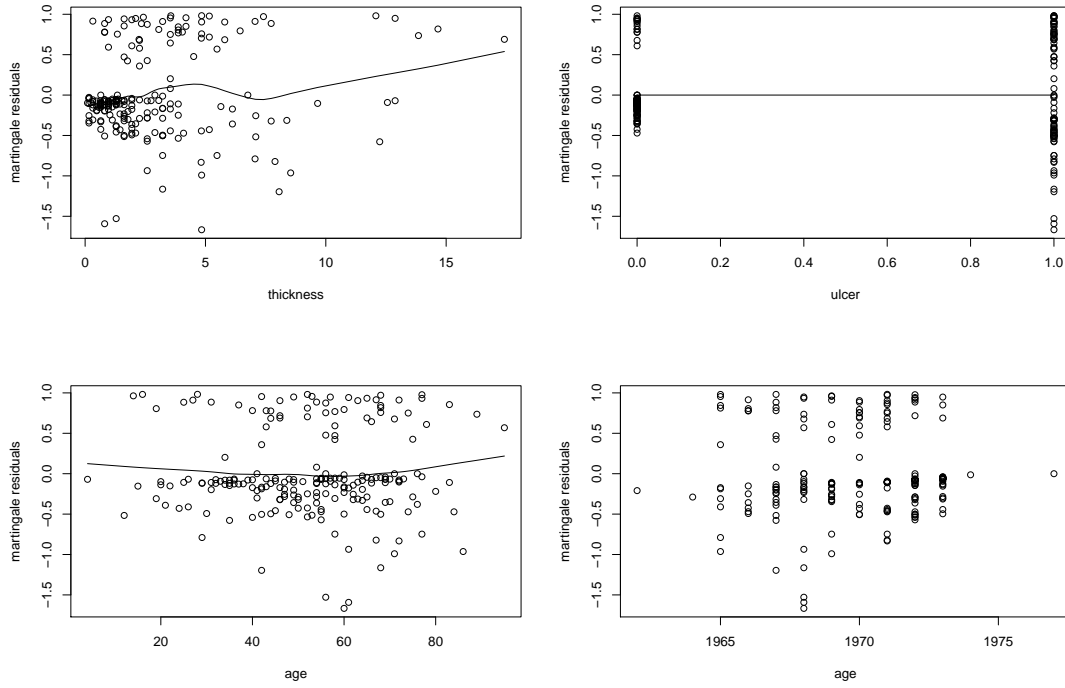


Figure 13 – Tracé des résidus martingale en fonction des prédicteurs pour le modèle `cox.Melanoma.cov`

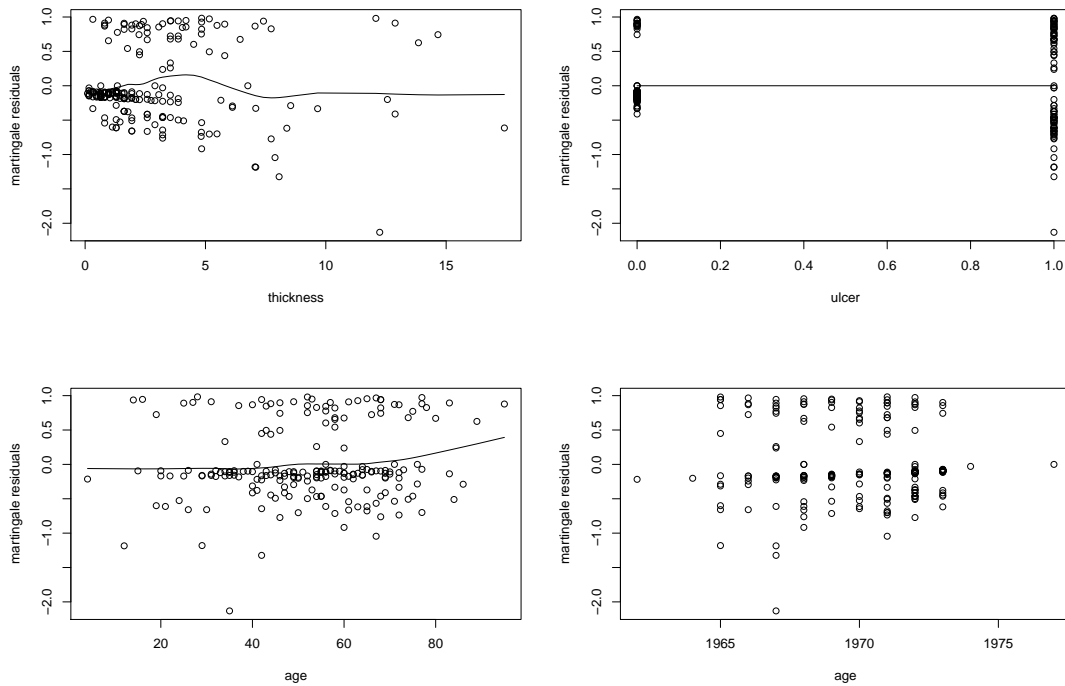


Figure 14 – Tracé des résidus martingale en fonction des prédicteurs pour le modèle `cox.Melanoma.cov2`

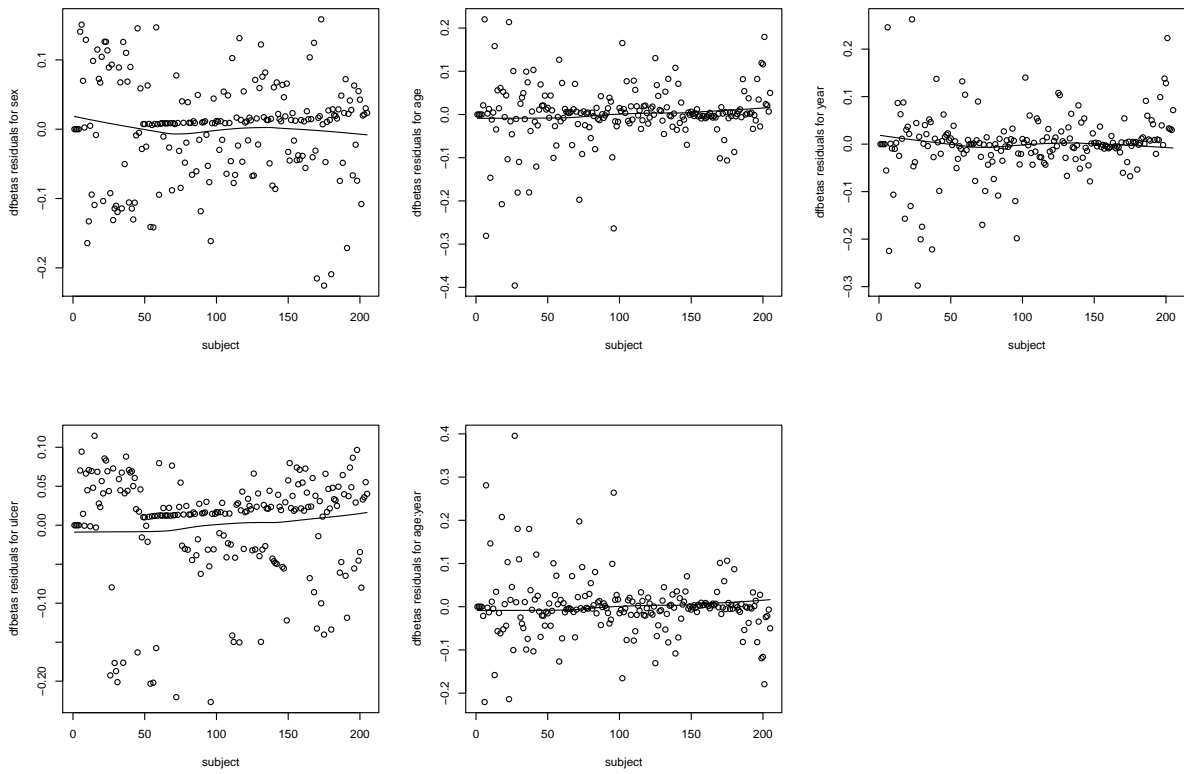


Figure 15 – Tracé des dfbetas en fonction des prédicteurs pour le modèle `cox.Melanoma.cov`

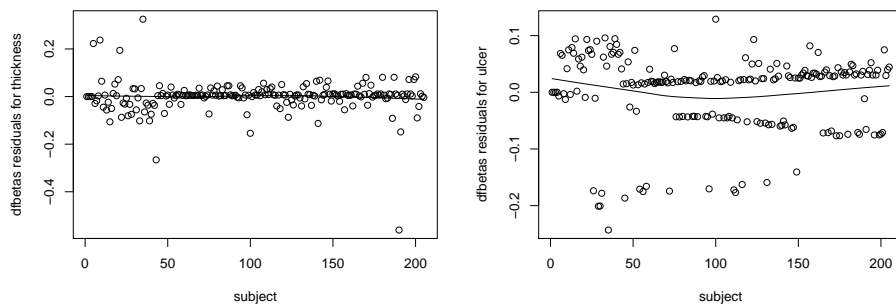


Figure 16 – Tracé des dfbetas en fonction des prédicteurs pour le modèle `cox.Melanoma.cov2`

Estimateur de Kaplan–Meier de la survie et intervalles de confiance

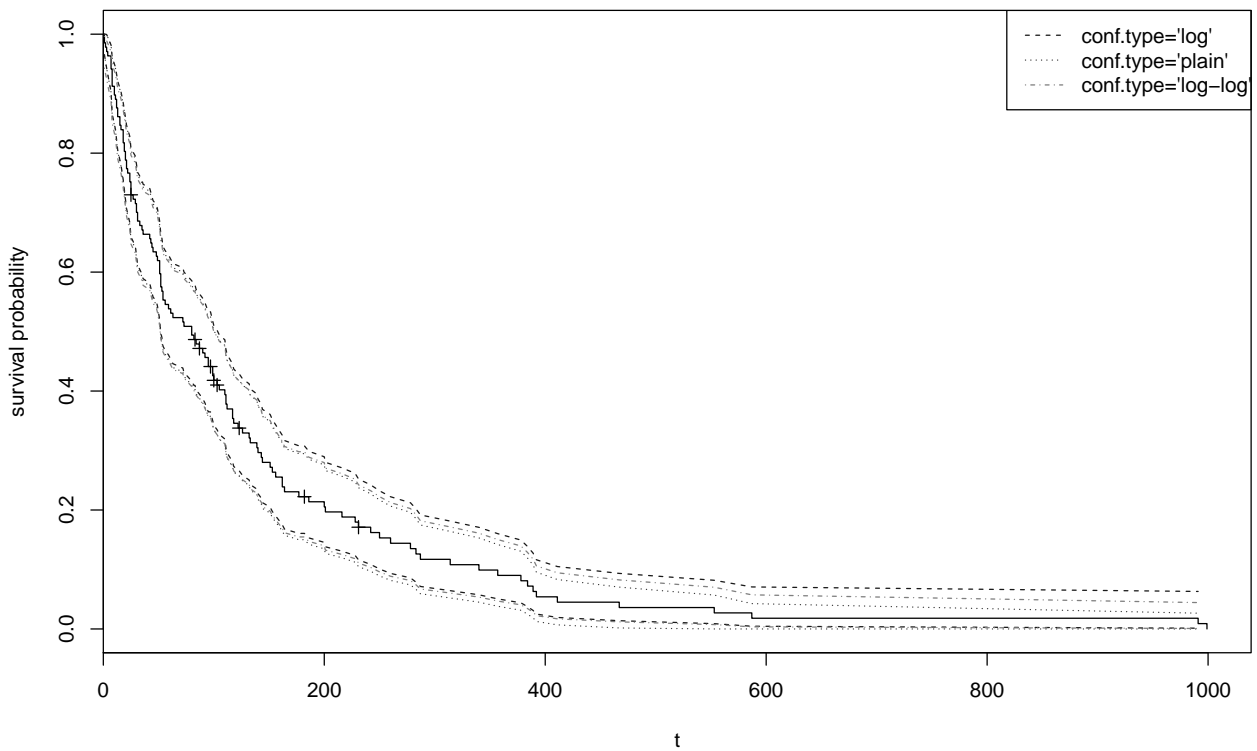


Figure 17 – Tracé des courbes de survie estimées avec différents types d'intervalles de confiance, jeu de données veteran

