

Introduction au logiciel R

Capsule: Courbe de Kaplan-Meier

Analyse de données d'incidence avec le logiciel R

Auteur: Benoît Mâsse

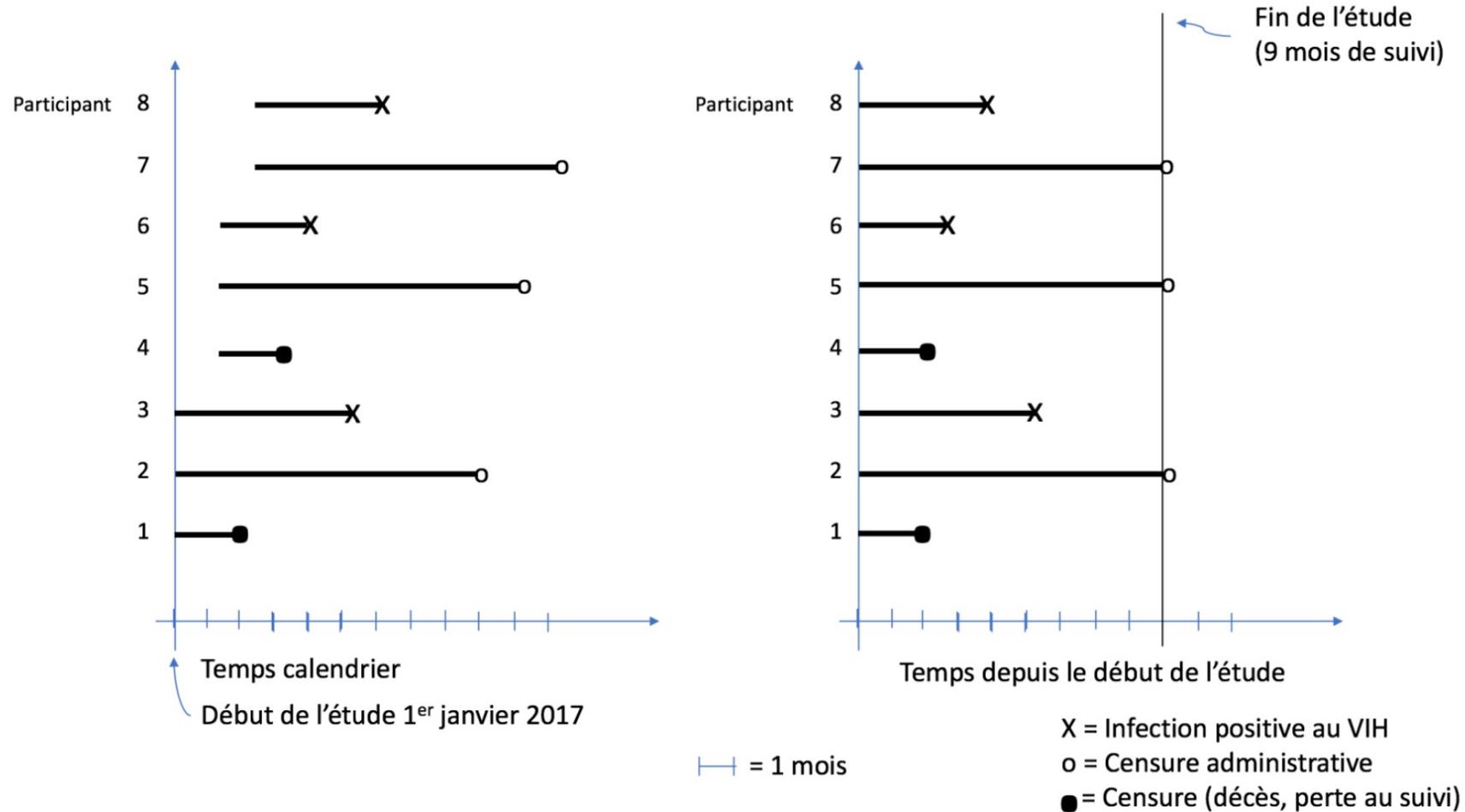
Incidence

- Bien souvent en recherche clinique, on s'intéresse à la fois à l'occurrence d'un évènement durant le suivi mais aussi au moment de l'occurrence de cet évènement. Quelques exemples:
 - Vaccin contre l'infection au VIH: on s'intéresse à réduire le risque d'acquisition d'un virus. Après l'administration d'un vaccin, on suivra les participants en les testant durant le suivi pour la présence du VIH. Pour les participants qui auront acquis le VIH, on notera le moment de l'infection (ex: 6 mois après la vaccination).
 - Traitement contre le cancer: on s'intéresse à réduire le risque de décès causé par un cancer. Après le début des traitements, on suivra les patients et on notera la date de décès des patients qui décèderont durant le suivi. Avec cette date, on pourra déterminer la survie des patients après le début des traitements (ex: 3 ans après le début des traitements)

Censure

- Dans la plupart des contextes, le suivi des participants ne peut se faire indéfiniment et les études ont des périodes de suivi maximum (ex: 5 ans). À la fin de ce suivi, il est possible que l'occurrence de l'évènement n'a pas eu lieu.
- Le suivi d'un participant a un début et une fin.
 - Si la fin du suivi arrive avant l'occurrence de l'évènement: censure à droite
 - La censure à droite est celle qui est la plus fréquente en recherche clinique
 - D'autres types de censure: censure à gauche et censure par intervalle

