

Analyse de Survie

- | | | |
|------|-----------------------------------|----|
| I. | Rappels de PACES | EC |
| II. | Régressions multiples et Cox | EC |
| III. | Modèle de Cox | CP |
| IV. | Fiche d'interprétation Cox | GF |
| V. | Exemples (articles scientifiques) | GF |
| VI. | Exemples (analyse de cas) | EC |

Analyse de Survie

Rappels de PACES

- I. Introduction
- II. Généralités sur la Survie
- III. Tracer la courbe de survie
- IV. Effectuer un test statistique

1- Introduction

1a- Contexte

- Maladies graves :
 - Augmentation du risque de décès, donc diminution de l'espérance de vie
 - Meilleur traitement : celui qui diminue ce risque
- Problème de « survie » concerne tout événement dans le temps:
 - Délai jusqu'au décès
 - Délai jusqu'à la première rechute
 - Durée de couverture assurancielle sans accident
 - Attente d'un client au guichet de la Poste
 - Etc.
- Définition :
 - soit la variable aléatoire T =délai entre l'inclusion et l'événement
 - ex : délai entre le diagnostic et le décès

1- Introduction

1b- Gravité : événement ou temps ?

- **⚠** l'événement peut être observé ou non, sur une durée variable.
- Quelle est la situation la plus favorable :
 - Alternative n°1 (délai =, événement ≠) :
 - Mourir au bout de 5 ans
 - Ne pas mourir pendant 5 ans
 - Alternative n°2 (délai ≠, événement =) :
 - Mourir au bout de 5 ans
 - Mourir au bout de 10 ans
 - Alternative n°3 (délai ≠, événement =) :
 - Ne pas mourir pendant 5 ans (suite inconnue)
 - Ne pas mourir pendant 10 ans (suite inconnue)
 - Alternative n°4 (délai ≠, événement ≠) :
 - Mourir au bout de 10 ans
 - Ne pas mourir pendant 5 ans (suite inconnue)
- => impossible de dissocier temps et événement !

1- Introduction

1c- Particularités de T

- Rappel : T =délai entre l'inclusion et l'événement
- Distribution particulière :
 - Toujours positive
 - Distribution non normale
 - Cas n°1 : l'événement est observé au bout d'un délai d : $t_i = d$
 - Cas n°2 : l'événement n'est toujours pas observé au bout d'un délai d : on sait seulement que $t_i > d$
 - Cas 2a : patient suivi jusqu'au bout de l'étude sans décès = « exclu vivant »
 - Cas 2b : patient perdu en cours d'étude, mais vivant aux dernières nouvelles = « perdu de vue »
 - Cas 2a&2b : problème statistique identique : on parle de « censure à droite » car $t_i > d$ et t_i partiellement inconnu (terme non péjoratif)
 - Cas 2a&2b : problème méthodologique différent : pourquoi a-t-on perdu certains individus de vue ? (exemple : arrêt du traitement du fait de l'apparition d'effets indésirables)

1- Introduction

1d- Méthodes classiques inadaptées

- Analyser *moyenne*(T) :
 - Impossible s'il existe des valeurs censurées
 - Test de comparaison (Student) : même problème + distribution non normale
- Analyser *médiane*(T) :
 - Impossible s'il existe des valeurs censurées parmi la première moitié :
 - Ex1 : $T=\{1 ; 2 ; >5\}$: médiane=2
 - Ex2 : $T=\{1 ; >2 ; 5\}$: médiane inconnue
 - Test de comparaison des médianes : même problème
 - Idem pour les autres quantiles
- Analyser proportion : $p(T > t_0)$:
 - Impossible s'il existe des valeurs censurées avant cette borne t_0 :
 - Ex1 : $T=\{1 ; 2 ; >5\}$: $p(T>3)=1/3$
 - Ex2 : $T=\{1 ; >2 ; 5\}$: $p(T>3)$ inconnue
 - D'ailleurs, à l'extrême, lorsque délai très élevé, 100% d'événement
 - Test de comparaison (Chi^2) : même problème

2- Généralités sur la Survie

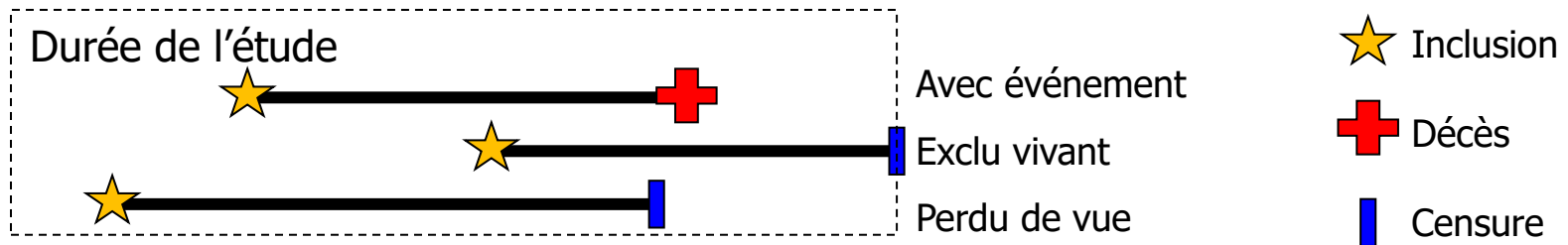
2a- Méthodes pour l'analyse de Survie

- Analyse descriptive :
 - Paramétrique possible en spécifiant la loi de T
 - Ex : Loi exponentielle
 - Non paramétrique :
 - Estimation de Kaplan-Meier (très utilisé en médecine)
 - Estimation actuarielle (hors programme)
- Comparaison statistique de deux groupes :
 - Test non paramétrique : Test du Log Rank
 - Test semi-paramétrique : Modèle de Cox avec une variable explicative binaire
- Analyse multivariée (ajustement sur plusieurs variables explicatives) :
 - Test semi-paramétrique : Modèle de Cox

2- Généralités sur la Survie

2b- Notion de censure

- Terme non péjoratif : « censure » = « information incomplète » (le patient n'est pas supprimé)
- Censure à droite (très fréquente en médecine) :
 - Certaines données t_i sont partiellement inconnues : on sait seulement que $t_i > d$
 - Autant pour les **perdus de vue** que les **exclus vivant**



- Censure à gauche (hors programme)
 - Certains t_i partiellement inconnus : on sait seulement que $t_i < d$
 - Ex : Un patient est inclus en 2005, pas de diabète type 2. Revu en 2010, un diabète type 2 asymptomatique est diagnostiqué : $t_i < 5$ ans
- Censure à gauche et à droite (hors programme)

2- Généralités sur la Survie

2c- Fonctions décrivant la survie

- Probabilité instantanée de décès $h(t)$
 - $h(t) = \frac{P(t \leq T \leq (t+\Delta t) \mid T \geq t)}{\Delta t}$ où Δt tend 0
- Fonction de décès $R(t)$
 - $R(t) = P(T \leq t)$
 - (si pas de censure, c'est la fonction de répartition de T)
- Fonction de survie $S(t)$
 - Probabilité de survivre au moins jusqu'à la date t
 - $S(t) = P(T \geq t) = 1 - R(t)$
 - Peut être estimée sur un échantillon : « courbe de survie »

3- Tracer la courbe de survie

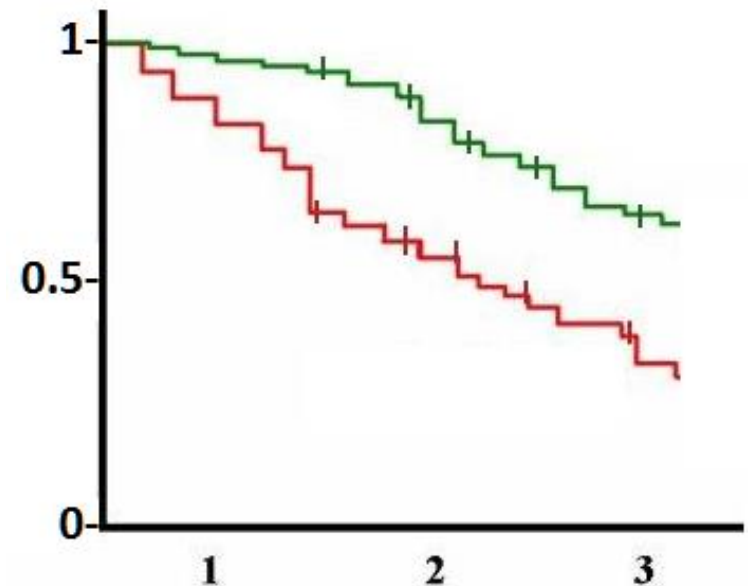
Estimateur de Kaplan-Meier

- Estimation graphique de la courbe de survie
- Méthode de Kaplan-Meier, utilisable avec (ou sans) censure à droite :
 - $S(t)$ estimée par $\hat{S}(t_i) = \prod_{j=1}^i (1 - \frac{d_j}{n_j})$
 - d_j =nb de sujets décédés à la date t_j
 - n_j =nb de sujets en vie juste avant la date t_j
 - Si simultanés, traiter les décès avant les censures

3- Tracer la courbe de survie

Estimateur de Kaplan-Meier

- Représente estimation de $S(t)$ en fonction de t
- Escalier descendant, démarre à 1, descend vers 0 (l'atteint ou non)
- Superpositions possibles (ici rouge=placebo, vert=traitement efficace)
- Aspect d'un décès :
 - Marche descendante
- Aspect d'une censure :
 - Pas de marche descendante (sauf si décès simultané)
 - Peut être signalée par une coche verticale ou un rond
 - Si un seul décès par marche, les marches suivantes sont plus hautes que les précédentes



Décès :

Censure :

n= 6 5 4 3

 -1/6 -1/4

4- Effectuer un test statistique

4a- Comparer 2 groupes : Log Rank

- Principes du test :
 - H_0 : les taux de décès des deux groupes sont identiques
 - Calcul des effectifs théoriques obtenus sous H_0
 - Calcul d'une métrique d'écart entre les effectifs observés et théoriques, qui suit une loi du Chi^2
 - Rejet de H_0 au-delà du seuil donné par la table du Chi^2 (ou si $p < 5\%$)

4- Effectuer un test statistique

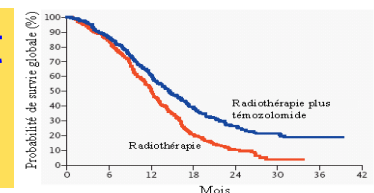
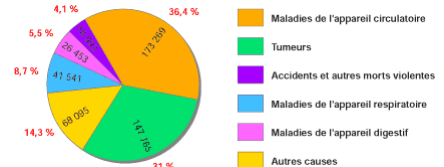
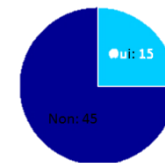
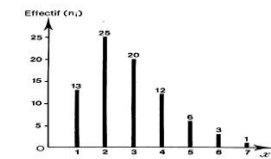
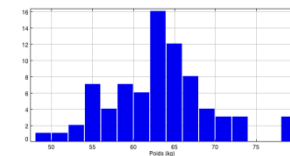
4b- Modèle de Cox

- Avec une variable X binaire :
 - Estime la fonction $h(t)$ dans les deux groupes
 - Cherche un facteur multiplicateur constant (« hypothèse des risques proportionnels »)
 - Ce facteur est-il significativement différent de 1 ?
- Avec d'autres variables explicatives :
 - On peut ajouter les variables souhaitées
 - Intérêt : au moins ajuster sur l'âge des patients (plus on est âgé, plus on meurt...)

