

## Correction TP Analyse de Survie : Modèle de Cox

Importation des packages et données :

```
library(survival)
library(survminer)
```

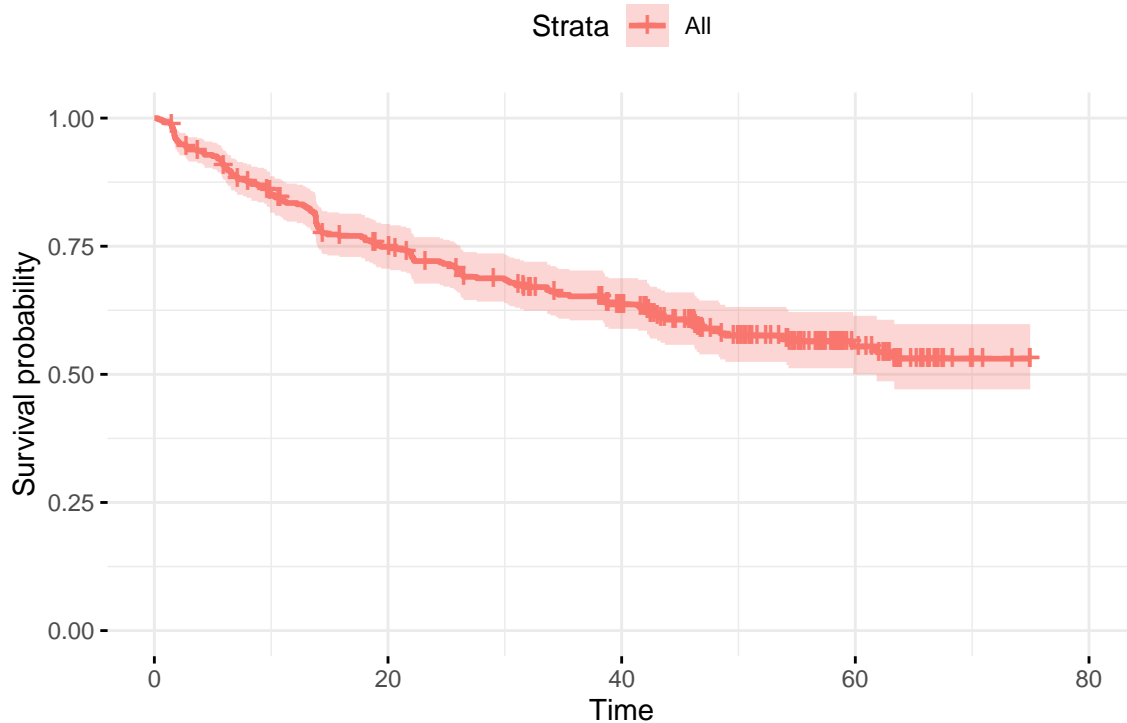
```
data(diabetic)
head(diabetic)
```

```
##   id laser age  eye trt risk  time status
## 1  5 argon  28 left  0   9 46.23      0
## 2  5 argon  28 right 1   9 46.23      0
## 3 14 xenon  12 left  1   8 42.50      0
## 4 14 xenon  12 right 0   6 31.30      1
## 5 16 xenon   9 left  1  11 42.27      0
## 6 16 xenon   9 right 0  11 42.27      0
```