Exemple: Vaccin contre le VIH



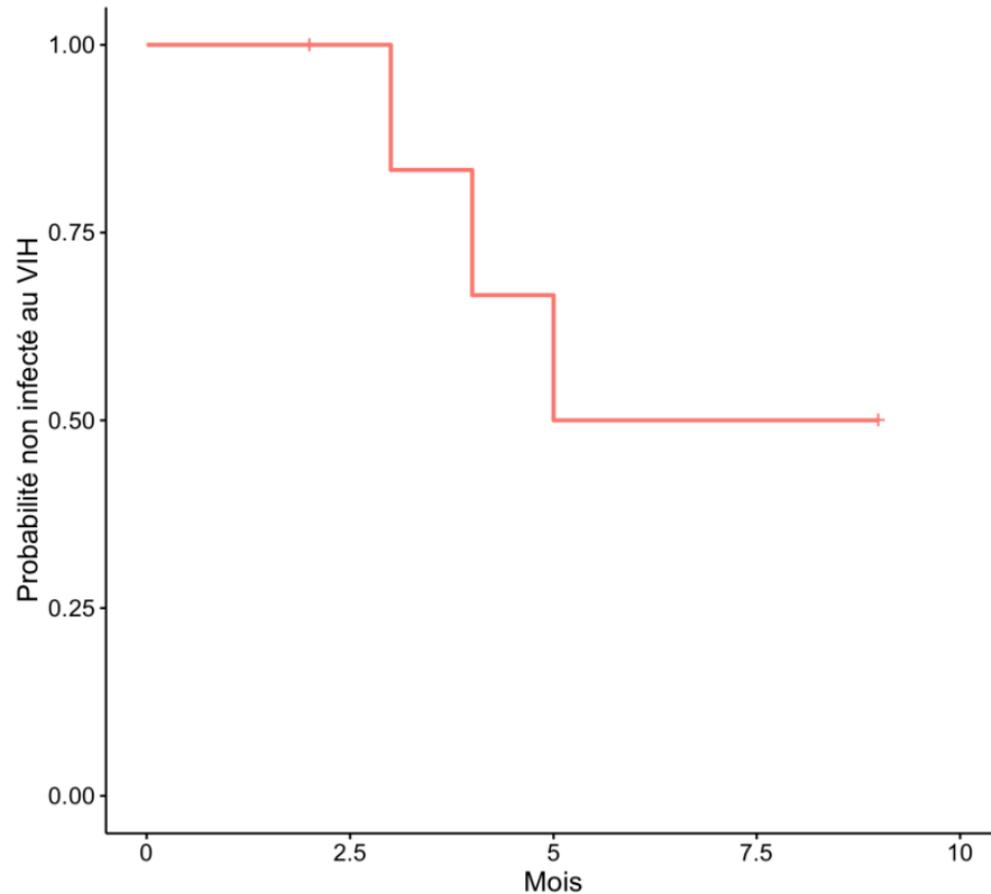
Description du suivi des 8 participants

Exemple: Vaccin contre le VIH

Participants	Date de randomisation (= date de vaccination)	Date du premier test positif	Remarques	Temps de suivi	Censure administrative	Autre censure (décès, perte au suivi)	Test positif Infection au VIH
1	1er janvier 2017		Participant décédé le 1er mars 2017	2 mois		x	
2	1er janvier 2017			9 mois	x		
3	1er janvier 2017	1er juin 2017		5 mois			x
4	1er février 2017		Participant perdu au suivi le 1er avril 2017	2 mois		x	
5	1er février 2017			9 mois	x		
6	1er février 2017	1er mai 2017		3 mois			x
7	1er mars 2017			9 mois	x		
8	1er mars 2017	1er juillet 2017		4 mois			x

Participant	Temps de suivi avant l'infection ou la censure (mois)
1	2+
2	9+
3	5
4	2+
5	9+
6	3
7	9+
8	4