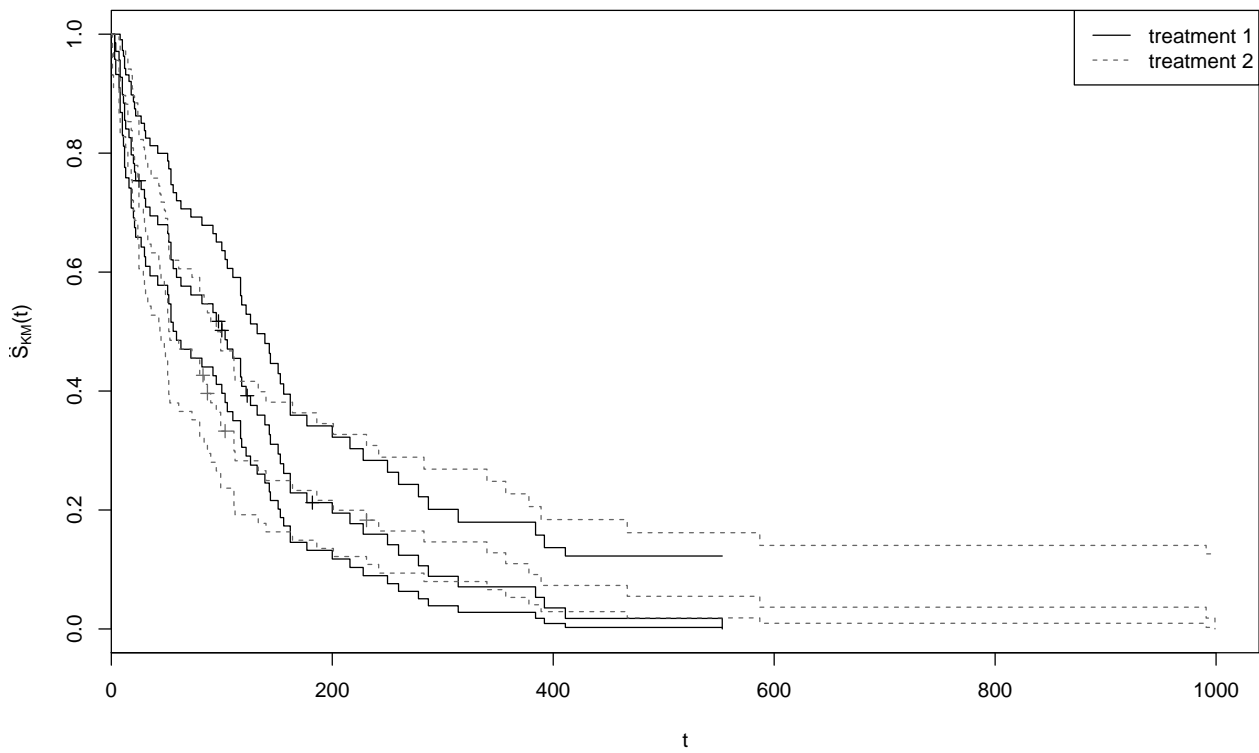


Figure 18 – Tracé des courbes de survie estimées par traitement avec intervalles de confiance, jeu de données veteran

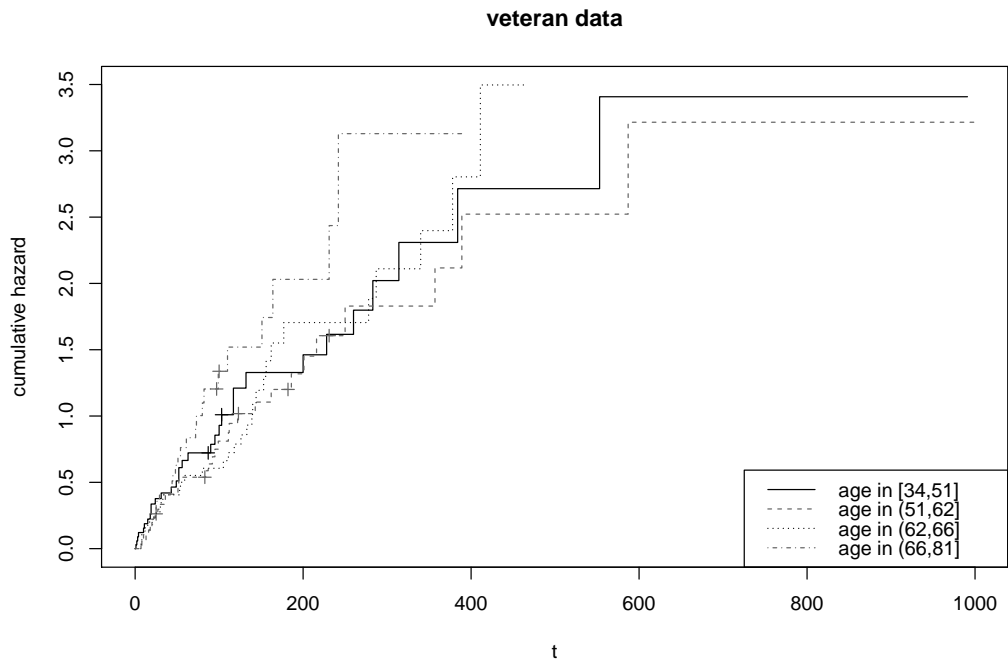


Figure 19 – Tracé des courbes de hasard cumulé estimées par groupe d'âge, jeu de données veteran

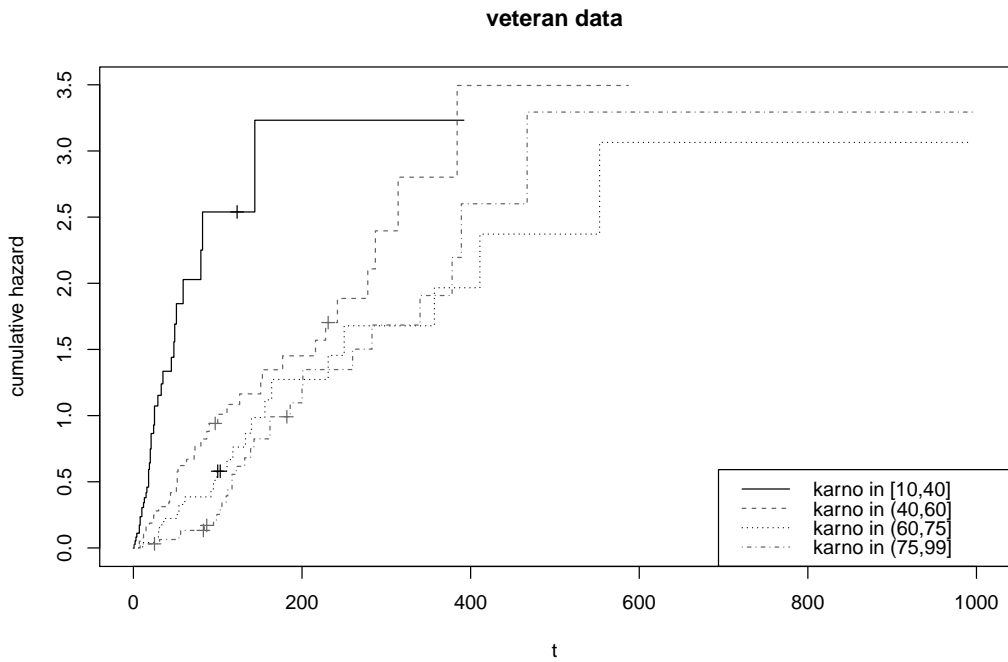


Figure 20 – Tracé des courbes de par groupe de hasard cumulé par score de Karnofsky, jeu de données veteran

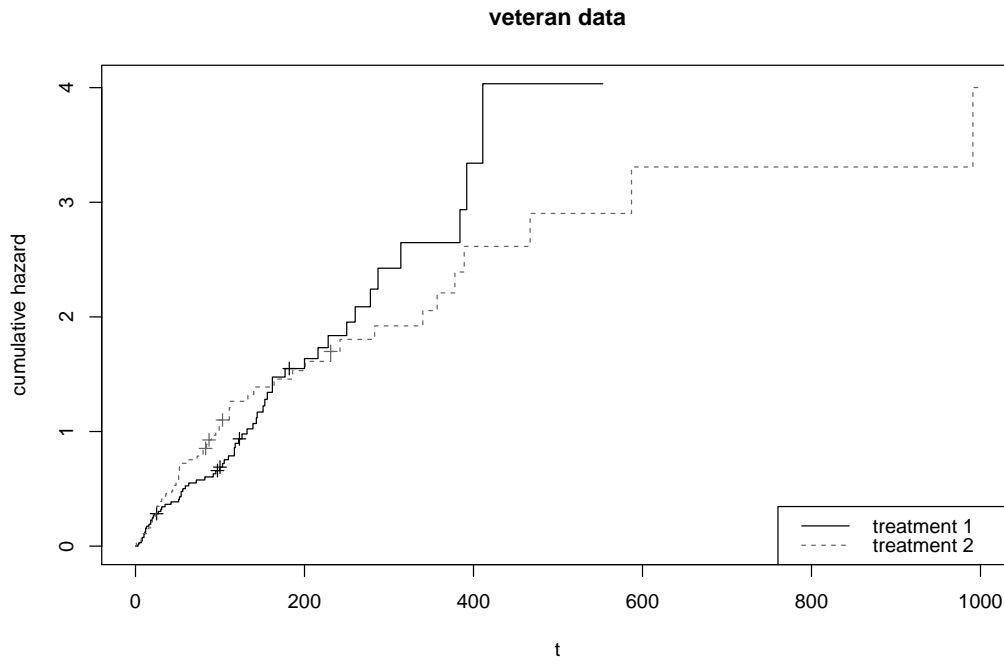


Figure 21 – Tracé des courbes de hasard cumulé estimées par groupe de traitement, jeu de données veteran

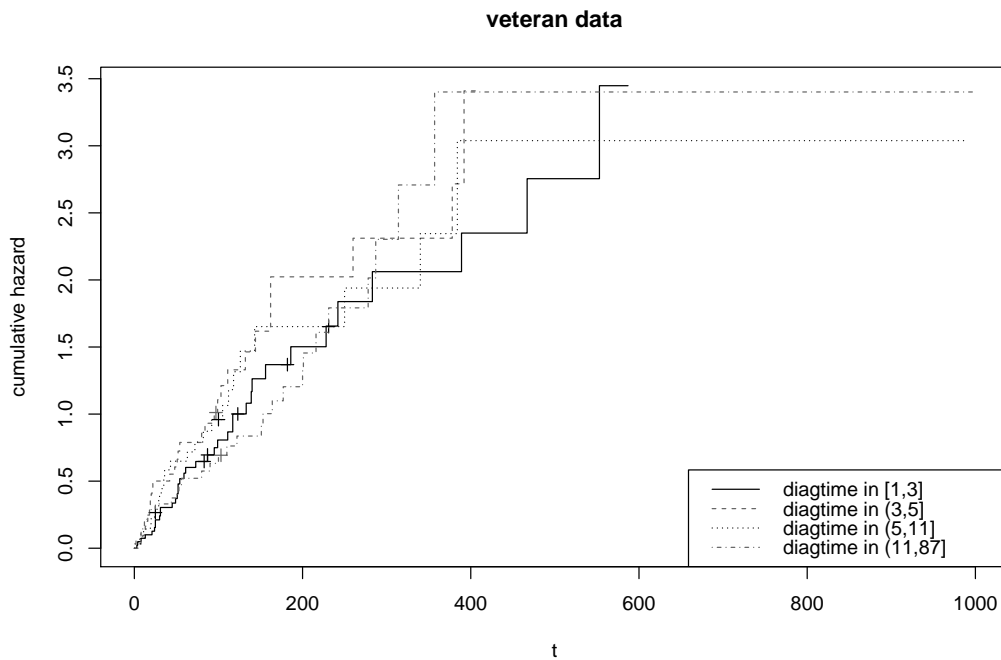


Figure 22 – Tracé des courbes de hasard cumulé estimées par groupe de `diagtime`, jeu de données veteran