### En pratique

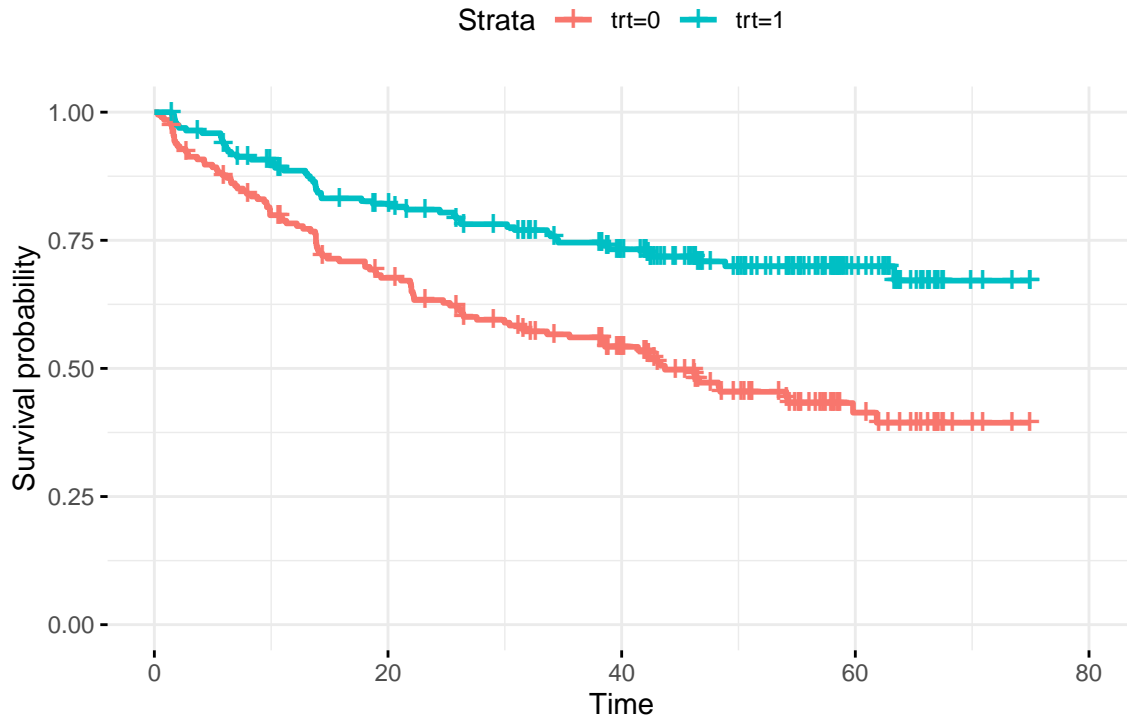
On regarde dans un premier l'allure de la courbe de survie des données globales, sans utilisation des variables explicatives.

```
fit.global <- survfit(Surv(time,status) ~ 1,data = diabetic)
ggsurvplot(fit.global, ggtheme = theme_minimal())
```



On compare ensuite les courbes de survie pour les patients traités ou non

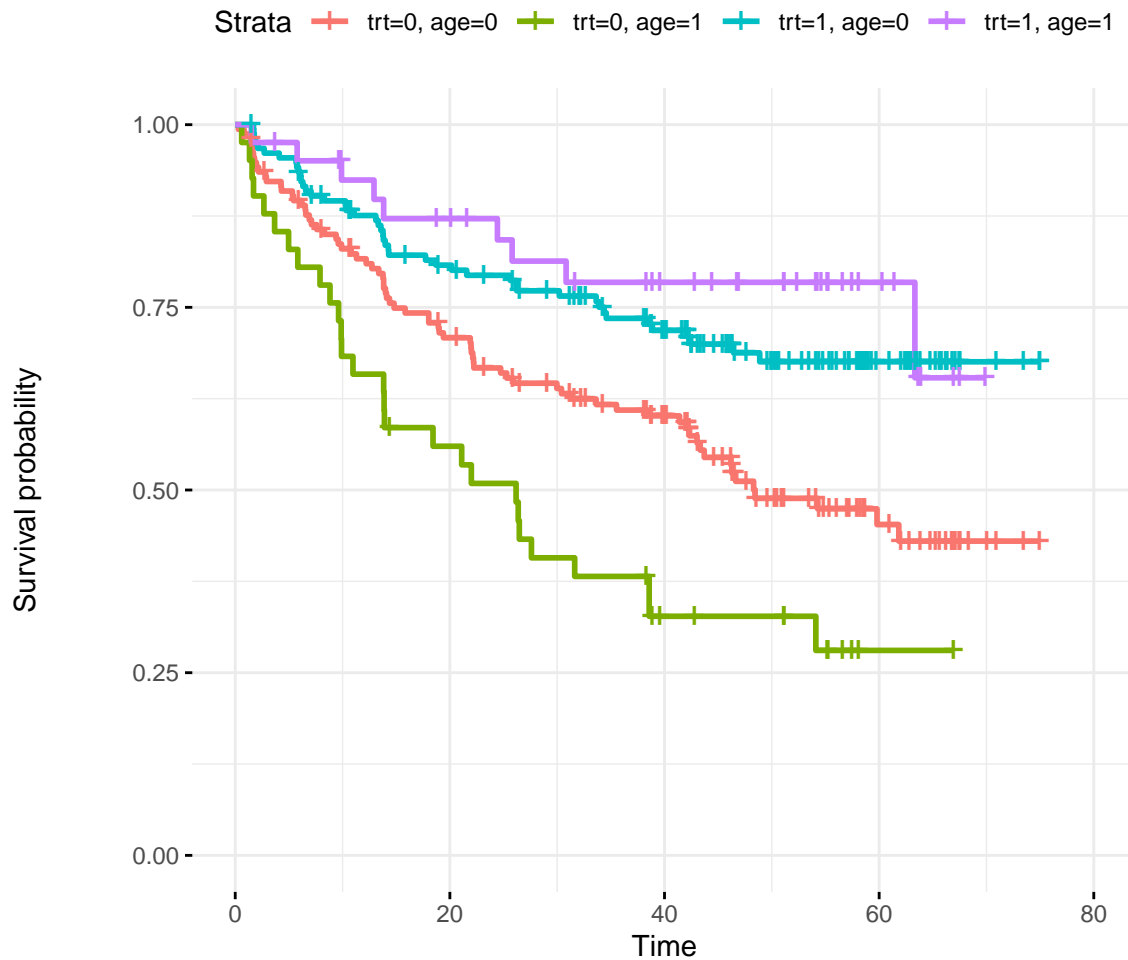
```
fit.trait <- survfit(Surv(time,status) ~ trt,data = diabetic)
ggsurvplot(fit.trait, ggtheme = theme_minimal())
```