Exemple: Vaccin contre le VIH

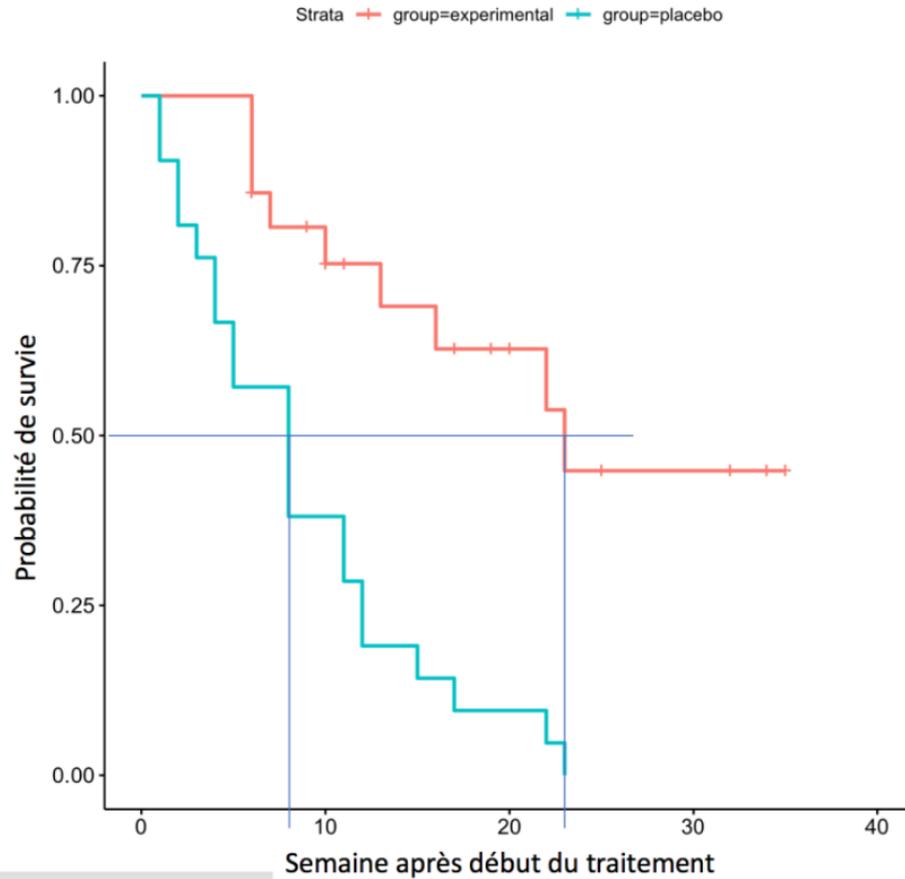


Courbe de Kaplan-Meier pour les 8 participants

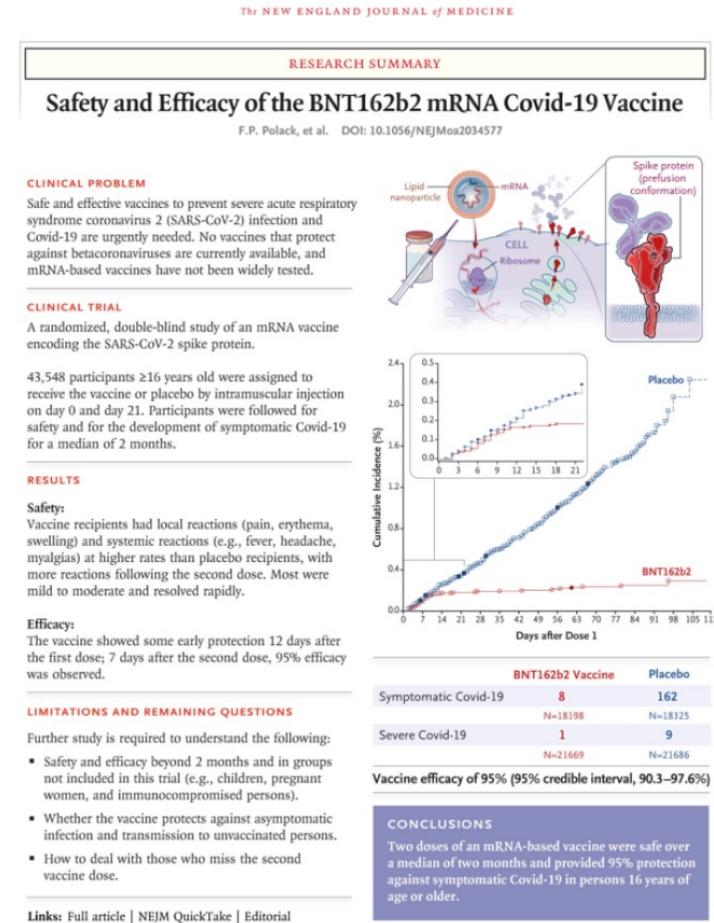
Courbe de Kaplan-Meier : fonction en escalier

- Axe des X: temps en mois depuis l'administration du vaccin
- Axe des Y: proportion de patients qui n'ont pas acquis une infection au VIH
- Au moment de la vaccination 8/8 (100%) participants sont non infectés.
- À 3 mois, un premier participant s'infecte mais il ne reste que 6 participants puisque que 2 participants sont censurés à 2 mois (un décès et une perte au suivi): 5/6 (83.3%)
- L'escalier s'arrête à 9 mois car l'étude avait une durée de 9 mois
- On peut déduire le temps médian d'infection: ≥ 5 mois

Autres exemples



Courbes Kaplan-Meier par groupe



Courbes K-M pour le vaccin de Pfizer-BioNTech

Format des données pour analyse avec R

Pour des données d'incidence avec censure à droite, on utilise 2 variables:

- Temps : la durée du suivi entre le début du suivi et le temps de censure ou de l'occurrence de l'évènement
- Indicateur 0/1 de l'évènement:
« 1 » si l'occurrence de l'évènement a eu lieu ou
« 0 » si le suivi est censuré

	(mois)
1	2+
2	9+
3	5
4	2+
5	9+
6	3
7	9+
8	4

Identifiant	Temps	Censure
1	2	0
2	9	0
3	5	1
4	2	0
5	9	0
6	3	1
7	9	0
8	4	1

Production des courbes de Kaplan-Meier avec R

- Monte les packages *quantreg* et *survival*

```
> library(quantreg)
> library(survival)
> library(ggplot2)
> library(ggsurvfit)
> library(gfsummary)
```

- Monte la base de données *uis*
- *uis*: essai clinique randomisé sur le traitement de la toxicomanie

```
> data(uis)
> attach(uis)
> help(uis)
```

Production des courbes de Kaplan-Meier avec R

- Commande qui remplace les 0 et 1 par des étiquettes lisibles

```
> TREATc <- ifelse(TREAT==0,"Traitement court","Traitement long")
```