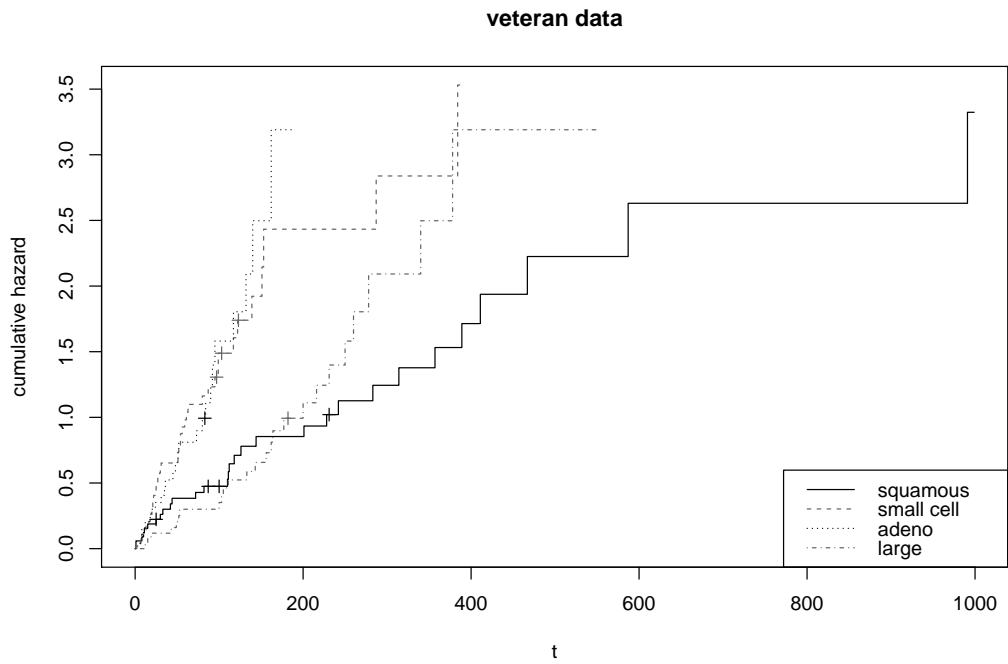


Figure 23 – Tracé des courbes de hasard cumulé estimées par type histologique de tumeur, jeu de données **veteran**

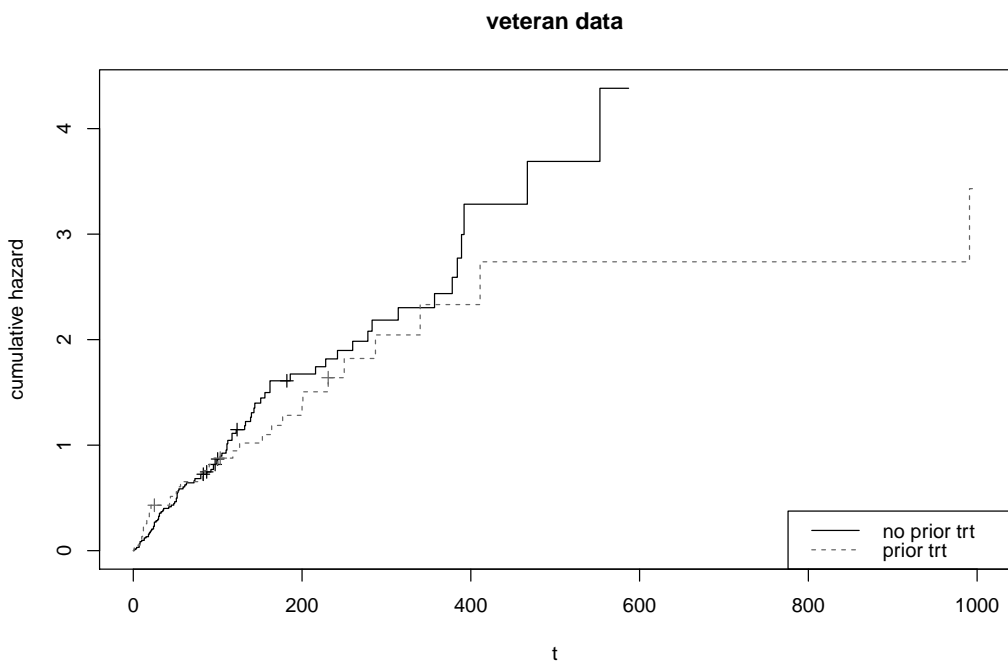


Figure 24 – Tracé des courbes de hasard cumulé estimées selon la prise de traitement antérieur ou non, jeu de données **veteran**

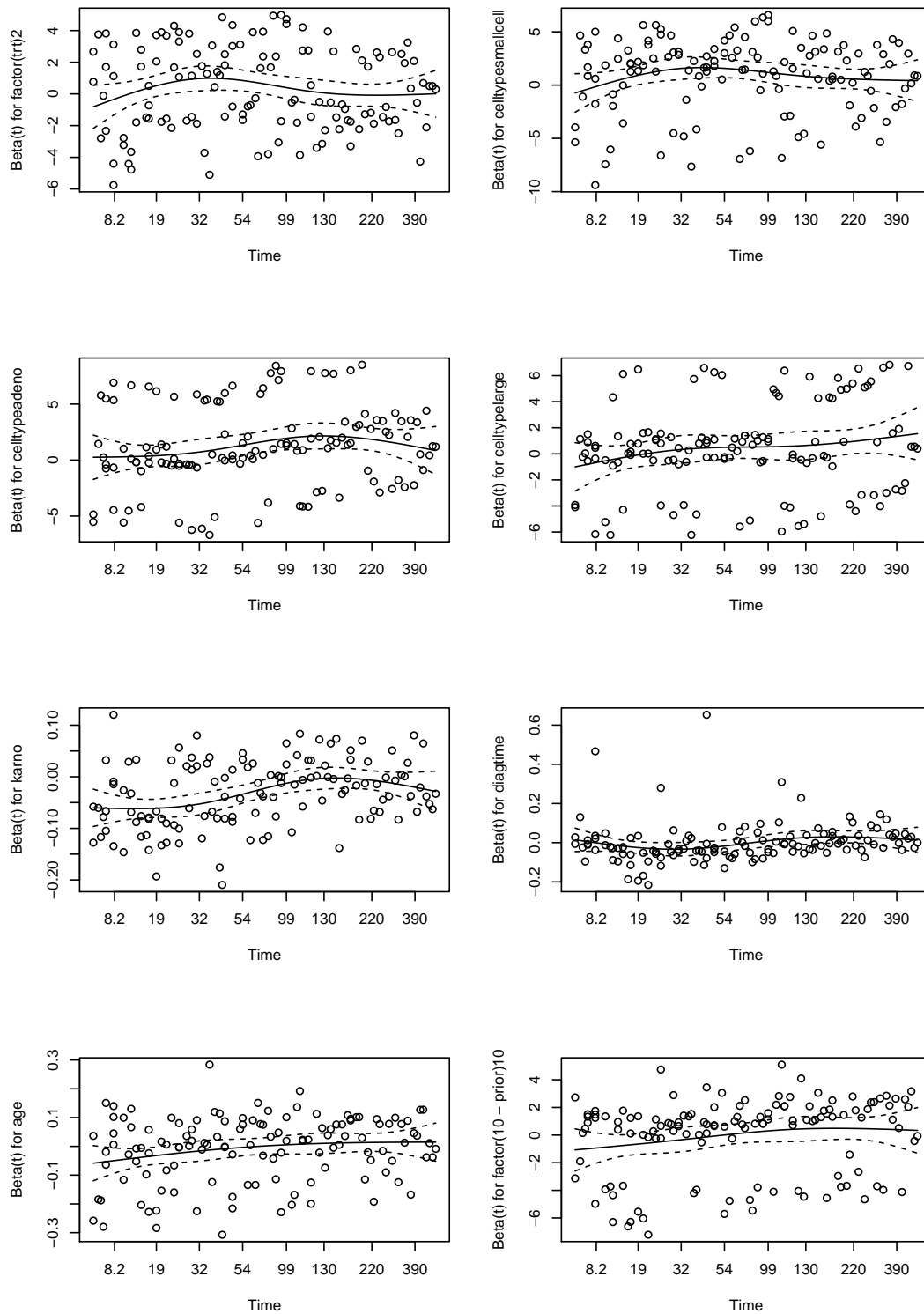


Figure 25 – Tracé des résidus de Schoenfeld normalisés en fonction du temps, jeu de données veteran