Comme on pourrait s'y attendre, les patients traités ont une courbe de survie plus élevée que celle des patients non traité, et cela quelque soit le temps de survie. On veut maintenant regarder la même chose mais en prenant compte l'âge des patients que l'on divise en deux catégories : plus ou moins 35 ans.

```
diabetic$age <- as.numeric(diabetic$age >= 35)

fit.trait.age <- survfit(Surv(time,status) ~ trt + age,data = diabetic)
ggsurvplot(fit.trait.age, risk.table = TRUE, ggtheme = theme_minimal())
```



Number at risk						
Strata	trt=0, age=0	156	104	72	21	0
	trt=0, age=1	41	22	10	1	0
	trt=1, age=0	156	117	83	25	0
	trt=1, age=1	41	32	23	8	0
		0	20	40	60	80
		Time				

Là où les courbes de survie des patients non traités semblent clairement différentes, on peut se poser la question de l'influence de l'âge sur les patients traités. A noter que la différence entre les courbes de survie pour les patients traités de plus ou moins de 35 ans peut s'expliquer par la différence du nombre d'individus dans les deux populations. On peut donc effectuer un test de Log-rank pour déterminer si les courbes de survie ne sont pas identiques en fonction de l'âge

```
survdif(Surv(time,status) ~ trt + age, data = subset(diabetic,trt==0))
```

```
## Call:
## survdif(formula = Surv(time, status) ~ trt + age, data = subset(diabetic,
##   trt == 0))
##
```

```
##               N Observed Expected (O-E)^2/E (O-E)^2/V
## trt=0, age=0 156         73      83.1      1.23      6.97
## trt=0, age=1  41         28      17.9      5.70      6.97
##
## Chisq= 7 on 1 degrees of freedom, p= 0.008
survdifff(Surv(time,status) ~ trt + age, data = subset(diabetic,trt==1))
```

```
## Call:
## survdiff(formula = Surv(time, status) ~ trt + age, data = subset(diabetic,
##   trt == 1))
##
##               N Observed Expected (O-E)^2/E (O-E)^2/V
## trt=1, age=0 156         45      42.5      0.151      0.706
## trt=1, age=1  41          9      11.5      0.555      0.706
##
## Chisq= 0.7 on 1 degrees of freedom, p= 0.4
```

On constate alors que la  $p$  valeur est particulièrement élevée pour les patients traités, ce qui indique donc fortement que l'âge ne semble pas influencer la qualité du traitement.