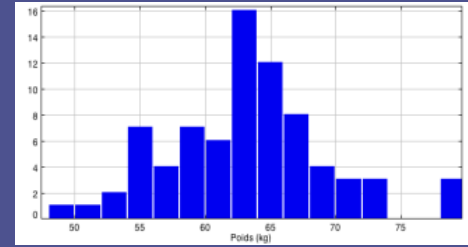
Le modèle de Cox, une sorte de régression multiple

Les régressions multiples

- Permettent d'expliquer une variable Y à l'aide de plusieurs variables X_i
- Y variable quantitative
=> régression linéaire ex : poids
- Y variable de décompte
=> régression de Poisson ex : nb hospitalisations
- Y binaire
=> régression logistique ex : malade 1/0
- Y qualitative
=> régression logistique multinomiale ex : cause décès
- Y événement dans le temps
=> modèle de Cox ex : survie sous traitement



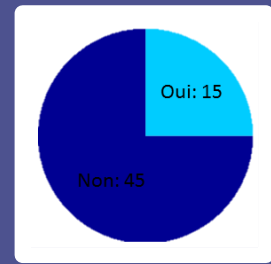
Coefficients de la régression linéaire



■ Régression linéaire :

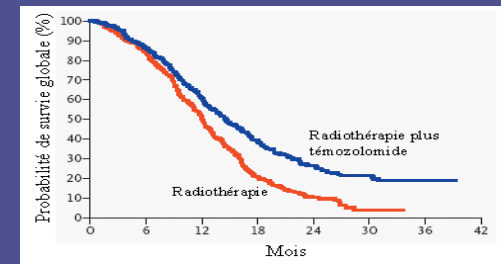
- *Ex : expliquer le poids (kg) par une combinaison de mesures anthropométriques X_i (cm)*
- Chaque coefficient β_i indique **la contribution linéaire ajustée** de X_i pour expliquer Y
- *Ex : $\beta_{taille}=1.04$ s'interprète : « quand la taille augmente d'1 cm, en moyenne le poids augmente d'1.04 kg »*

Coefficients de la régression logistique...



- Régression logistique :
 - *Ex : expliquer le statut « bronchite chronique » 0/1 par une combinaison de facteurs de risque X_i*
 - Chaque exponentielle de coefficient e^{β_i} donne l'**odds ratio ajusté** de X_i pour expliquer Y
 - *Ex : $e^{\beta_{\text{tabac}}}=8,5$ s'interprète : « l'odds ratio ajusté du tabac et de la bronchite chronique vaut 8,5 »*
 - NB : cette notion correspond à une observation statique (type cas-témoin) et ne tient pas compte de la temporalité de l'exposition et de l'événement

Coefficients du modèle de Cox



- Modèle de Cox (sorte de « régression de survie ») :
 - *Ex : expliquer la probabilité instantanée de décès par une combinaison de facteurs de risque X_i*
 - Chaque exponentielle e^{β_i} donne le **hazard ratio ajusté** (risque relatif instantané) de X_i pour expliquer Y
 - *Ex : $e^{\beta_{tabac}}=1.24$ s'interprète : « à chaque instant, les fumeurs ont 1.24 fois plus de chances de mourir »*
 - /!\ ne pas directement conclure sur la durée de survie : elle en résulte, mais relation complexe

Analyse de cas

Données : analyse des ~10 500 séjours d'une année complète d'activité d'un centre hospitalier

Chaque séjour :

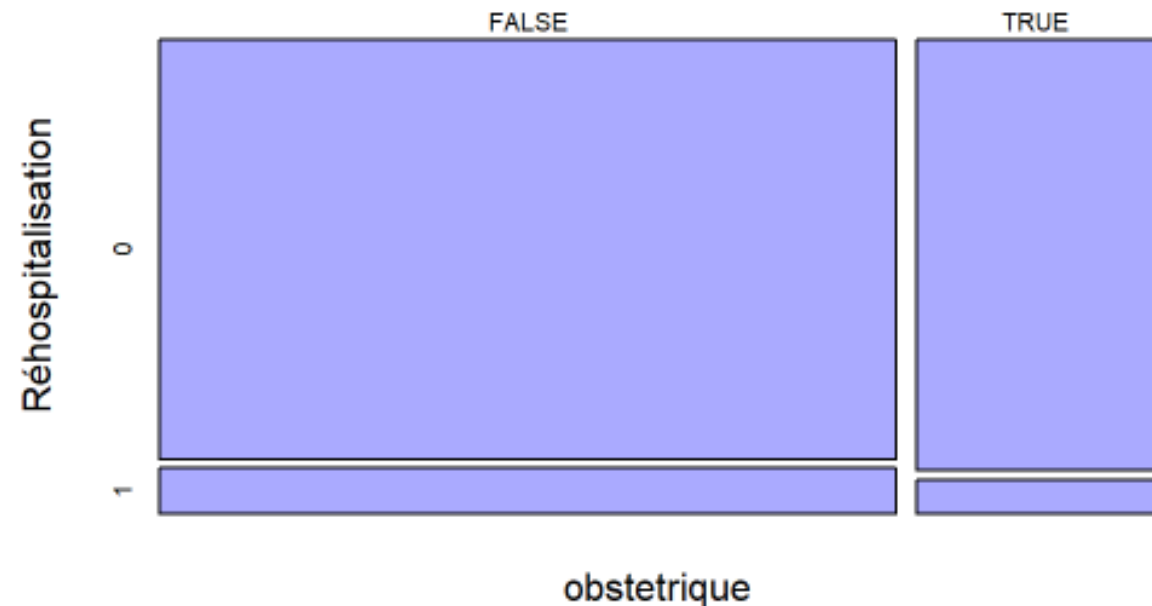
- Durée de séjour
- Réhospitalisation (0/1, avec délai)
- Décès en sortie de séjour (0/1)
- Discipline (cancéro, obstétrique, etc.)
- Age, sexe

Ex 1 – réhospitalisations obstétrique vs les autres

- Situation :
 - Patients suivis pendant une durée variable.
 - Pour certains, on observe une réhospitalisation (délai connu).
 - Pour d'autres, on n'observe pas de réhospitalisation (délai de suivi connu)
- Question : les patients d'obstétrique sont-ils plus ou moins réhospitalisés que les autres ?

Ex 1 – réhospitalisations obstétrique vs les autres

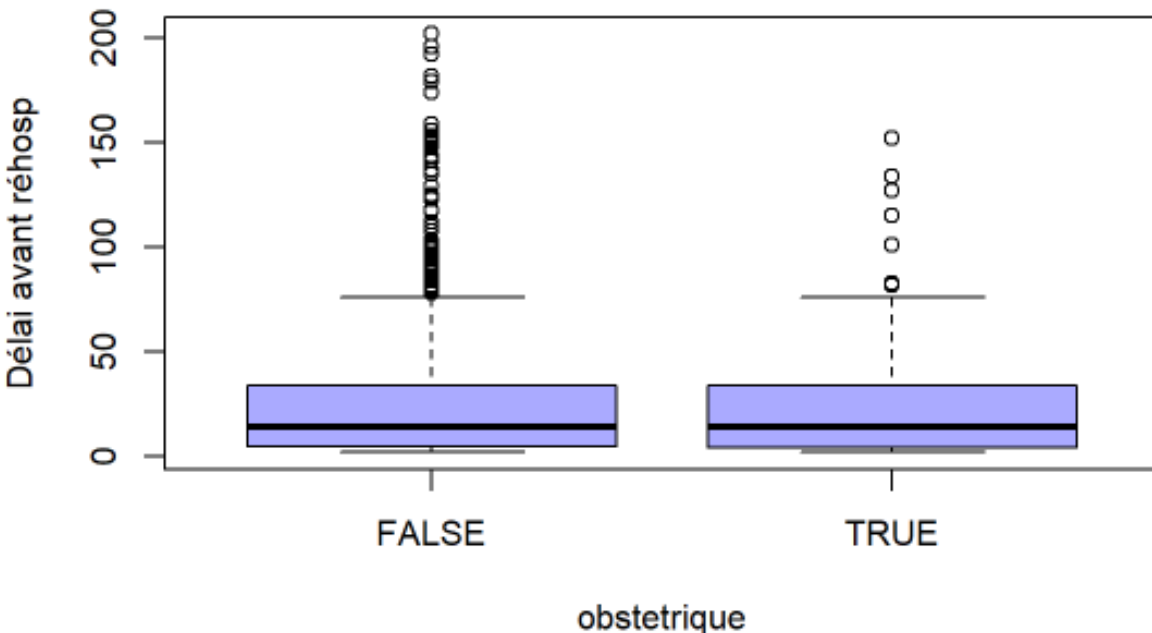
- Première approche : taux brut de réhospitalisation
- Test du Khi^2 : $p=0.00037$