- Commande *survfit()* pour le calcul des courbes de K-M

- Courbe pour tous les participants de l'étude

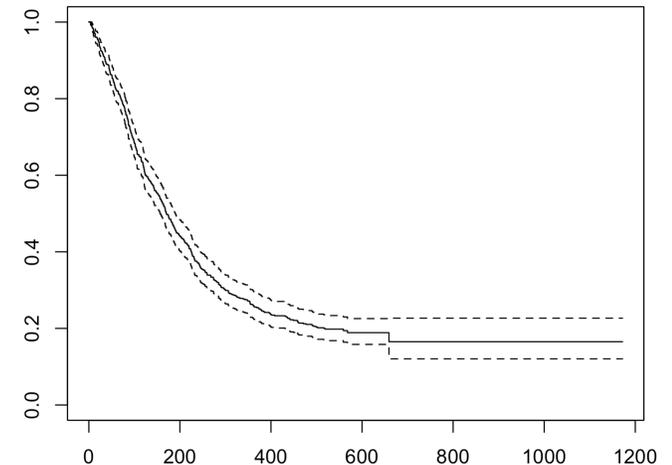
```
> f1 <- survfit(Surv(TIME, CENSOR) ~ 1)
```

```
> print(f1)
```

```
> plot(f1)
```

```
Call: survfit(formula = Surv(TIME, CENSOR) ~ 1)
```

```
      n events median 0.95LCL 0.95UCL  
[1,] 575    464   170    156    190
```



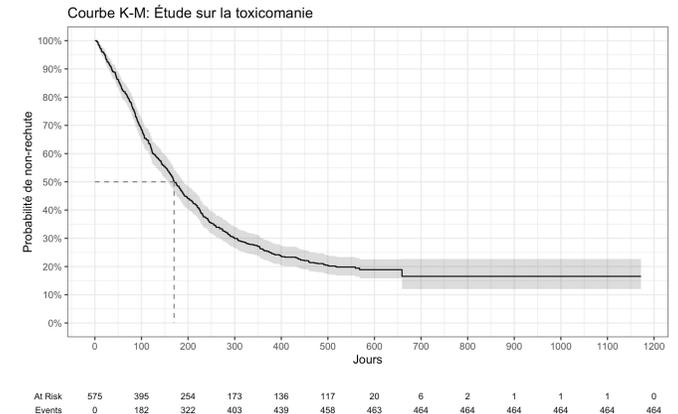
Production des courbes de Kaplan-Meier avec R

- Monte les packages `ggsurvfit`, `gtsummary`, et `ggplot2`

```
> library(ggplot2)
> library(ggsurvfit)
> library(gtsummary)
```

- Création de courbe avec étiquette, 95% IC, et table

```
> ggsurvfit(f1) +
  labs(title="Courbe K-M: Étude sur la toxicomanie",x="Jours",y="Probabilité de non-rechute") +
  add_confidence_interval() +
  add_risktable() +
  add_quantile(y_value = 0.5, color = "gray50") +
  labs(x="Jours", y="Probabilité de non-rechute") +
  scale_y_continuous(label = scales::percent, breaks = seq(0, 1, by = 0.1))+
  scale_x_continuous(breaks=seq(0,1200,by=100))
```



Production des courbes de Kaplan-Meier avec R

- Création de courbes K-M pour plusieurs groupes

```
> f2 <- survfit2(Surv(TIME, CENSOR) ~ TREATc)
```

```
> print(f2)
```

```
Call: survfit(formula = Surv(TIME, CENSOR) ~ TREATc)
```

	n	events	median	0.95LCL	0.95UCL
TREATc=Traitement court	289	239	142	120	166
TREATc=Traitement long	286	225	202	180	232

```
> ggsvrfits(f2) +
```

```
labs(title="Courbes K-M par traitement: Étude sur la toxicomanie",x="Jours",y="Probabilité de non-rechute") +
```

```
add_confidence_interval() +
```

```
add_risktable() +
```