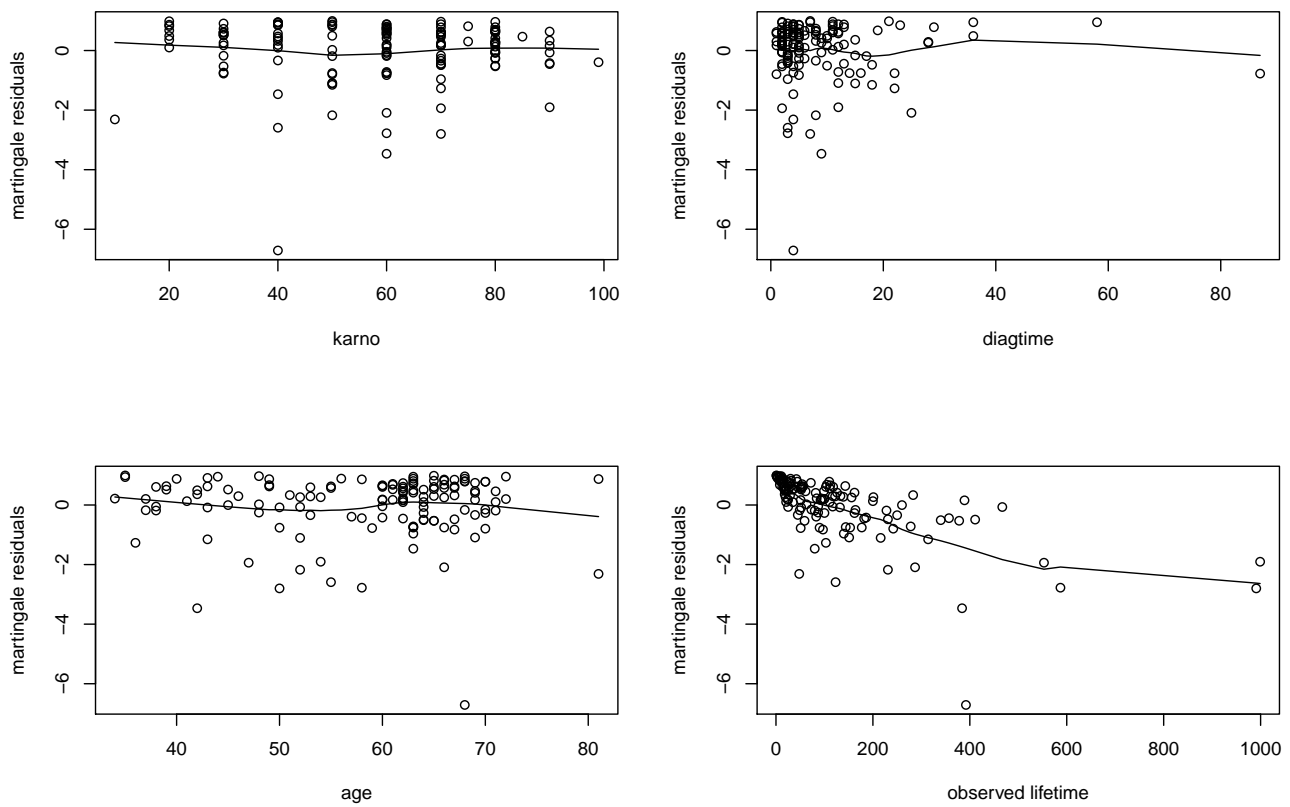


Figure 26 – Tracés des résidus de type martingale, jeu de données *veteran*

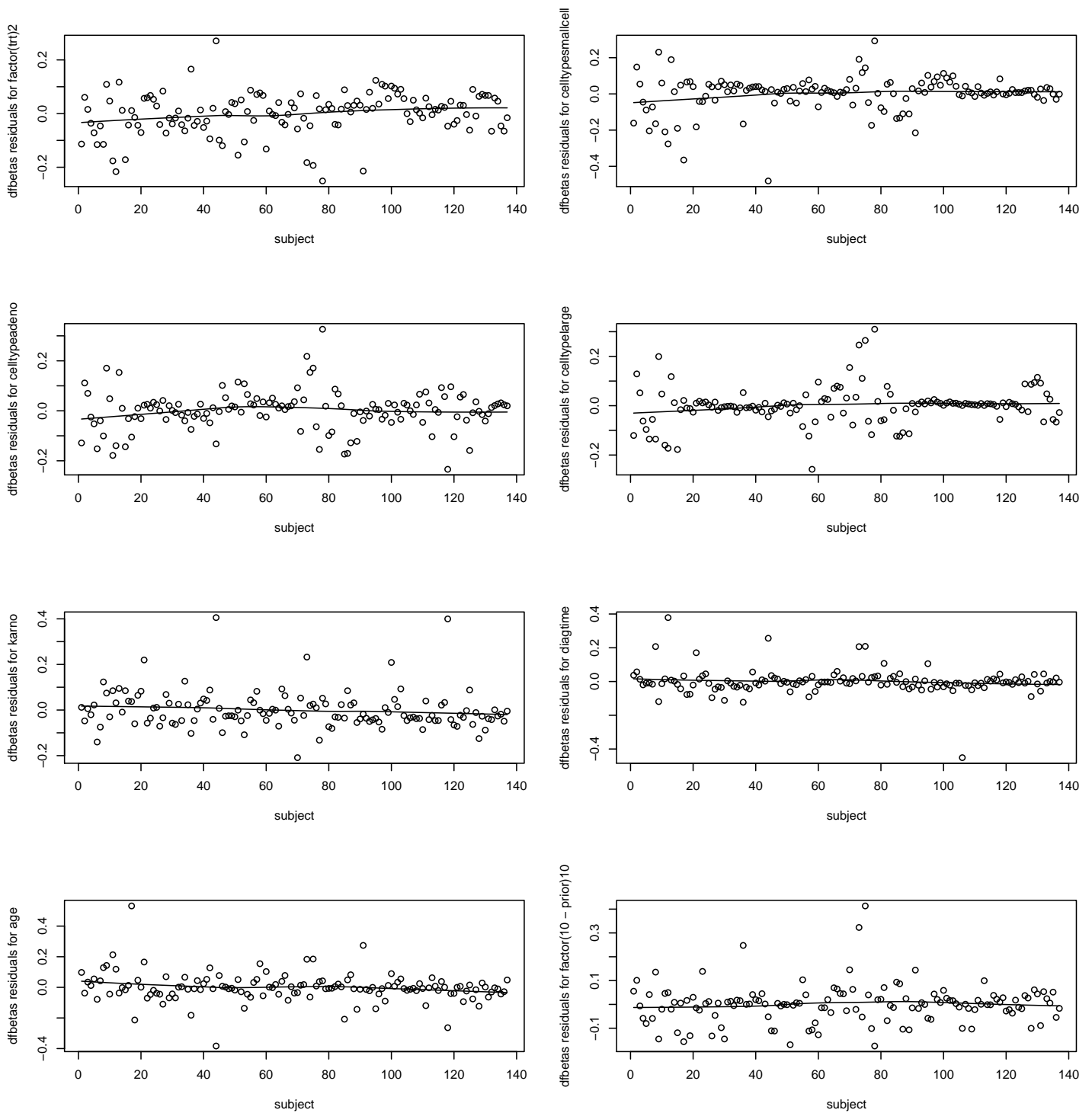


Figure 27 – Tracé des dfbetas par coefficient, jeu de données **veteran**

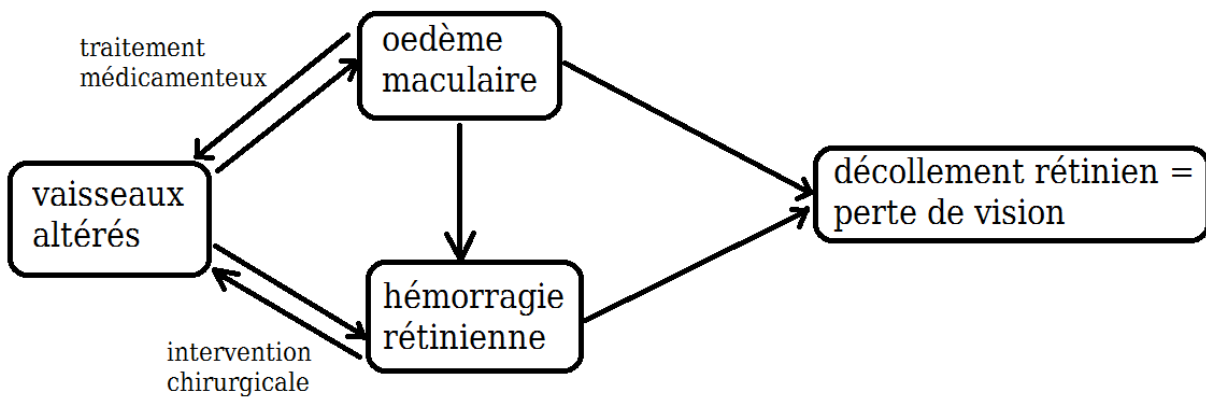


Figure 28 – Evolution simplifiée de la rétinopathie diabétique

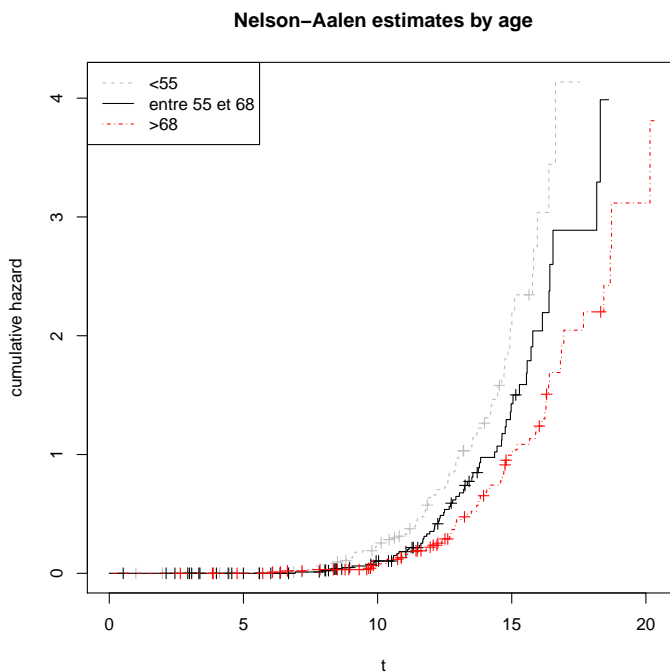


Figure 29 – Tracé des courbes de hasard cumulé estimées par groupe d'âge

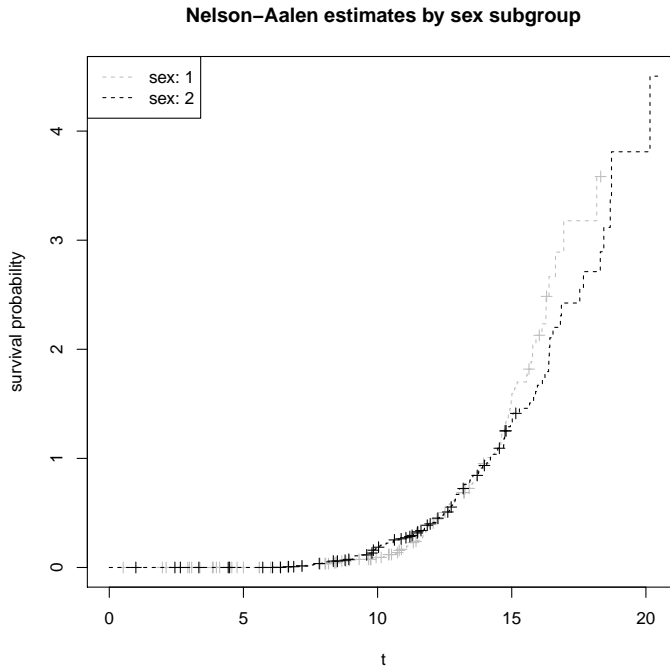


Figure 30 – Tracé des courbes de hasard cumulé estimées par sexe

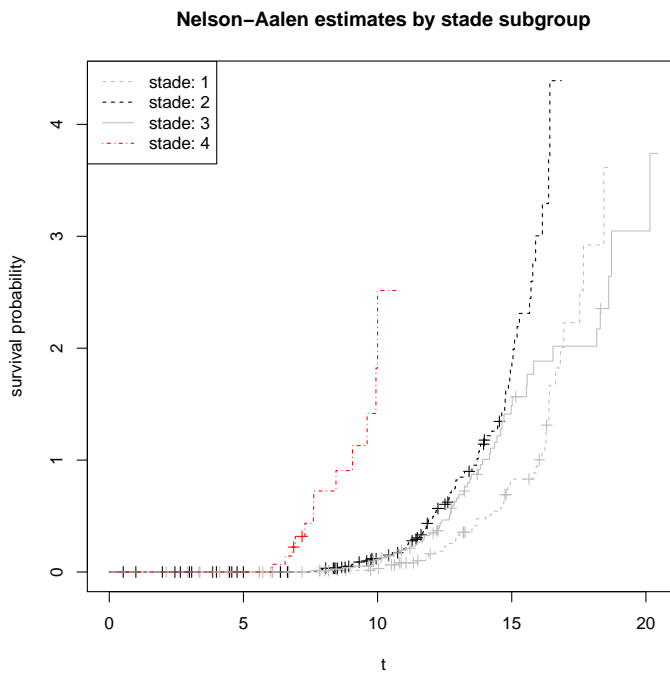


Figure 31 – Tracé des courbes de hasard cumulé estimées par stade

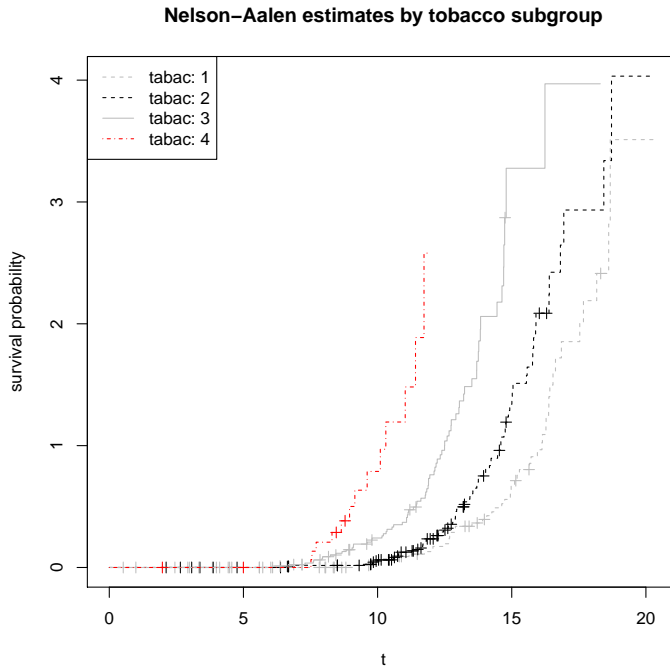


Figure 32 – Tracé des courbes de hasard cumulé estimées par degré de tabagisme

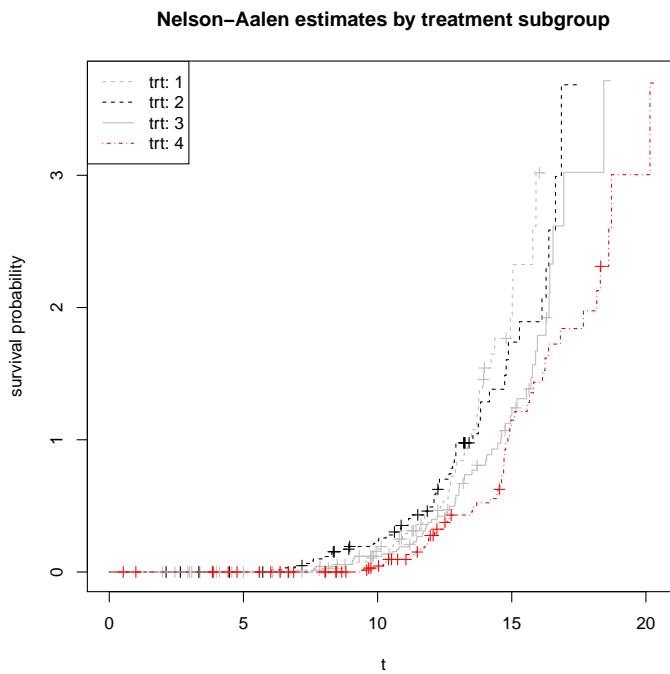


Figure 33 – Tracé des courbes de hasard cumulé estimées par traitement

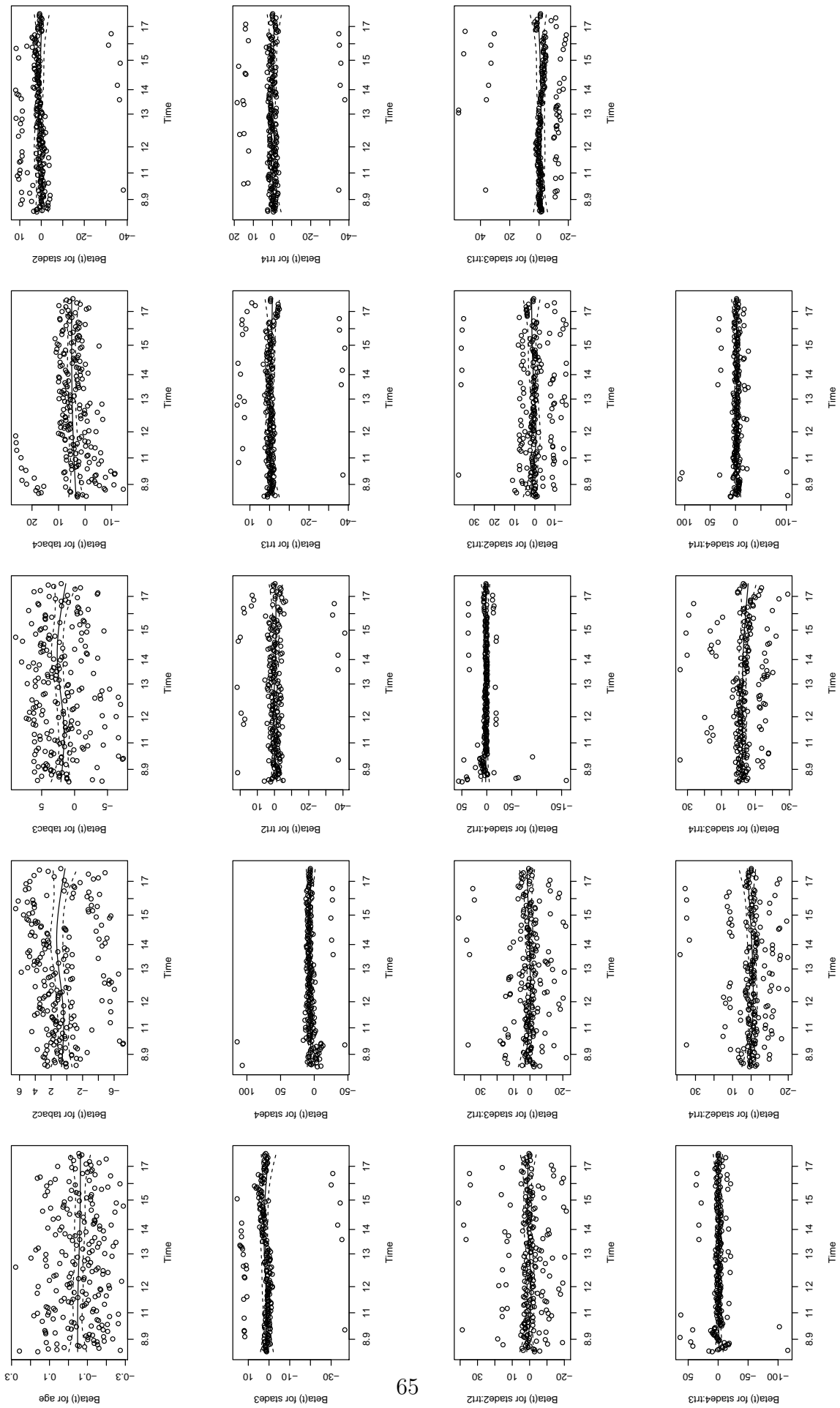


Figure 34 – Tracé des résidus de Schoenfeld normalisés en fonction du temps

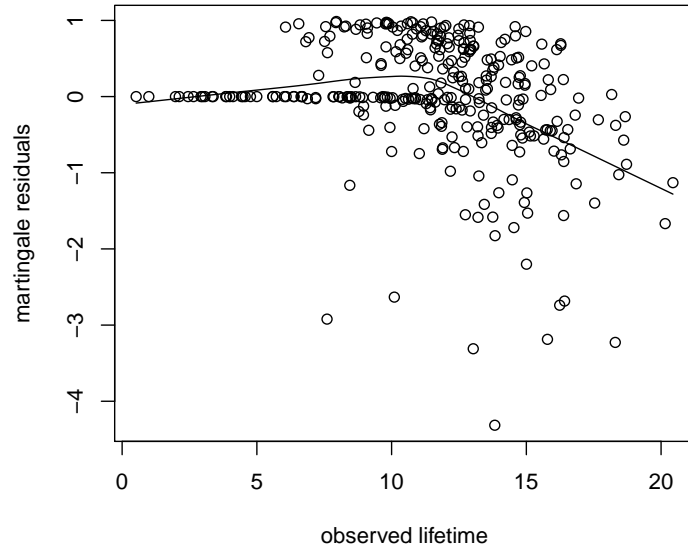
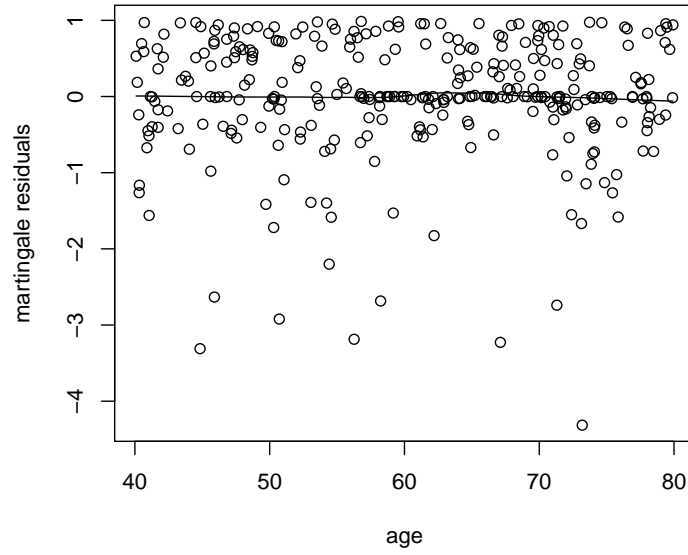


Figure 35 – Tracés des résidus de type martingale

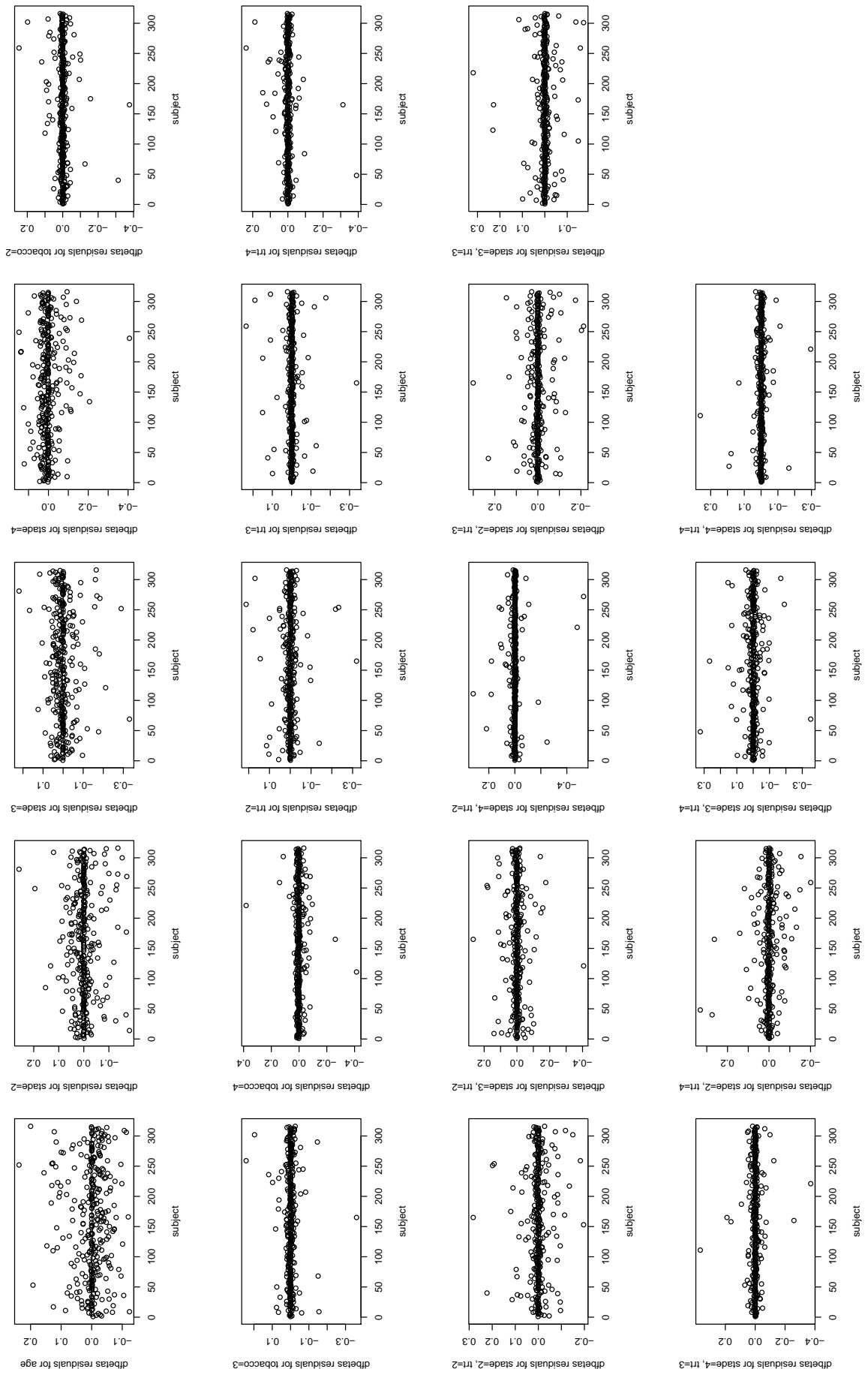


Figure 36 – Tracé des dfbetas par coefficient

n° de patient	observation
1	oedème maculaire au bout de 3 mois résorbé par traitement médicamenteux, hémorragie rétinienne au bout de 1 an, décollement rétinien au bout de 16 mois entrée dans l'étude le 1er mars 2000
2	oedème maculaire au bout de 3 mois, décollement rétinien au bout de 6 mois entrée dans l'étude le 1er mars 2000
3	oedème maculaire au bout de 6 mois résorbé par traitement médicamenteux, oedème maculaire au bout de 10 mois, décollement rétinien au bout de 16 mois entrée dans l'étude le 1er mars 2000
4	rien ne se produit pendant l'étude entrée dans l'étude le 1er juin 2000
5	déménagement pour mutation du conjoint au bout de 2 ans entrée dans l'étude le 1er mars 2000