	Autres spé	Obstétrique
Pas de réhosp.	90.2%	95.6%
Réhosp	9.8%	7.4%

Ex 1 – réhospitalisations obstétrique vs les autres

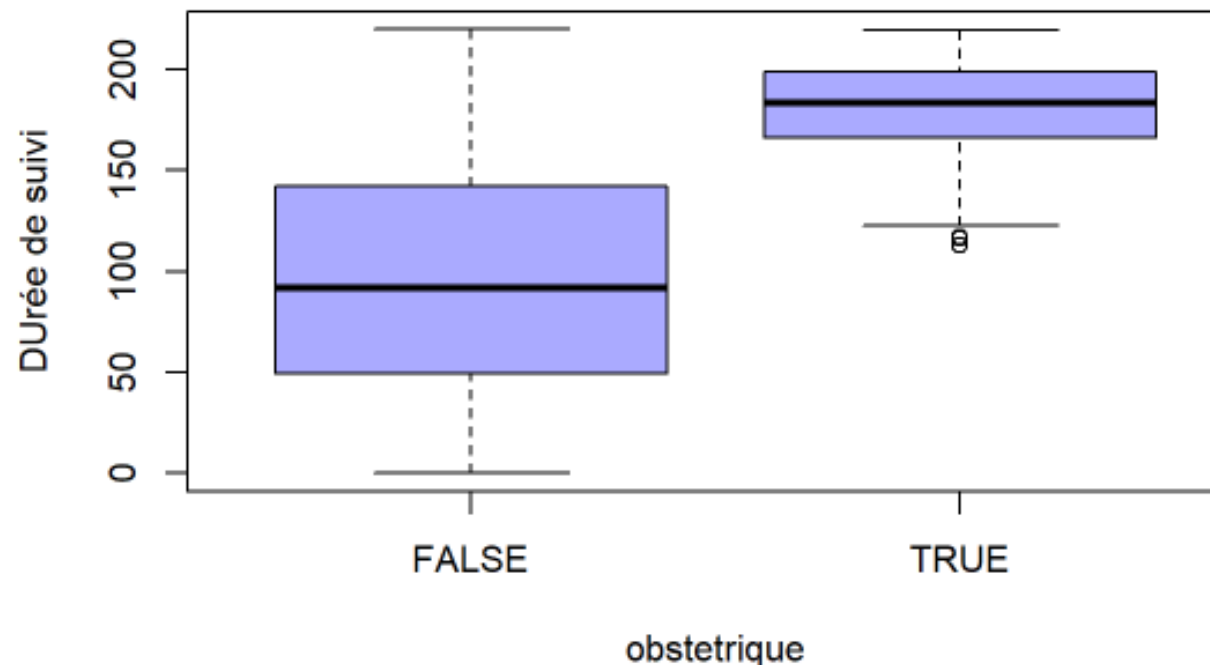
- Et ce délai de réhospitalisation est-il le même ?
- Deuxième approche : comparaison chez les seules personnes réhospitalisées
- Test de Student : $p=0.18$



	Autres spé	Obstétrique
Délai moyen	26,7j	23,5j

Ex 1 – réhospitalisations obstétrique vs les autres

- Mais problème de fond...
- Comparaison de la durée de suivi chez les seules personnes qui n'ont pas été réhospitalisées...
- Test de Student : $p=0$



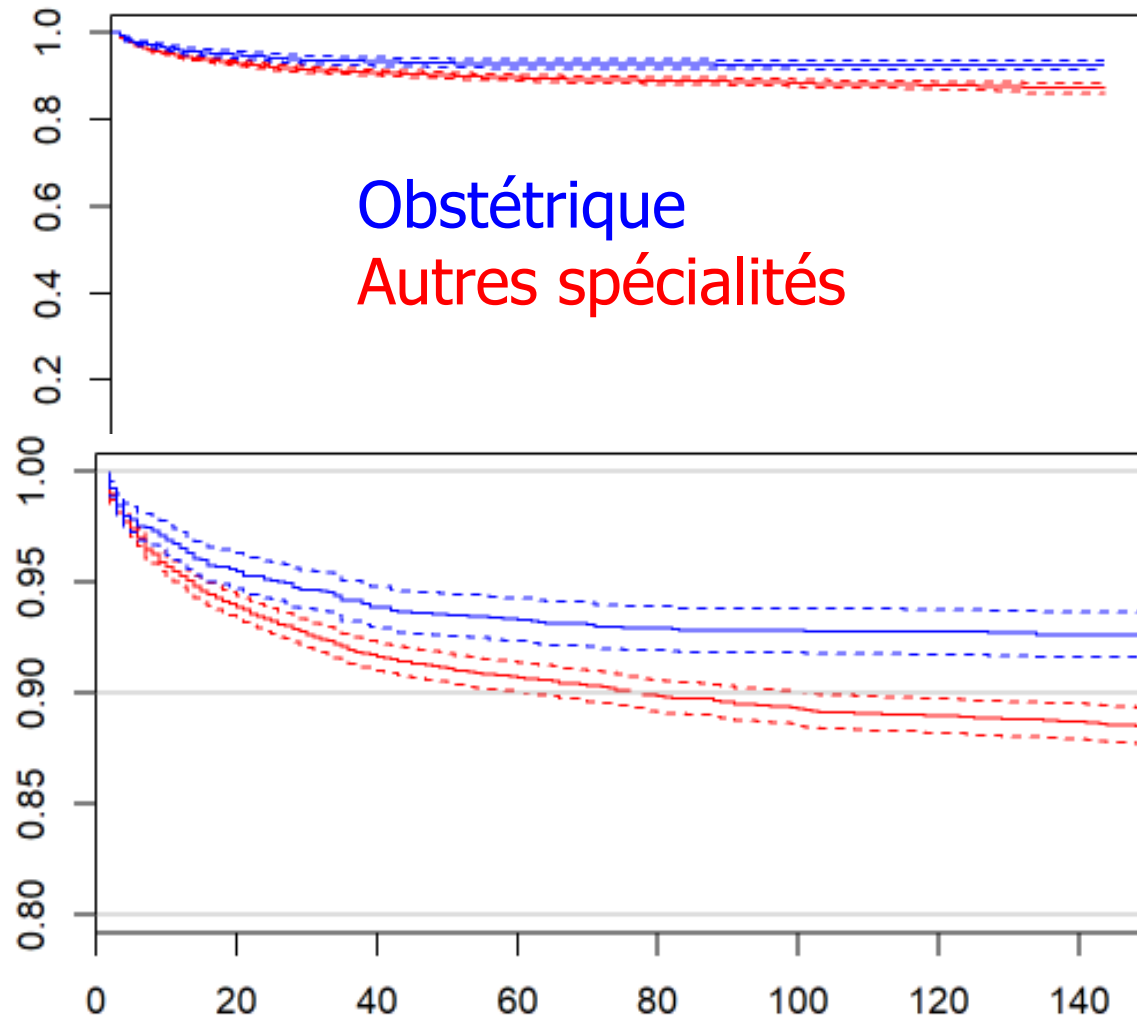
	Autres spé	Obstétrique
Suivi moyen	97.6j	181.8j

Ex 1 – réhospitalisations obstétrique vs les autres

- Il est donc impossible d'interpréter le taux brut de réhospitalisations :
 - La question d'être « plus réhospitalisé » n'est pas dissociable d'être « réhospitalisé plus tôt »
 - Les patients n'ont pas tous été suivis aussi longtemps
 - Les patientes d'obstétrique sont suivies plus longtemps que les autres (on pourrait donc sous-estimer le taux de réhospitalisation des autres patients)

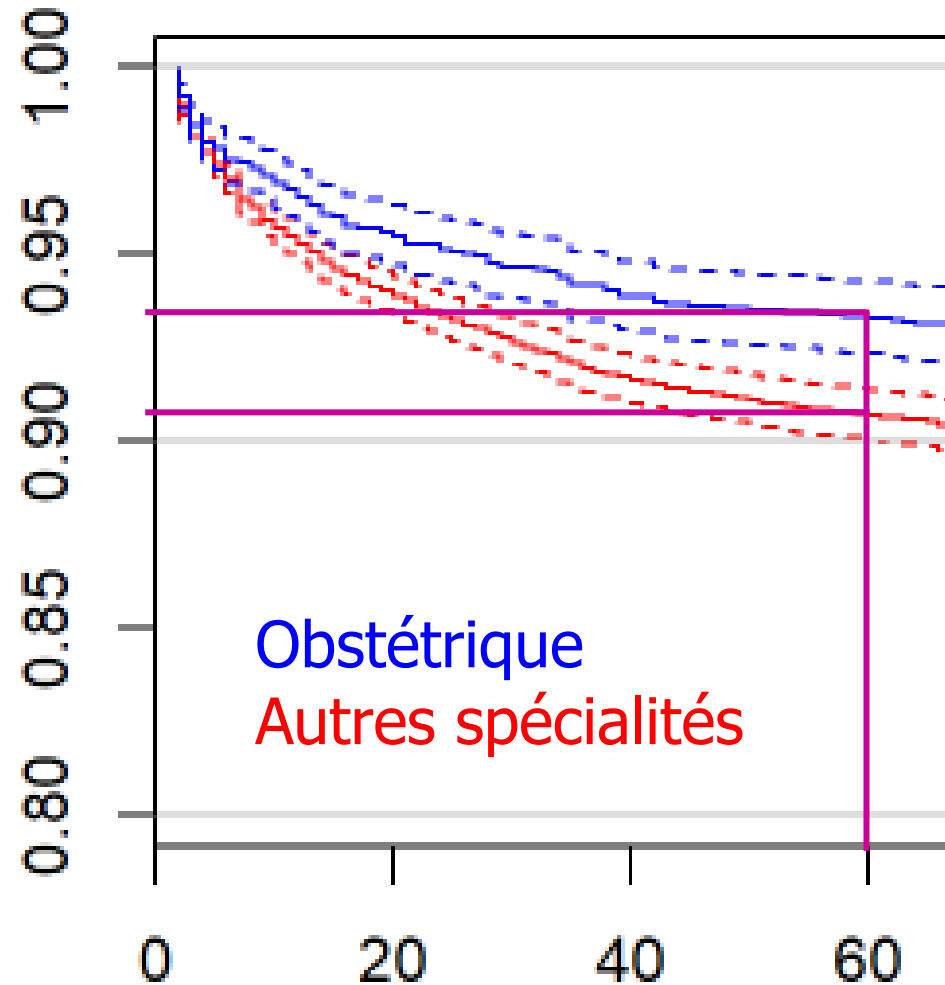
Ex 1 – réhospitalisations obstétrique vs les autres

- Courbe de Kaplan Meier.
- Evénement = réhospitalisation
- Délai :
 - Si réhospitalisés, délai jusqu'à la réhospitalisation
 - Sinon, durée de suivi sans réhospitalisation