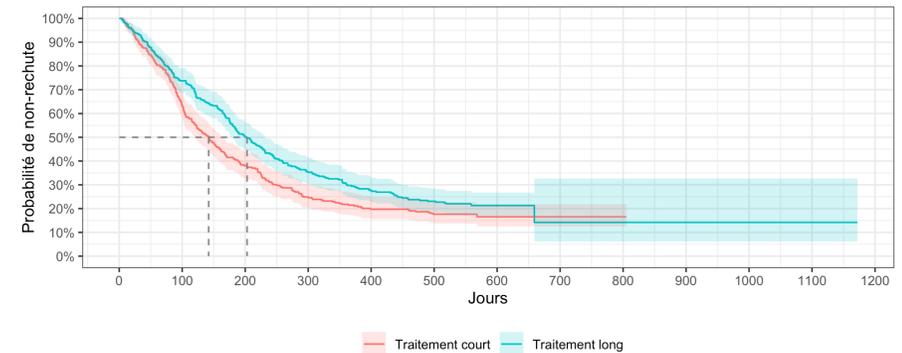
```
add_quantile(y_value = 0.5, color = "gray50") +
```

```
labs(x="Jours", y="Probabilité de non-rechute") +
```

```
scale_y_continuous(label = scales::percent, breaks = seq(0, 1, by = 0.1))+
```

```
scale_x_continuous(breaks=seq(0,1200,by=100))
```

Courbes K-M par traitement: Étude sur la toxicomanie



Production des courbes de Kaplan-Meier avec R

- Détermination de différents points sur les courbes de K-M

```
> tbl_survfit(f2,probs=0.4,label_header="Nombre de jours pour atteindre le 40ième percentile")
```

Characteristic	Nombre de jours pour atteindre le 40 ^{ième} percentile
TREATc	
Traitement court	104 (98, 124)
Traitement long	166 (136, 184)

```
> tbl_survfit(f2,time=200,label_header="Percentile au jour 200")
```

Characteristic	Percentile au jour 200
TREATc	
Traitement court	38% (33%, 44%)
Traitement long	50% (45%, 56%)

Comparaison de 2 courbes de K-M: test du logrank

- Test du logrank avec la fonction *survdiff*

```
> lr1 <- survdiff(Surv(TIME, CENSOR) ~ TREATc)
> print(lr1)
```

call:

```
survdiff(formula = Surv(TIME, CENSOR) ~ TREATc)
```

	N	Observed	Expected	(O-E) ² /E	(O-E) ² /V
TREATc=Traitement court	289	239	212	3.52	6.53
TREATc=Traitement long	286	225	252	2.96	6.53

Chisq= 6.5 on 1 degrees of freedom, p= 0.01

Comparaison de 2 courbes de K-M: test du logrank

- Test du logrank avec la fonction *coxph*

```
> lr2 <- coxph(Surv(TIME, CENSOR) ~ TREATc)
```

```
> summary(lr2)
```

```
Call:
```

```
coxph(formula = Surv(TIME, CENSOR) ~ TREATc)
```

```
n= 575, number of events= 464
```

	coef	exp(coef)	se(coef)	z	Pr(> z)
TREATcTraitement long	-0.23744	0.78864	0.09308	-2.551	0.0107 *

```
---
```

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

	exp(coef)	exp(-coef)	lower .95	upper .95
TREATcTraitement long	0.7886	1.268	0.6571	0.9465

```
Concordance= 0.538 (se = 0.012 )
```

```
Likelihood ratio test= 6.51 on 1 df, p=0.01
```

```
wald test = 6.51 on 1 df, p=0.01
```

```
Score (logrank) test = 6.54 on 1 df, p=0.01
```

Analyses multivariées avec modèle de Cox

- Ex: Ajustement pour le score de dépression BECK

```
> lr3 <- coxph(Surv(TIME, CENSOR) ~ TREATc + BECK)
> summary(lr3)
```

Call:

```
coxph(formula = Surv(TIME, CENSOR) ~ TREATc + BECK)
```

```
n= 575, number of events= 464
```

	coef	exp(coef)	se(coef)	z	Pr(> z)	
TREATcTraitement long	-0.236335	0.789516	0.093079	-2.539	0.0111	*
BECK	0.010280	1.010333	0.004849	2.120	0.0340	*

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

	exp(coef)	exp(-coef)	lower .95	upper .95
TREATcTraitement long	0.7895	1.2666	0.6579	0.9475
BECK	1.0103	0.9898	1.0008	1.0200

Concordance= 0.559 (se = 0.014)

Likelihood ratio test= 10.94 on 2 df, p=0.004

Wald test = 10.96 on 2 df, p=0.004

Score (logrank) test = 11 on 2 df, p=0.004