6	oedème maculaire au bout de 6 mois, décollement rétinien au bout de 24 mois entrée dans l'étude le 1er mars 2000
7	oedème maculaire au bout de 6 mois, décollement rétinien au bout de 24 mois entrée dans l'étude le 1er mars 2000
8	perdu de vue au bout de 2 ans entrée dans l'étude le 1er mars 2000
9	oedème maculaire au bout de 18 mois, perdu de vue au bout de 3 ans entrée dans l'étude le 1er mars 2000
10	oedème maculaire au bout de 18 mois, décollement rétinien au bout de 24 mois entrée dans l'étude le 1er mars 2000
11	oedème maculaire au bout de 18 mois, décollement rétinien au bout de 21 mois entrée dans l'étude le 1er mars 2000
12	hémorragie rétinienne au bout de 1 an traitée avec succès, hémorragie rétinienne au bout de 18 mois, décollement rétinien au bout de 2,5 ans entrée dans l'étude le 1er mars 2000
13	hémorragie rétinienne au bout de 1 an traitée avec succès, hémorragie rétinienne au bout de 20 mois, décollement rétinien au bout de 2 ans entrée dans l'étude le 1er avril 2000
14	hémorragie rétinienne au bout de 1 an, décollement rétinien au bout de 2,5 ans entrée dans l'étude le 1er mars 2000
15	hémorragie rétinienne au bout de 16 mois traitée avec succès, décès accidentel à 3 ans entrée dans l'étude le 1er mars 2000
16	hémorragie rétinienne au bout de 16 mois traitée avec succès entrée dans l'étude le 1er juin 2000
17	hémorragie rétinienne au bout de 16 mois traitée avec succès entrée dans l'étude le 1er juin 2000
18	oedème maculaire au bout de 3 ans résorbé par traitement médicamenteux entrée dans l'étude le 1er juillet 2000
19	oedème maculaire au bout de 3 ans résorbé par traitement médicamenteux entrée dans l'étude le 1er juillet 2000
20	oedème maculaire au bout de 4 mois résorbé par traitement médicamenteux, hémorragie rétinienne au bout de 1 an traitée avec succès, oedème maculaire au bout de 18 mois, décollement rétinien au bout de 24 mois entrée dans l'étude le 1er mars 2000