Ex 1 – réhospitalisations obstétrique vs les autres

- Courbe de Kaplan Meier
- Estimation du taux de réhospitalisations à 60 jours :
 - 6,7% en obstétrique
 - 9,3% dans les autres spécialités
- NB : estimation possible grâce à KM alors que 29% des patients sont suivis moins de 60j



Ex 1 – réhospitalisations obstétrique vs les autres

- Test du Log Rank :

```
survdifff(formula = obj_surv ~ df$obstetrique)
n=10441
```

	N	Observed	Expected
obstetrique=0	7889	771	690
obstetrique=1	2552	189	270

Chisq= 34.7 on 1 degrees of freedom, $p=3.85e-09$

- Modèle de Cox avec une seule variable explicative binaire

```
coxph(formula = obj_surv ~ obstetrique)
n= 10441, number of events= 960
```

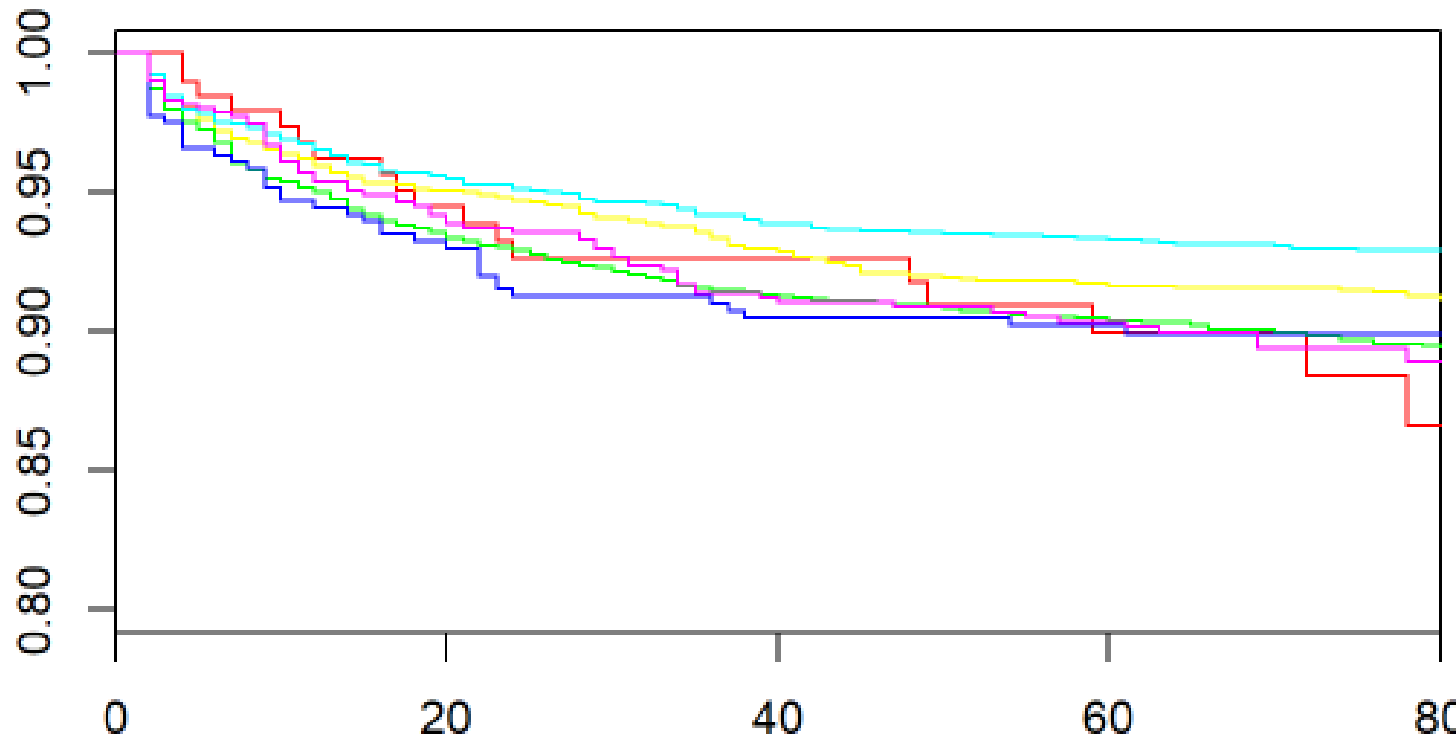
Hazard ratio (exp(beta)):

$obstetrique=1 : 0.62 [0.528 ; 0.728]$

Likelihood ratio test=37.2 on 1 df, $p=1.04e-09$

Ex 2 – réhospitalisations selon le type de séjour

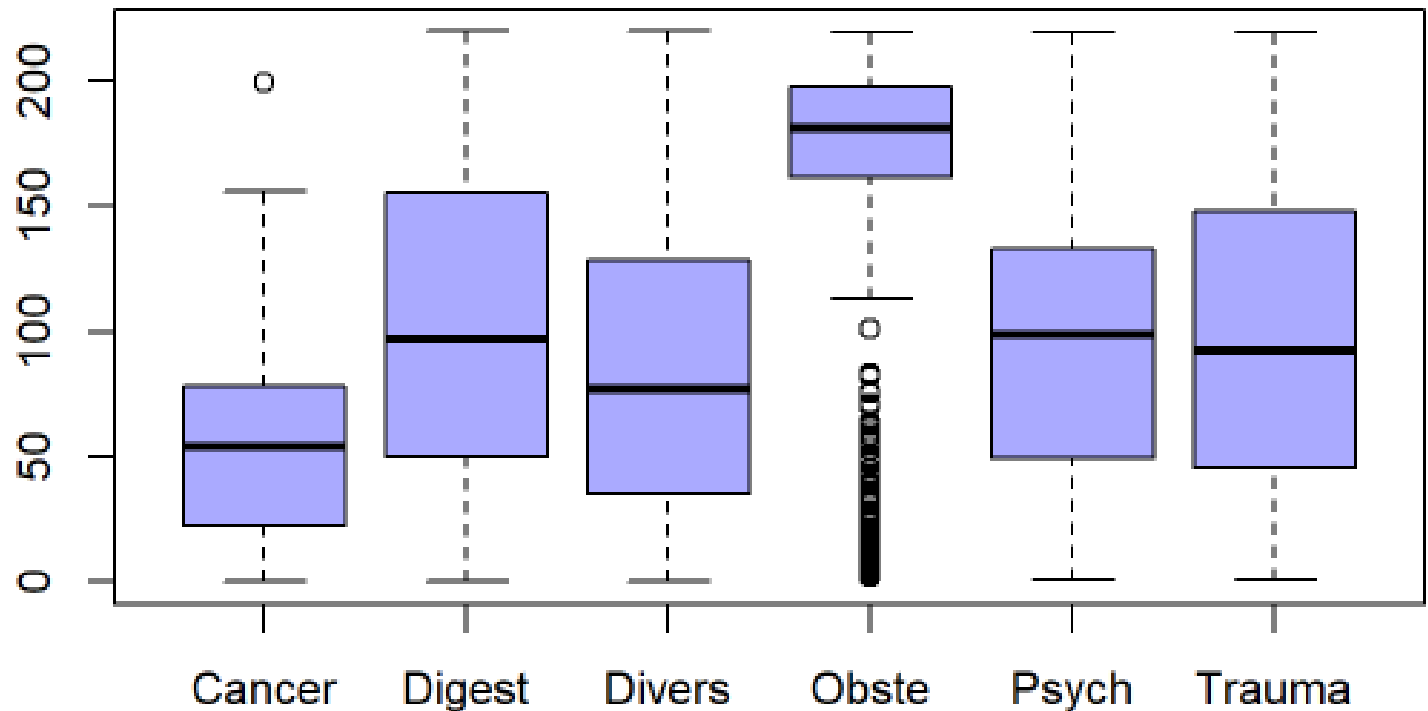
- 6 types : Cancers, Digestif, Divers, Obstétrique, Psychiatrie, Traumatismes
- Courbes de Kaplan Meier



Commentaire :
Nette différence
en apparence

Ex 2 – réhospitalisations selon le type de séjour

- Durées de suivi des patients



Commentaire :
Durées de suivi très différentes
=> méthodes de survie incontournables

Ex 2 – réhospitalisations selon le type de séjour

- Test du Logrank : $p= 4.4e-08$
- Modèle de Cox avec 1 variable explicative qualitative : $p=1.8e-08$

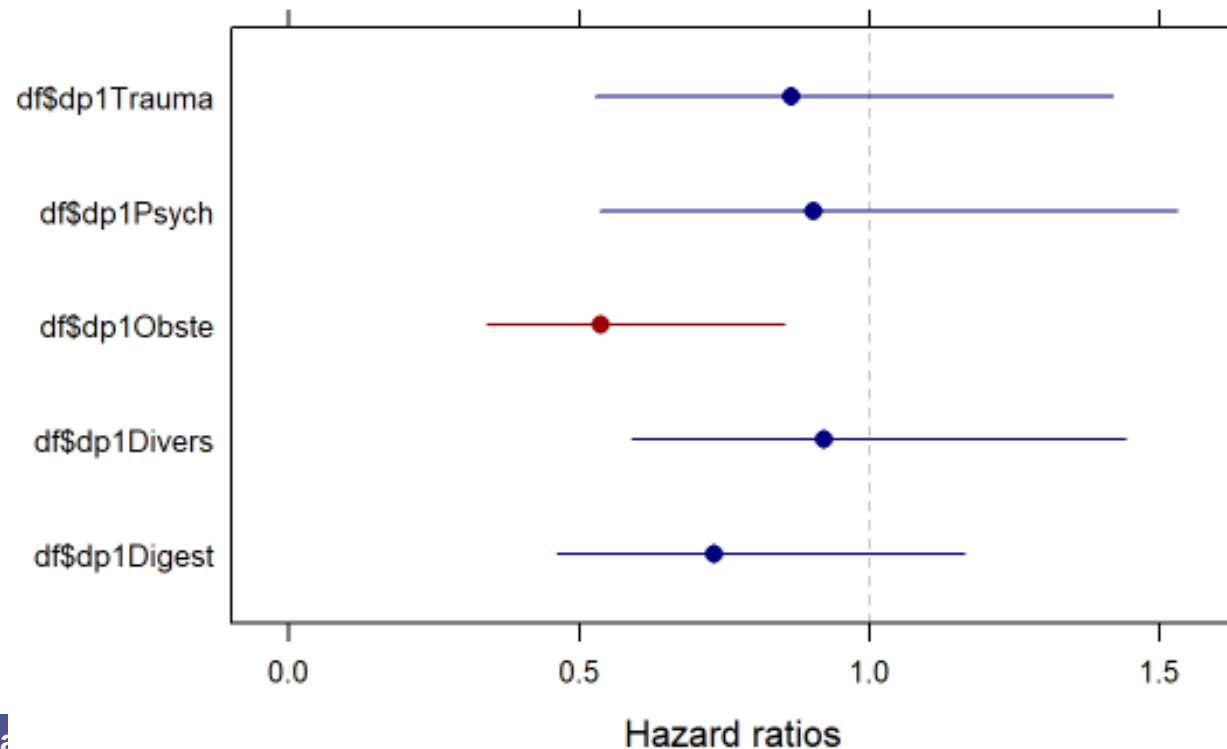
Hazard ratios et int. de conf. à 95% :

Trauma	: 0.865	[0.527 ; 1.42]
Psych	: 0.905	[0.535 ; 1.53]
Obste	: 0.538	[0.339 ; 0.854]
Divers	: 0.922	[0.589 ; 1.44]
Digest	: 0.732	[0.46 ; 1.16]

Commentaire :

P valeurs différentes mais même ordre de grandeur.
Cox : estime HR et fournit IC95%

Cancers pris comme référence, donc HR Obstétrique modifié.
Facteurs de confusion ?