Table 14 – Données recueillies lors de l'étude sur la rétinopathie diabétique

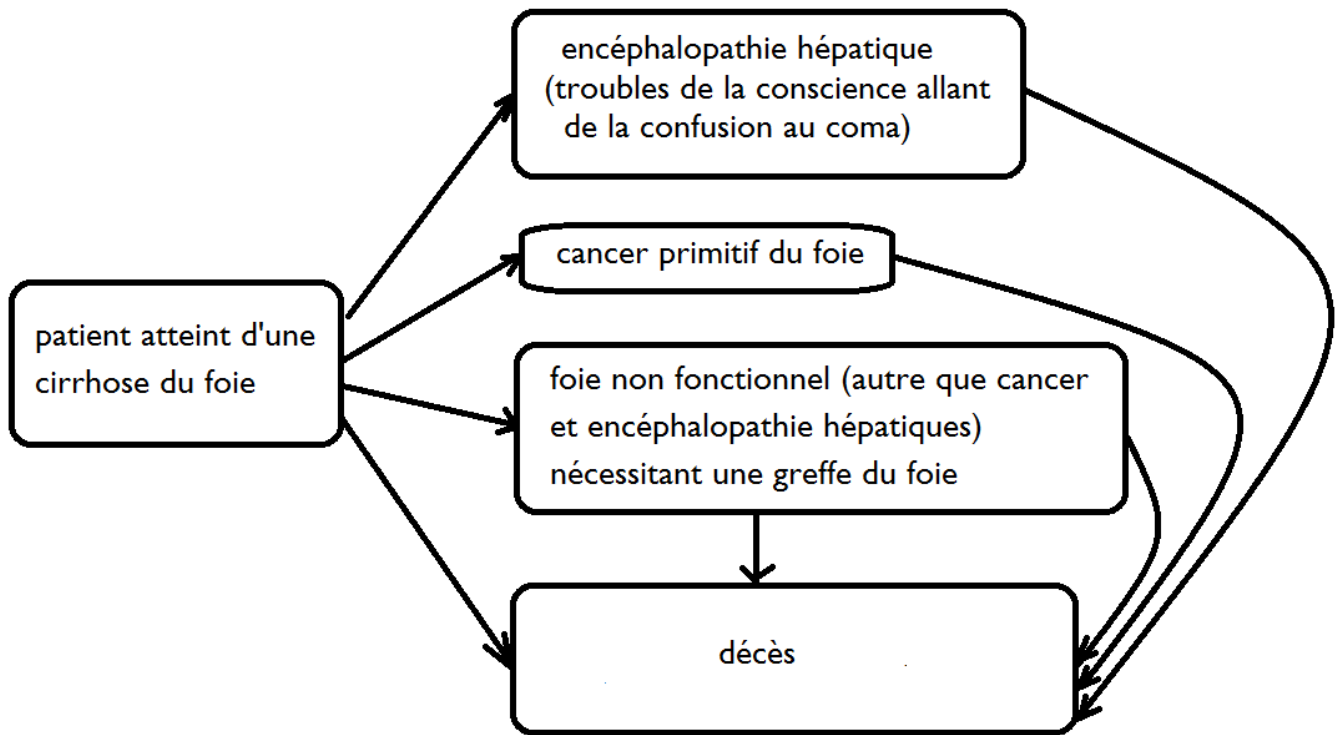


Figure 37 – Evolution simplifiée de la cirrhose du foie

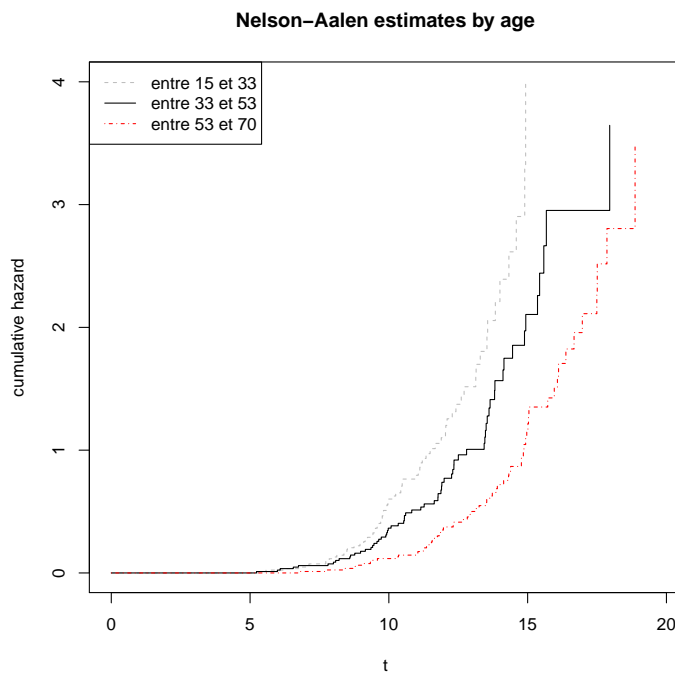


Figure 38 – Tracé des courbes de hazard cumulé estimées par groupe d'âge

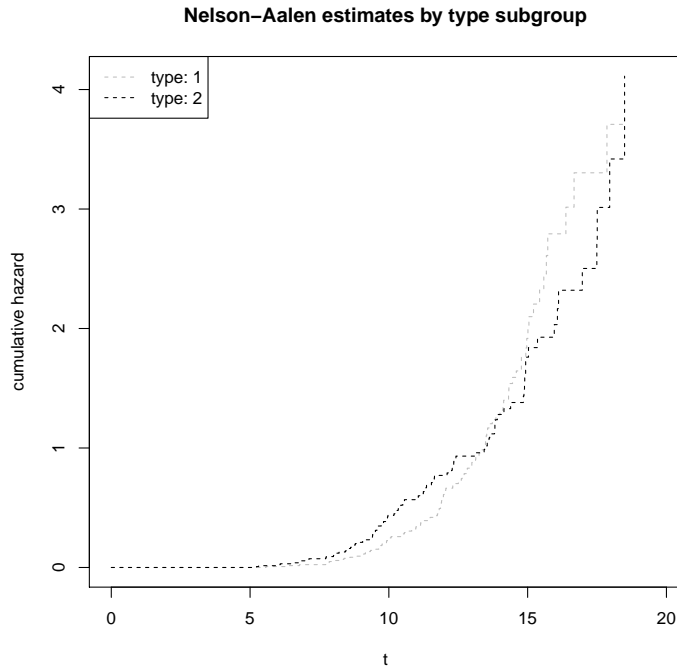


Figure 39 – Tracé des courbes de hasard cumulé estimées par type de LNH

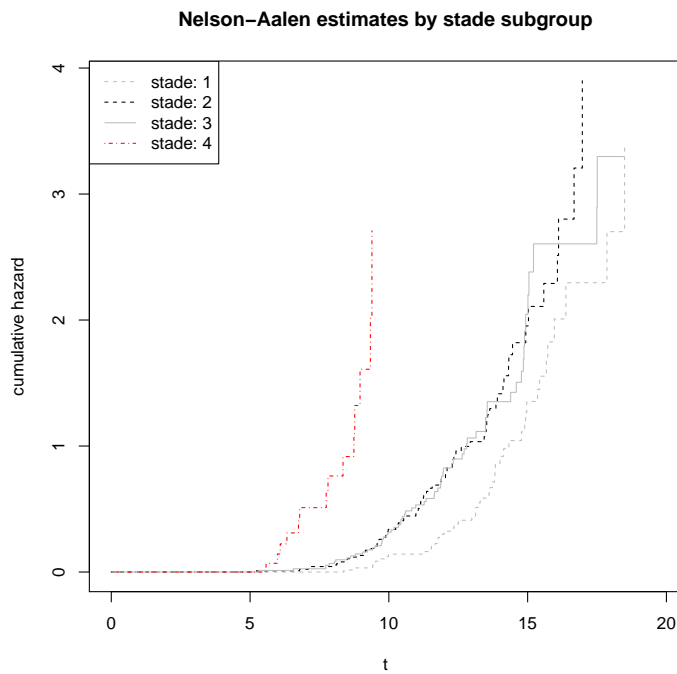


Figure 40 – Tracé des courbes de hasard cumulé estimées par stade

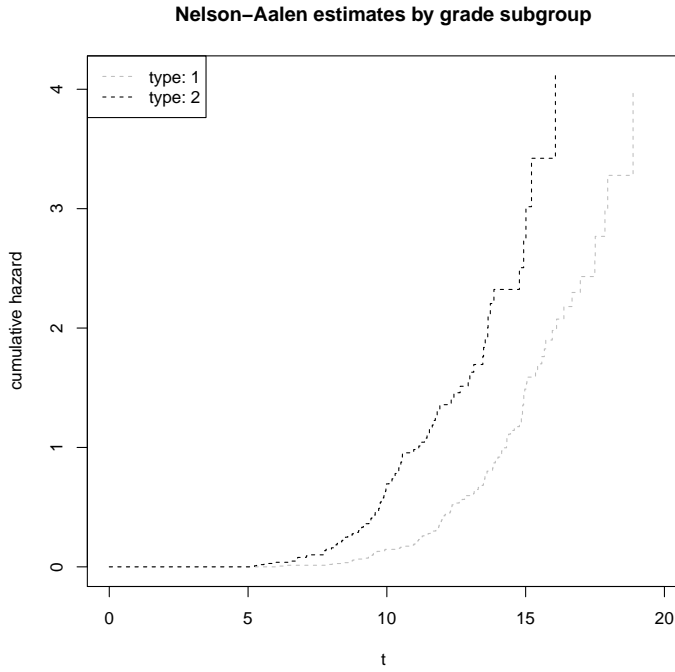


Figure 41 – Tracé des courbes de hasard cumulé estimées par grade

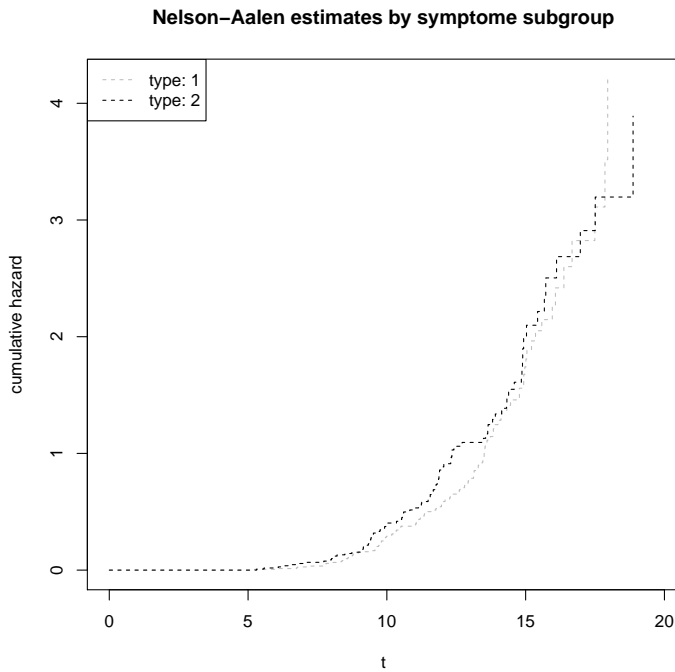


Figure 42 – Tracé des courbes de hasard cumulé estimées selon la présence ou non des symptômes évoqués