Ex 2 – réhospitalisations selon le type de séjour

- Modèle de Cox avec 1 variable explicative qualitative + âge + sexe : $p=1.8e-08$

Hazard ratios et int. de conf. à 95% :

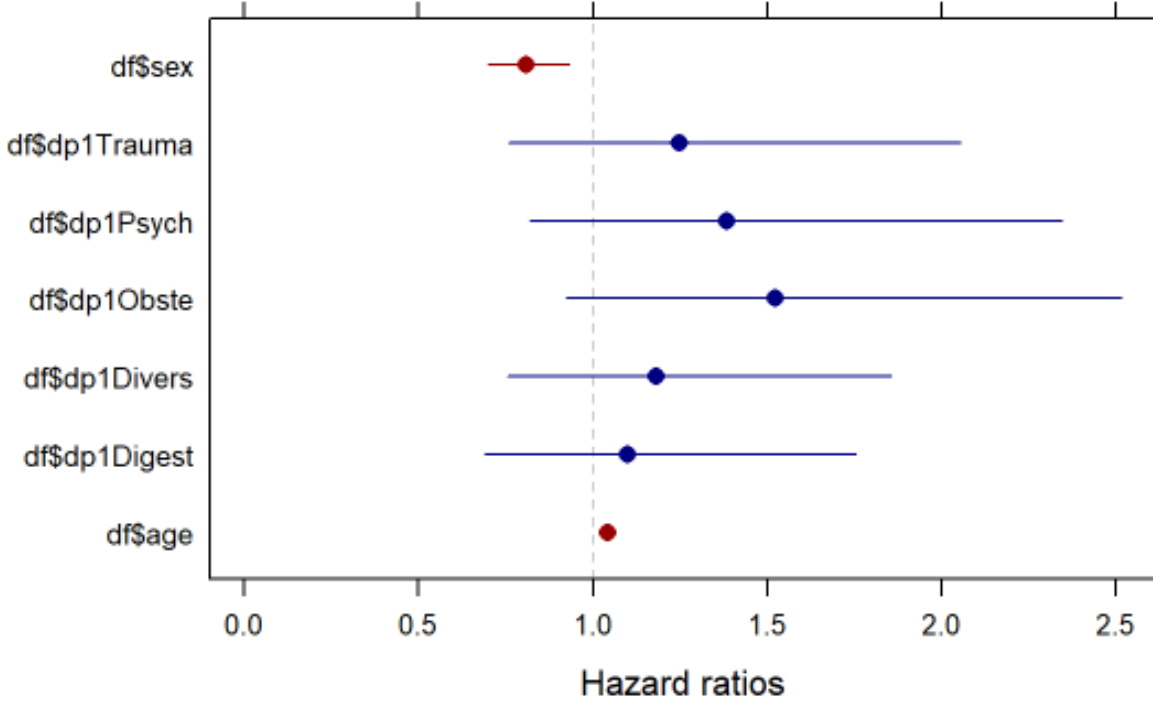
Sexe (ref=H):	0.807	[0.699 ; 0.932]
Diag=Trauma :	1.25	[0.759 ; 2.06]
Diag=Psych :	1.38	[0.817 ; 2.35]
Diag=Obste :	1.52	[0.924 ; 2.52]
Diag=Divers :	1.18	[0.754 ; 1.86]
Diag=Digest :	1.1	[0.687 ; 1.75]
Age :	1.04	[1.04 ; 1.05]

Commentaire :

L'obstétrique n'est plus un facteur discriminant : en fait il s'agissait de patients jeunes...

=> Toujours vérifier si âge et sexe pris en compte

Un changement d'échelle de l'âge serait utile



Ex 2 – réhospitalisations selon le type de séjour

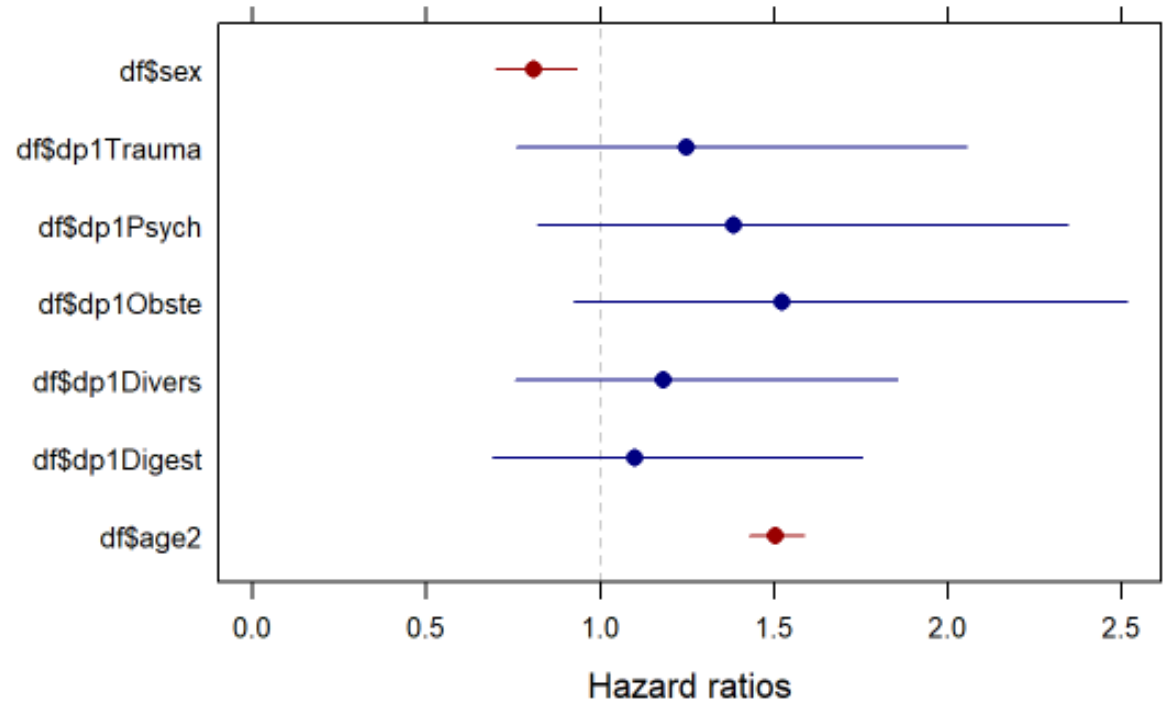
- $\text{age2} = \text{age}/10$
- Modèle de Cox avec 1 variable explicative qualitative + age2 + sexe : $p=1.8e-08$

Hazard ratios et int. de conf. à 95% :

Sexe :	0.807	[0.699 ; 0.932]
Diag=Trauma :	1.25	[0.759 ; 2.06]
Diag=Psych :	1.38	[0.817 ; 2.35]
Diag=Obste :	1.52	[0.924 ; 2.52]
Diag=Divers :	1.18	[0.754 ; 1.86]
Diag=Digest :	1.1	[0.687 ; 1.75]
Age2 :	1.5	[1.43 ; 1.59]

Commentaire :

Le HR de l'âge devient interprétable : risque multiplié par 1,5 lorsque l'âge augmente de 10 ans. Rien d'autre ne change



Ex 2 – réhospitalisations selon le type de séjour

- Même modèle avec sélection « stepwise »

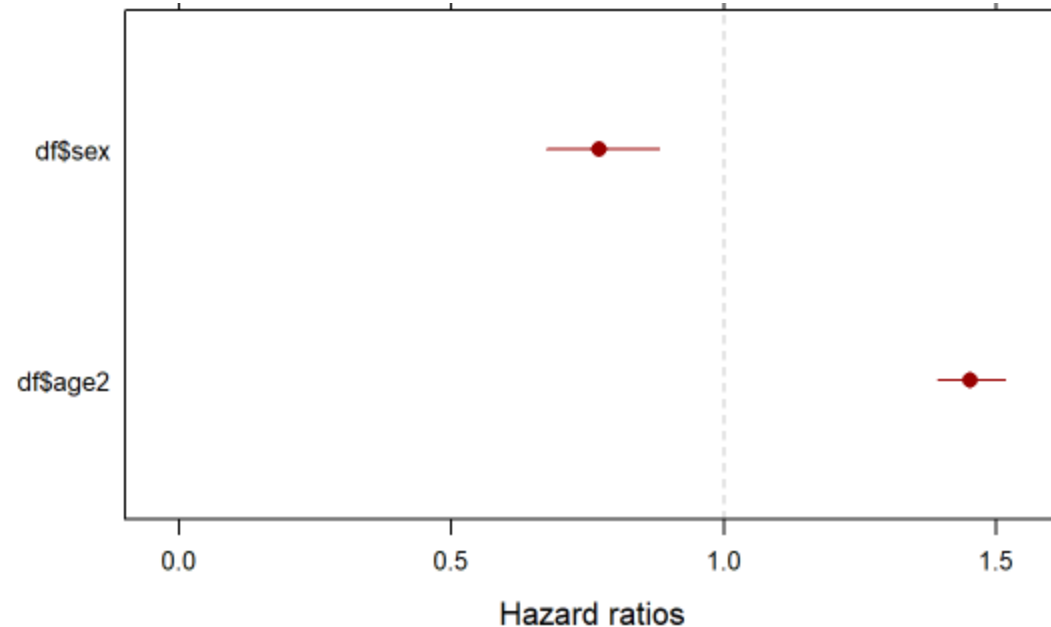
Hazard ratios et intervalle de confiance à 95% :

sexe : 0.77 [0.673 ; 0.881]

age2 : 1.45 [1.39 ; 1.52]