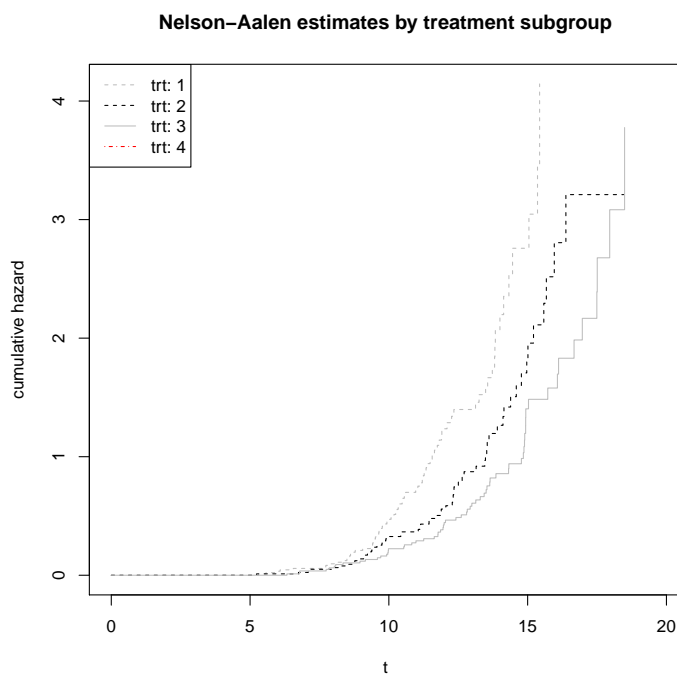


Figure 43 – Tracé des courbes de hasard cumulé estimées par traitement

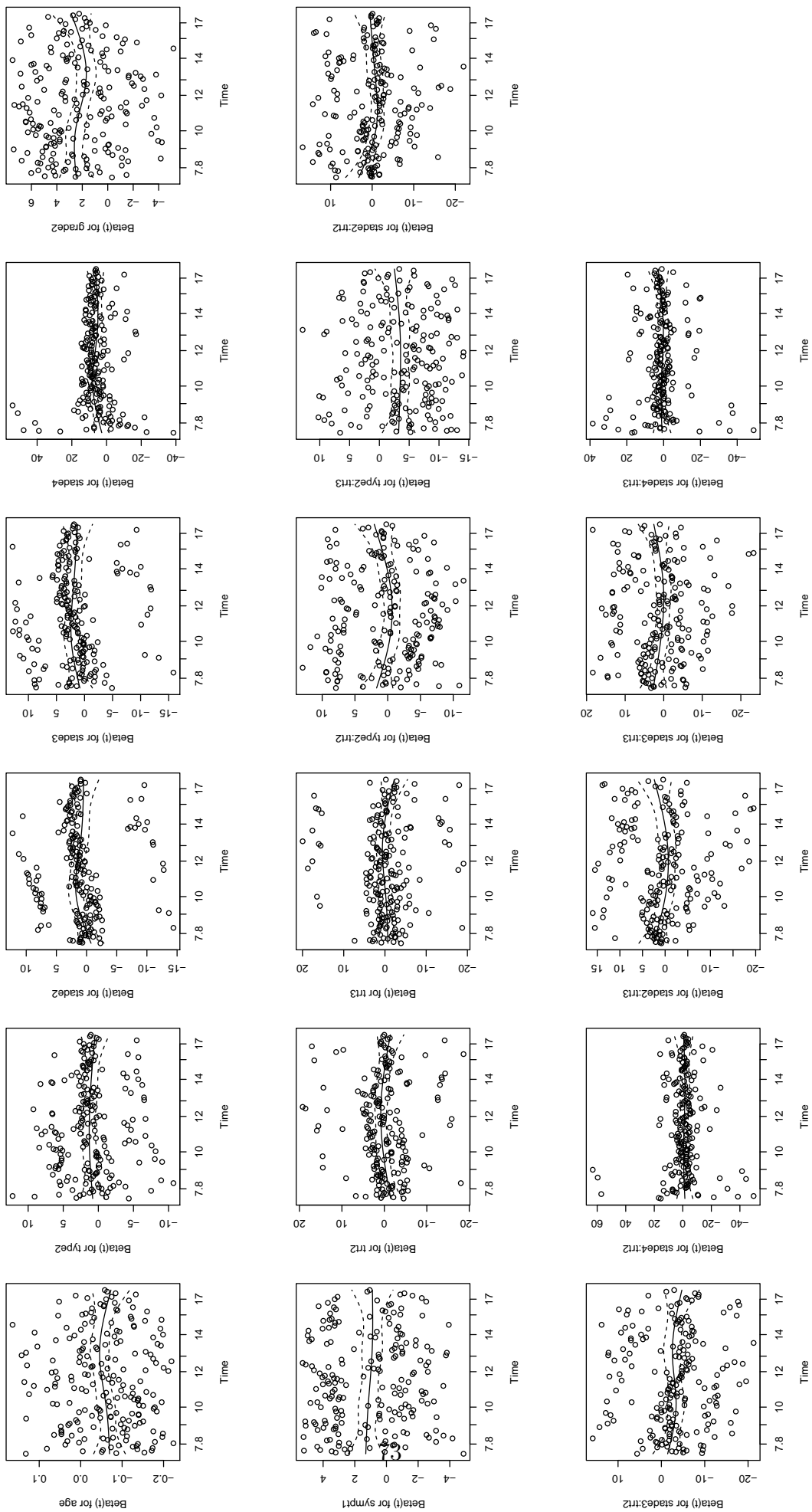


Figure 44 – Tracé des résidus de Schoenfeld normalisés en fonction du temps

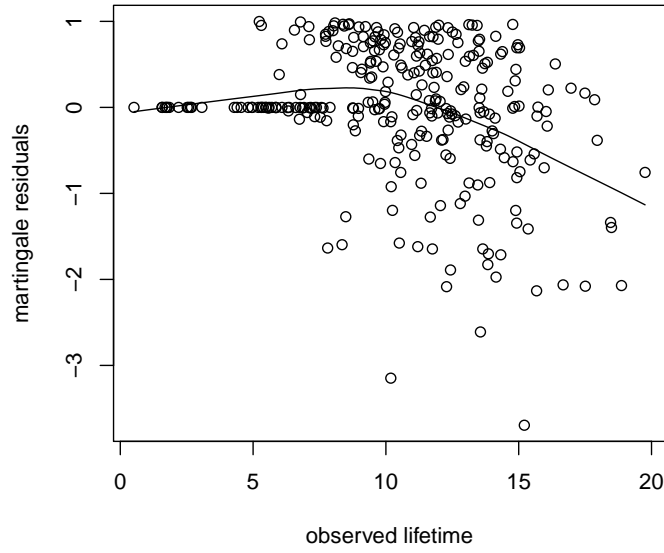
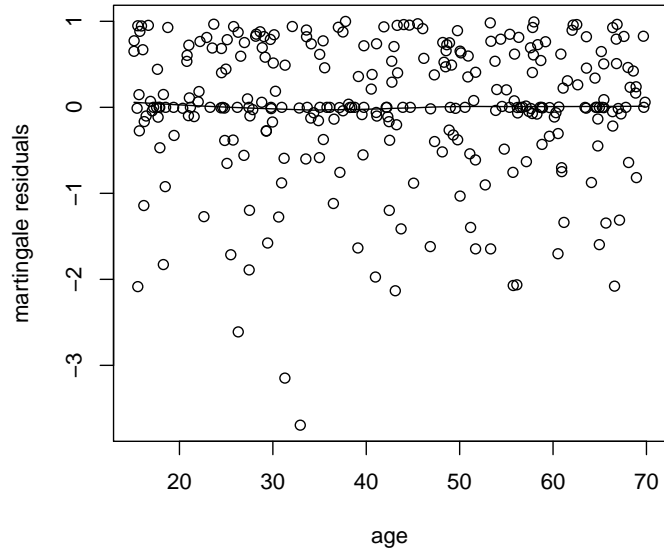


Figure 45 – Tracés des résidus de type martingale

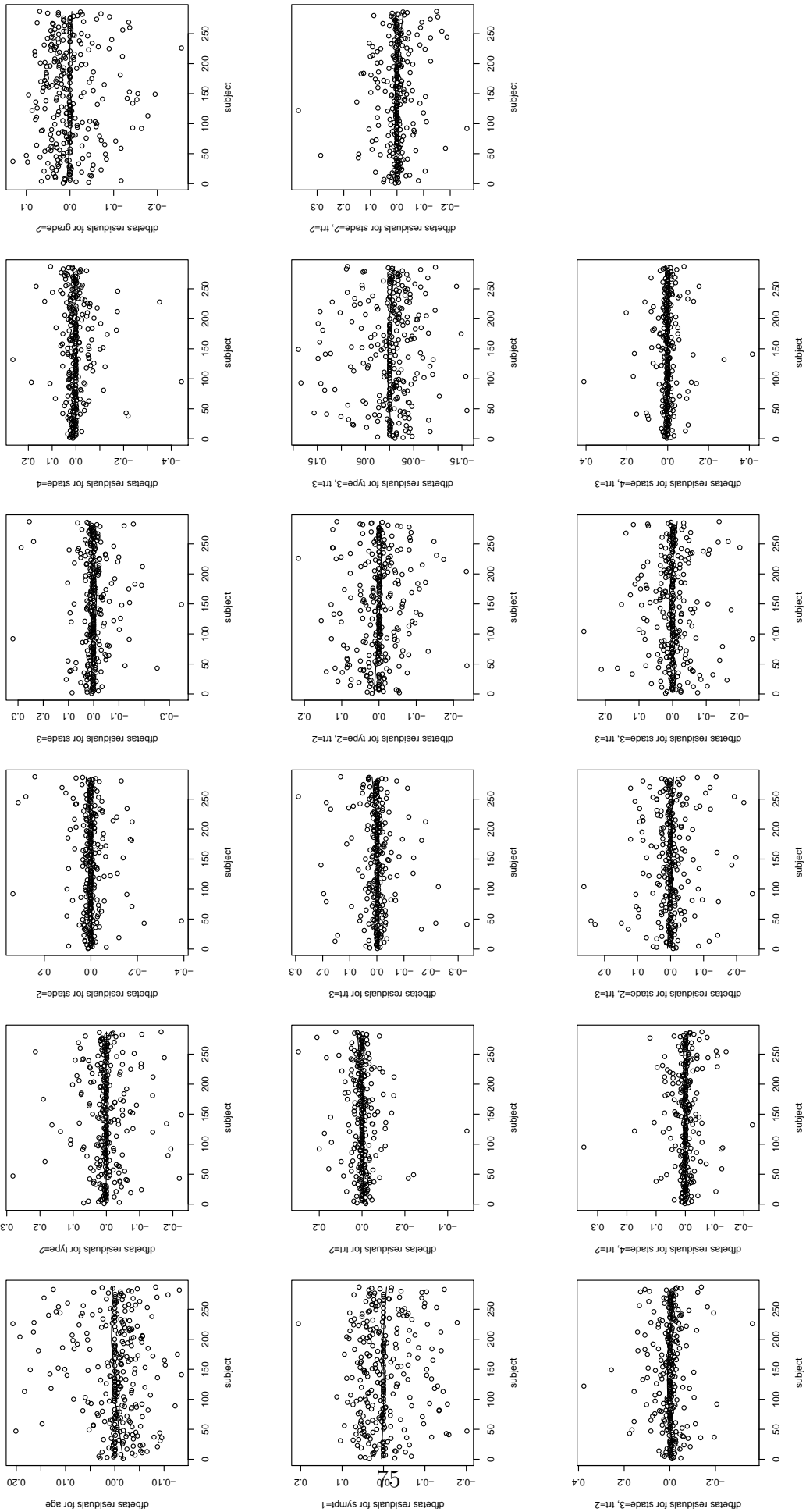


Figure 46 – Tracé des dfbetas

patient	observation	âge	CC	sexe	TL	traitement
1	rechute au bout de 3 ans 6 mois décès au bout de 4 ans	15	à bon risque	H	5000	A
2	RGH aigüe au bout de 6 mois décès au bout de 1 an	62	à bon risque	H	8000	B
3	rechute au bout de 1 an 6 mois décès au bout de 2 ans	51	à risque moyen	H	15000	B
4	rechute au bout de 1 an 6 mois décès au bout de 2 ans	58	à bon risque	F	5000	B
5	décès accidentel au bout de 18 mois	43	à bon risque	H	7000	B
6	rechute au bout de 2 ans décès au bout de 2 ans 10 mois	71	à bon risque	H	5000	A
7	RGH aigüe au bout de 6 mois décès au bout de 1 an	61	à risque moyen	F	8000	A
8	perdue de vue au bout de 2 ans	62	à mauvais risque	F	20000	A
9	RGH chronique au bout de 13 mois rechute au bout de 14 mois décès au bout de 2 ans	57	à risque moyen	F	12000	A
10	décès au bout de 20 mois	56	à mauvais risque	H	12000	B
11	RGH aigüe au bout de 6 mois rechute au bout de 5 ans encore en vie à la fin de l'étude entrée au 01/12/1990	40	à bon risque	H	5000	B
12	rechute au bout de 1 an 6 mois perdu de vue au bout de 2 ans encore en vie à la fin de l'étude entrée au 01/03/1991	60	à mauvais risque	H	8000	A
13	décès au bout de 8 mois	22	à bon risque	H	3000	B
14	rechute au bout de 2,5 ans perdu de vue au bout de 3 ans	76	à mauvais risque	H	10000	A
15	perdu de vue au bout de 1 an 6 mois	69	à mauvais risque	H	15000	B
16	encore en vie à la fin de l'étude entrée au 01/07/1990	32	à bon risque	H	5000	A
17	rechute au bout de 3 ans 6 mois encore en vie à la fin de l'étude entrée au 01/04/1991	58	à bon risque	F	6000	B
18	RGH aigüe au bout de 6 mois RGH chronique au bout de 1 an 6 mois rechute au bout de 2 ans 4 mois décès au bout de 2 ans 8 mois	50	à bon risque	H	9000	B
19	rechute au bout de 1 an 8 mois	57	à risque moyen	H	14000	A
20	rechute au bout de 1 an 8 mois	56	à mauvais risque	H	25000	A