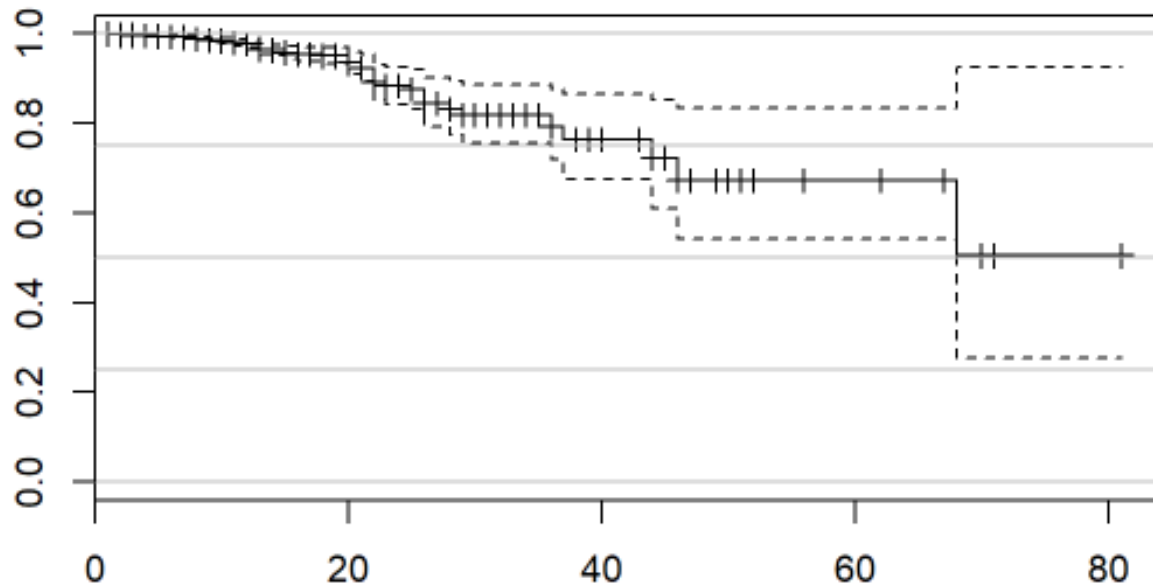
Commentaire :

Les procédures de sélection pas-à-pas fonctionnent de manière similaire au cas des régressions linéaires (cf. cours précédents)



Ex 3 – mortalité durant le séjour hospitalier

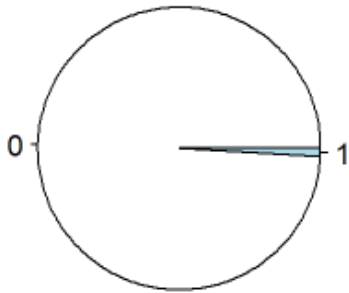
- Modélisation choisie :
 - Séjours hospitaliers
 - Événement : décès 0/1
 - Censure (à droite) : patients quittant le service pour rentrer à domicile
 - Délai : dans tous les cas, la durée de séjour
- Courbe de Kaplan-Meier (avec censures et IC95%)



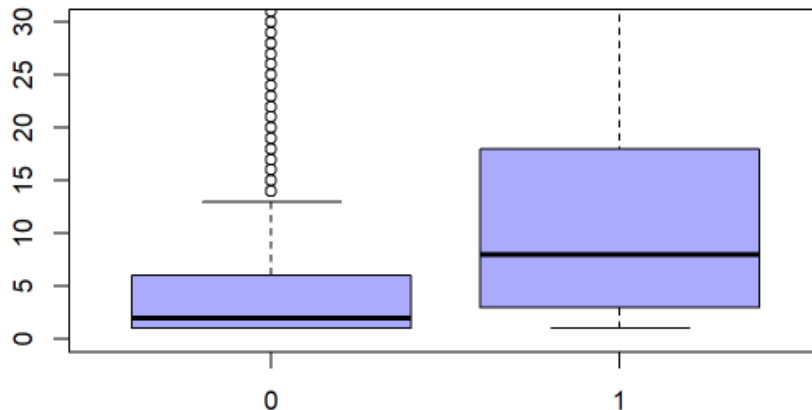
Commentaire :
La mortalité semble
considérable...

Ex 3 – mortalité durant le séjour hospitalier

- Analyses complémentaires :
 - proportion brute de morts : 0,85%



- durée de suivi en fonction du décès :



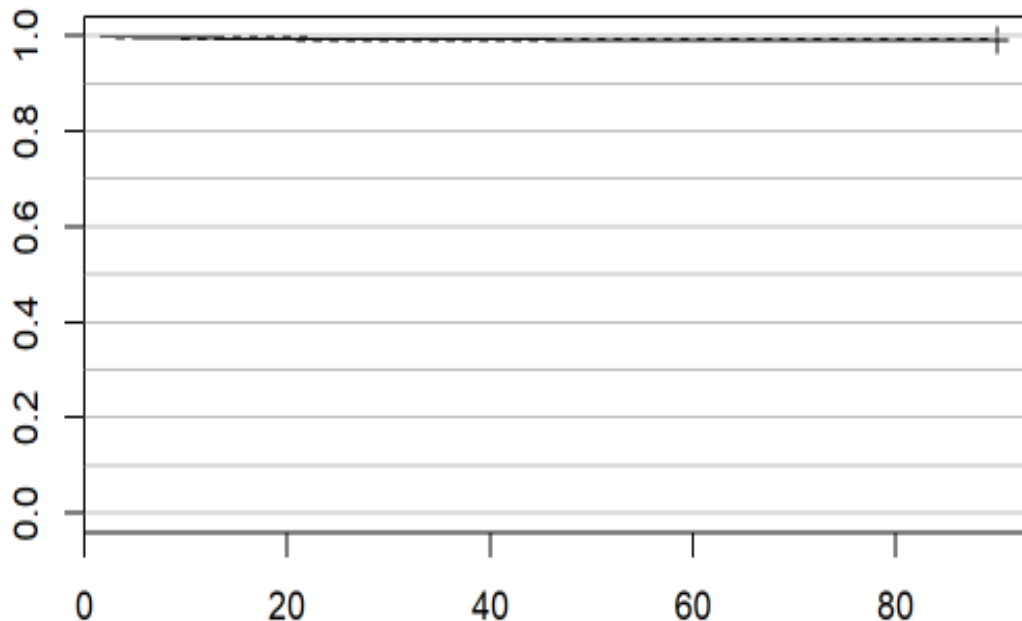
Commentaire :

De toute évidence, les « censurés » ne mourront pas : c'est parce qu'une personne va mieux qu'elle quitte l'hôpital. C'est parce qu'elle se dégrade qu'elle y reste. Les censures sont donc **informatives**.

L'approche survie n'est donc pas valide ici !!!

Ex 3 – mortalité durant le séjour hospitalier

- Pour s'en convaincre :
 - supposons que les personnes qui quittent le service ne meurent jamais dans les 90 jours suivants
 - Formellement, tous exclus vivants à 90j



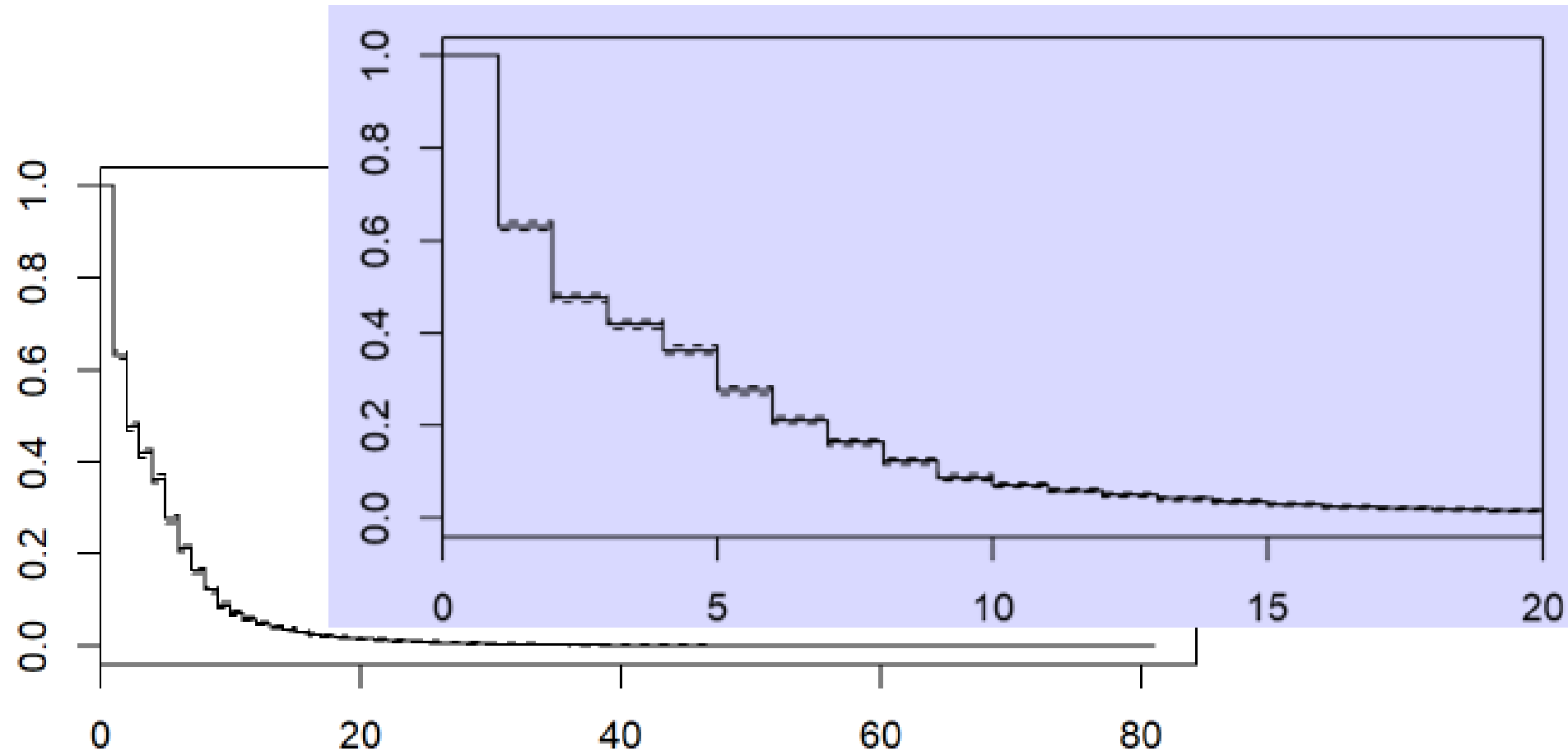
Commentaire :
Approche sans doute fautive : on ne sait pas réellement ce qu'il s'est passé durant 90j. Néanmoins, résultat plus vraisemblable que précédemment.
Attitude correcte : simple proportion de vivants sous un délai très court, ou au terme du séjour

Ex 4 – analyse de survie en l'absence de censure

- Situation :
 - Séjours hospitaliers
 - Événement = sortie de l'hôpital
 - Événement toujours observé (=1 toujours)
 - Délai = durée de séjour dans tous les cas
 - *De facto*, aucune censure dans cet exemple
- Objectif :
 - Découvrir les déterminants de la durée de séjour
 - Parmi : âge, sexe, type de séjour (obs, cancer, etc.)
- Méthodes : 2 possibilités, toutes deux licites
 - Durée de séjour = variable quantitative
=> régression linéaire multiple
 - Sortie = événements temps-dépendant
=> modèle de Cox (avec 0% de censures)

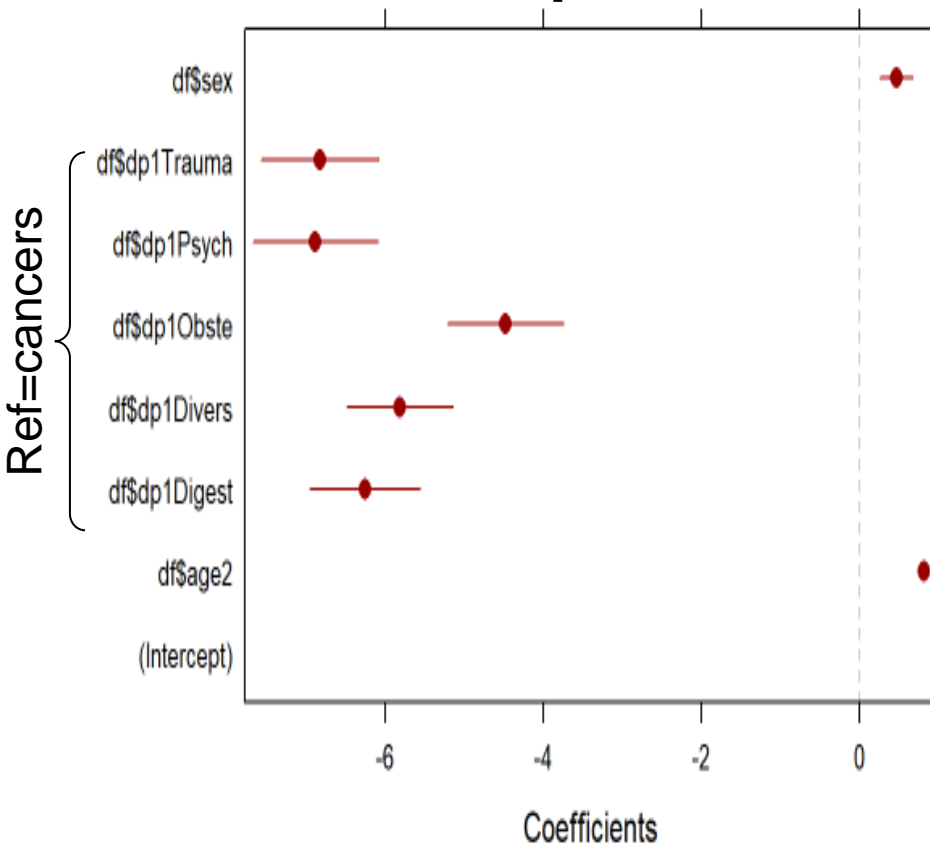
Ex 4 – analyse de survie en l'absence de censure

- Courbe de Kaplan-Meier (100% d'événements) :

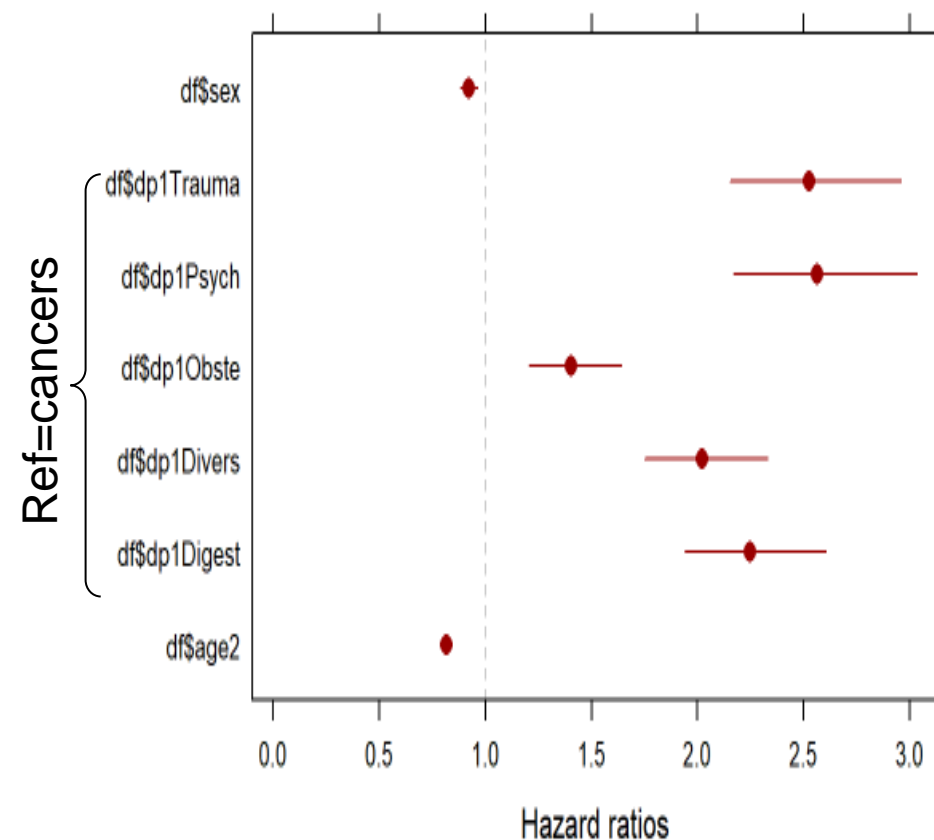


Ex 4 – analyse de survie en l'absence de censure

Régression linéaire multiple



Modèle de Cox

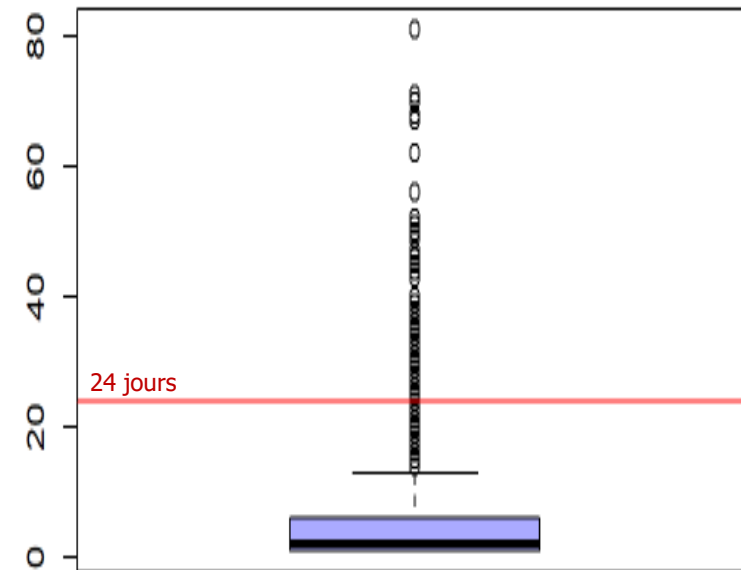
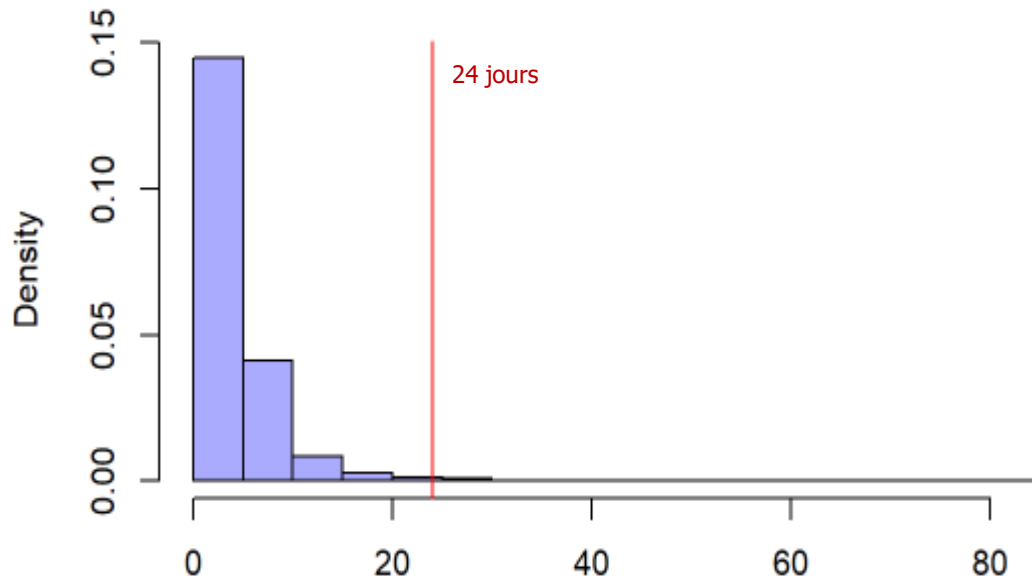