Table 15 – Observations réalisées pour les 20 patients atteints de leucémie. On note ici RGH = Réaction du greffon contre l'hôte, CC = changements chromosomiques, TL = taux de leucocytes au moment du diagnostic (exprimé en nombre de cellules par mm³).

n° de patient	observation
1	entrée dans l'étude le 1er mars 2000 récidive au bout de 3 ans, propagation extra-ganglionnaire au bout de 4 ans décès au bout de 5 ans
2	entrée dans l'étude le 1er mars 2000 récidive au bout de 5 ans, décès au bout de 6 ans
3	entrée dans l'étude le 1er mars 2000 récidive au bout de 6 ans, propagation extra-ganglionnaire au bout de 7 ans décès au bout de 8 ans
4	entrée dans l'étude le 1er juin 2000 rien ne se produit pendant l'étude
5	entrée dans l'étude le 1er mars 2000 déménagement pour mutation du conjoint au bout de 8 ans
6	entrée dans l'étude le 1er mars 2000 récidive au bout de 6 ans, décès au bout de 7 ans
7	entrée dans l'étude le 1er mars 2000 récidive au bout de 6 ans, décès au bout de 7 ans
8	entrée dans l'étude le 1er mars 2000 perdu de vue au bout de 2 ans
9	entrée dans l'étude le 1er mars 2000 récidive au bout de 2,5 ans, perdu de vue au bout de 3 ans
10	entrée dans l'étude le 1er mars 2000 récidive au bout de 2,5 ans, propagation extra-ganglionnaire au bout de 4 ans décès au bout de 5 ans
11	entrée dans l'étude le 1er mars 2000 récidive au bout de 2,5 ans, décès au bout de 3,5 ans
12	entrée dans l'étude le 1er mars 2000 propagation extra-ganglionnaire au bout de 7 ans décès au bout de 9,5 ans
13	entrée dans l'étude le 1er avril 2000 propagation extra-ganglionnaire au bout de 1 an, décès au bout de 2 ans
14	entrée dans l'étude le 1er mars 2000 propagation extra-ganglionnaire au bout de 1 an, décès au bout de 2 ans
15	entrée dans l'étude le 1er mars 2000 propagation extra-ganglionnaire au bout de 4 ans, décès au bout de 5 ans
16	entrée dans l'étude le 1er juin 2000 décès accidentel au bout de 3 ans
17	entrée dans l'étude le 1er mars 2001 propagation extra-ganglionnaire au bout de 8 ans
18	entrée dans l'étude le 1er décembre 2000 récidive au bout de 9 ans
19	entrée dans l'étude le 1er janvier 2001 récidive au bout de 9 ans
20	entrée dans l'étude le 1er mars 2000 récidive au bout de 2,5 ans, propagation extra-ganglionnaire au bout de 4 ans décès au bout de 6 ans et 6 mois
21	entrée dans l'étude le 1er février 2001 propagation extra-ganglionnaire au bout de 4 ans décès au bout de 5 ans
22	entrée dans l'étude le 1er juin 2000 décès au bout de 1,5 années
23	entrée dans l'étude le 1er juillet 2000 décès au bout de 1,5 années

Table 16 – Données recueillies lors de l'étude sur le lymphome non-hodgkinien

patient	observation	âge (en années)	IMC	présence d'hépatite virale	consommation d'alcool (en moyenne)	sexe	diabète ou autre syndrome métabolique
1	cancer au bout de 10 ans 6 mois décès au bout de 11 ans	54	29	non	excessive	H	oui
2	encéphalopathie au bout de 6 ans décès au bout de 10 ans	62	21.5	non	régulière	H	oui
3	décès au bout de 8 ans	51	21	oui	occasionnelle	H	non
4	cancer au bout de 10 ans décès au bout de 11 ans	58	21.4	oui	occasionnelle	F	non
5	décès accidentel au bout de 5 ans	53	23	oui	excessive	H	non
6	encéphalopathie au bout de 6 ans décès au bout de 10 ans	71	28	non	régulière	H	oui
7	encéphalopathie au bout de 12 ans décès au bout de 14 ans	61	20	oui	jamais	F	non
8	perdue de vue au bout de 6 ans	62	27.2	non	occasionnelle	F	oui
9	cancer au bout de 10 ans décès au bout de 10 ans 6 mois	57	20.8	oui	jamais	F	oui
10	décès au bout de 3 ans	56	25	oui	excessive	H	non
11	foie non fonctionnel au bout de 13 ans encore en vie à la fin de l'étude entrée au 01/12/1990	54	18	non	occasionnelle	H	oui
12	encéphalopathie au bout de 6 ans 6 mois perdu de vue au bout de 10 ans	60	25	oui	excessive	H	non
13	encore en vie à la fin de l'étude entrée au 01/03/1991	52	18.8	non	excessive	H	non
14	cancer au bout de 7 ans décès au bout de 7 ans 6 mois	56	19.9	non	occasionnelle	H	oui
15	foie non fonctionnel au bout de 5 ans perdu de vue au bout de 7 ans	69	23.4	non	jamais	H	oui
16	perdu de vue au bout de 8 ans	57	19.4	oui	excessive	H	non
17	encore en vie à la fin de l'étude entrée au 01/06/1990	58	20.9	oui	excessive	F	non
18	foie non fonctionnel au bout de 13 ans encore en vie à la fin de l'étude entrée au 01/12/1990	54	28	non	excessive	H	oui
19	cancer au bout de 10 ans décès au bout de 10 ans 6 mois	57	20.8	oui	occasionnelle	H	oui
20	décès au bout de 5 ans 6 mois	56	25	oui	excessive	H	non

Table 17 – Observations réalisées pour les 20 patients atteints de cirrhose du foie

patient	observation	âge	IMC	stade	consommation tabagique	sexe	histologie	traitement supplémentaire
1	rechute locale au bout de 1 an 6 mois décès au bout de 2 ans	54	29	II	excessive	H	à petites cellules	radiothérapie
2	métastases au bout de 6 mois décès au bout de 1 an	62	21.5	II	régulière	H	à petites cellules	immunothérapie
3	rechute locale au bout de 8 mois	51	21	III	occasionnelle	H	à petites cellules	immunothérapie
4	rechute locale au bout de 1 an 6 mois décès au bout de 2 ans	58	21.4	I	occasionnelle	F	adénocarcinome	immunothérapie
5	décès accidentel au bout de 18 mois	53	23	I	excessive	H	adénocarcinome	immunothérapie
6	rechute locale au bout de 2 ans décès au bout de 2 ans 10 mois	71	22	I	régulière	H	adénocarcinome	radiothérapie
7	métastases au bout de 6 mois décès au bout de 1 an	61	20	II	jamais	F	épidermoïde	radiothérapie
8	perdue de vue au bout de 2 ans	62	27.2	III	occasionnelle	F	épidermoïde	radiothérapie
9	rechute locale au bout de 10 mois métastases au bout de 14 mois décès au bout de 2 ans	57	20.8	II	jamais	F	épidermoïde	radiothérapie
10	décès au bout de 20 mois	56	25	II	excessive	H	épidermoïde	immunothérapie
11	rechute locale au bout de 24 mois encore en vie à la fin de l'étude entrée au 01/12/1990	54	18	I	occasionnelle	H	à grandes cellules	immunothérapie
12	rechute locale au bout de 1 an 6 mois perdu de vue au bout de 2 ans encore en vie à la fin de l'étude entrée au 01/03/1991	60	25	II	excessive	H	à grandes cellules	radiothérapie
13	rechute locale au bout de 2,5 ans perdu de vue au bout de 3 ans encore en vie à la fin de l'étude entrée au 01/06/1990	52	18.8	I	excessive	H	à grandes cellules	immunothérapie
14	décès au bout de 8 mois	56	19.9	III	occasionnelle	H	épidermoïde	radiothérapie
15	rechute locale au bout de 2,5 ans perdu de vue au bout de 3 ans encore en vie à la fin de l'étude entrée au 01/06/1990	69	23.4	III	jamais	H	à petites cellules	immunothérapie
16	rechute locale au bout de 1 an 6 mois perdu de vue au bout de 1 an 6 mois encore en vie à la fin de l'étude entrée au 01/06/1990	57	19.4	I	excessive	H	à grandes cellules	radiothérapie
17	rechute locale au bout de 2,5 ans encore en vie à la fin de l'étude entrée au 01/06/1990	58	20.9	I	excessive	F	à grandes cellules	immunothérapie
18	rechute locale au bout de 2,5 ans encore en vie à la fin de l'étude entrée au 01/12/1990	54	28	II	excessive	H	épidermoïde	immunothérapie
19	métastases au bout de 1 an décès au bout de 1 an 6 mois	57	20.8	II	occasionnelle	H	adénocarcinome	radiothérapie
20	décès au bout de 1 an 8 mois	56	25	II	excessive	H	à petites cellules	radiothérapie

Table 18 – Observations réalisées pour les patients n° 1 à 20 atteints de cancer primitif du poumon

patient	observation	âge	IMC	stade	consommation tabagique	sexe	histologie	traitement supplémentaire
21	rechute locale au bout de 24 mois encore en vie à la fin de l'étude entrée au 01/12/1990	64	18	II	occasionnelle	F	à grandes cellules	radiothérapie
22	métastases au bout de 1 an décès au bout de 1,5 ans	60	28	III	excessive	F	à grandes cellules	immunothérapie
23	encore en vie à la fin de l'étude entrée au 01/03/1990	52	18.8	I	jamais	F	à grandes cellules	immunothérapie
24	décès au bout de 8 mois	56	19.9	III	occasionnelle	H	épidermoïde	radiothérapie
25	rechute locale au bout de 2,5 ans perdu de vue au bout de 3 ans	63	23.4	II	jamais	H	à petites cellules	immunothérapie
26	perdu de vue au bout de 1 an 6 mois	57	19.4	II	excessive	H	à grandes cellules	radiothérapie
27	encore en vie à la fin de l'étude entrée au 01/06/1990	58	20.9	I	excessive	F	à grandes cellules	immunothérapie
28	rechute locale au bout de 2,5 ans encore en vie à la fin de l'étude entrée au 01/12/1990	64	28	II	excessive	H	épidermoïde	immunothérapie
29	métastases au bout de 1 an décès au bout de 1 an 6 mois	57	22.8	III	occasionnelle	F	adénocarcinome	radiothérapie
30	décès au bout de 1 an 8 mois	56	25	II	excessive	F	à petites cellules	radiothérapie
31	rechute locale au bout de 20 mois	57	24	II	excessive	H	à petites cellules	radiothérapie
32	métastases au bout de 6 mois décès au bout de 1 an	62	21.5	II	régulière	H	à petites cellules	immunothérapie
33	décès au bout de 8 mois	51	21	III	occasionnelle	H	à petites cellules	radiothérapie
34	rechute locale au bout de 1 an 6 mois décès au bout de 2 ans	58	21.4	I	occasionnelle	F	adénocarcinome	immunothérapie
35	décès accidentel au bout de 18 mois	53	23	I	jamais	H	adénocarcinome	immunothérapie
36	rechute locale au bout de 2 ans décès au bout de 2 ans 10 mois	71	22	I	régulière	H	adénocarcinome	radiothérapie
37	métastases au bout de 6 mois décès au bout de 1 an	61	20	II	occasionnelle	H	à petites cellules	radiothérapie
38	métastases au bout de 4 mois décès au bout de 8 mois	62	27.2	I	occasionnelle	F	épidermoïde	radiothérapie
39	rechute locale au bout de 10 mois rechute locale au bout de 14 mois	53	20.8	III	excessive	H	épidermoïde	immunothérapie
40	rechute locale au bout de 10 mois rechute locale au bout de 18 mois décès au bout de 2 ans	53	20.8	III	excessive	H	épidermoïde	immunothérapie

Table 19 – Observations réalisées pour les patients n° 21 à 40 atteints de cancer primitif du poumon