Commentaire :

RLM : interpréter *coeffs* en nombre de jours. Cox : interpréter e^{coeffs} en HR. => aspect « en miroir »

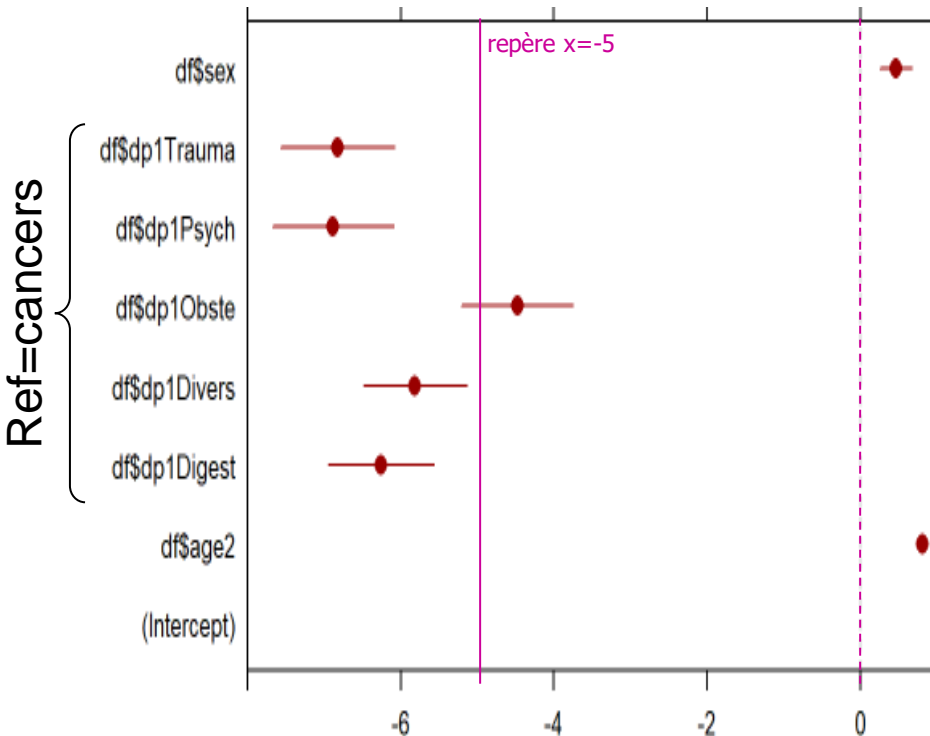
Ex 4 – analyse de survie en l'absence de censure

- Représentation graphique de la durée de séjour
- Quelques séjours extrêmes, jusqu'à 81 jours
- 99^{ème} percentile = 24 jours
- Par la suite, suppression arbitraire des séjours de plus de 24 jours (on passera de 10443 à 10339 séjours)

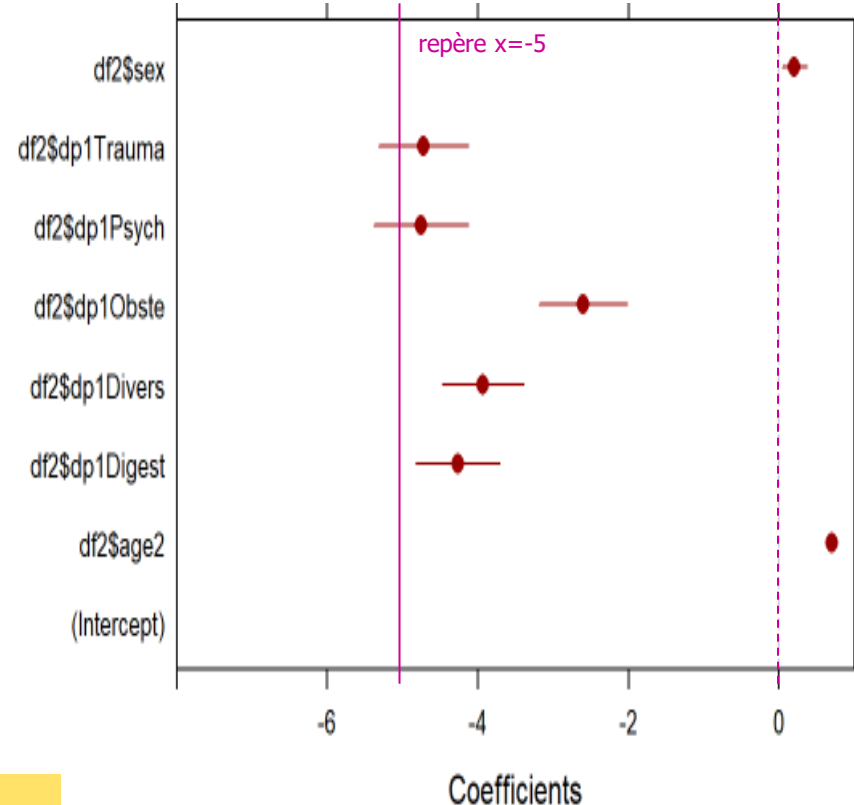


Ex 4 – analyse de survie en l'absence de censure

Régression linéaire multiple avant



RLM après suppression des 1% plus extrêmes

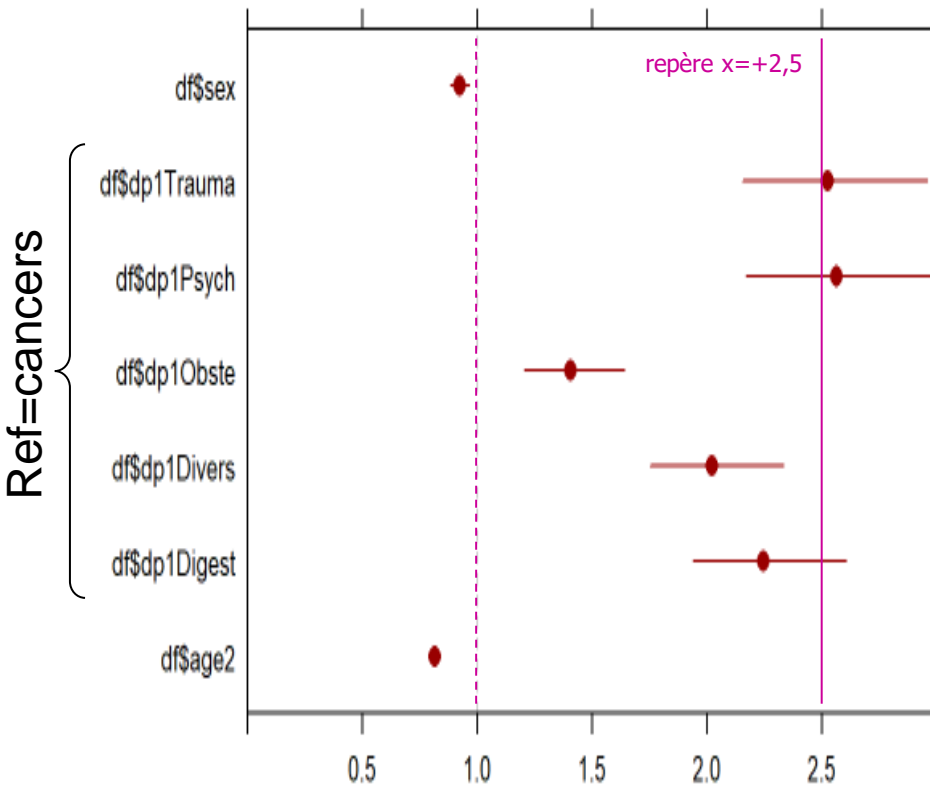


Commentaire :

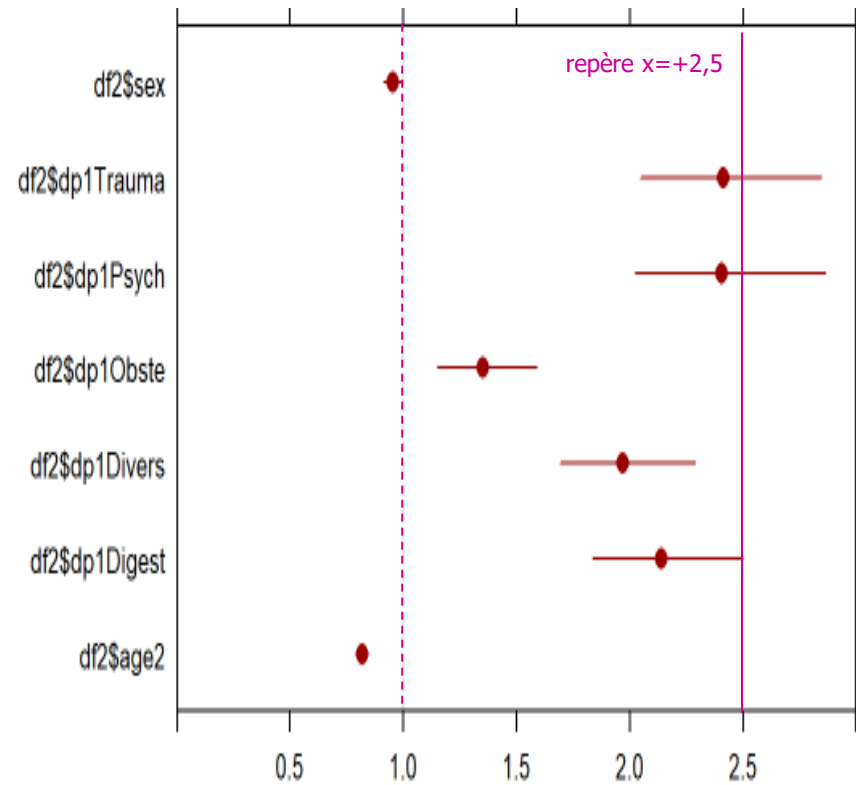
Les coefficients de la régression linéaire changent beaucoup => sensible aux valeurs extrêmes

Ex 4 – analyse de survie en l'absence de censure

Modèle de Cox avant



Cox après suppression des 1% plus extrêmes



Commentaire :

Les hazard ratios du modèle de Cox changent peu => peu sensible aux valeurs extrêmes.
L'analyse de survie est une approche « robuste » pour les durées, même lorsque toutes connues.