On souhaite maintenant savoir si ces dernières populations ne peuvent être ajustées à un modèle de Cox.

```
fit.cox <- coxph(Surv(time,status) ~ trt + age, data = diabetic)
summary(fit.cox)
```

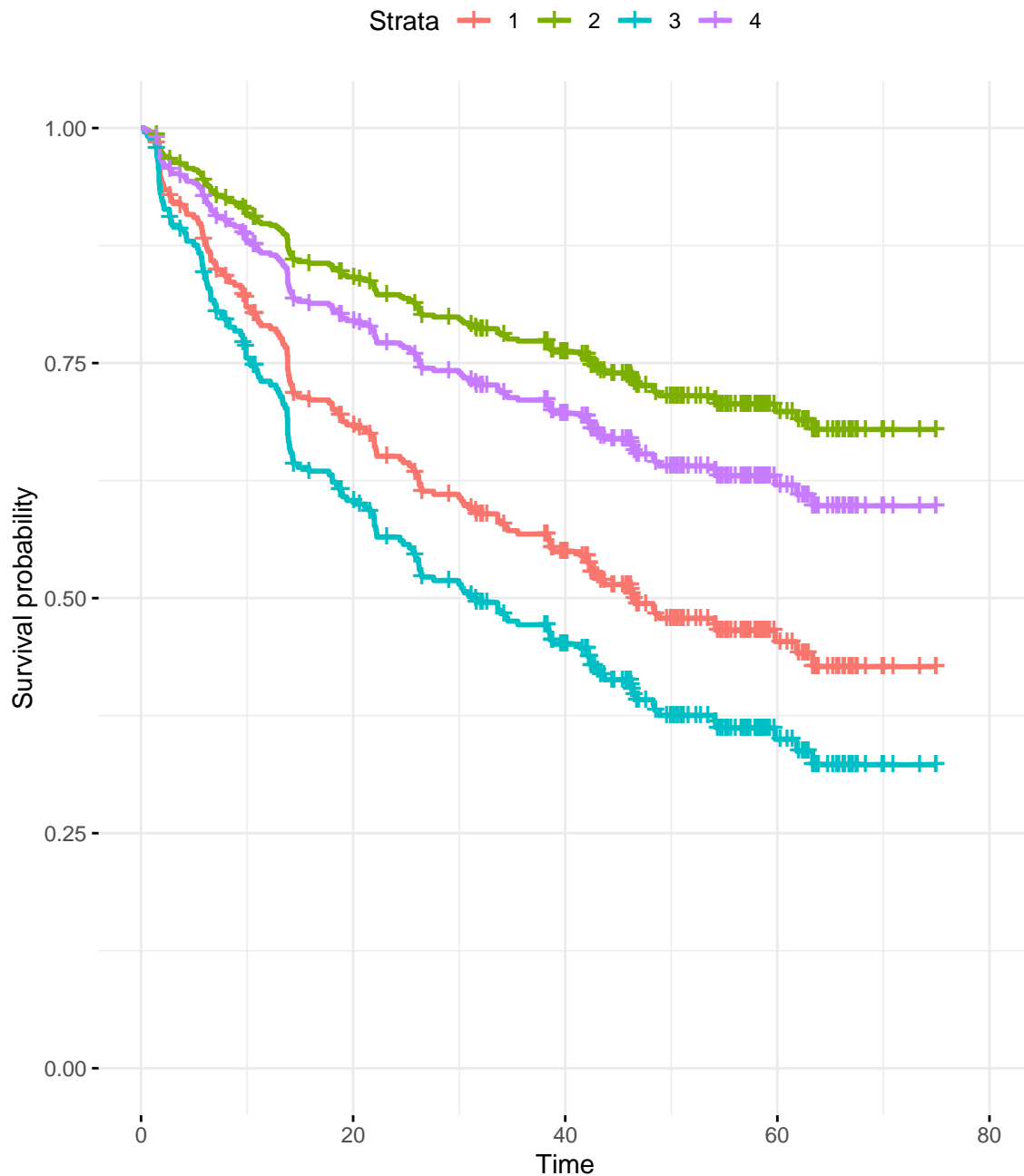
```
## Call:
## coxph(formula = Surv(time, status) ~ trt + age, data = diabetic)
##
## n= 394, number of events= 155
##
##      coef exp(coef) se(coef)      z Pr(>|z|)
## trt -0.7886   0.4545   0.1690 -4.665 3.08e-06 ***
## age  0.2848   1.3295   0.1888  1.508  0.132
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      exp(coef) exp(-coef) lower .95 upper .95
## trt    0.4545    2.2003    0.3263    0.633
## age    1.3295    0.7522    0.9182    1.925
##
## Concordance= 0.601 (se = 0.022 )
## Likelihood ratio test= 24.53 on 2 df,  p=5e-06
## Wald test               = 23.38 on 2 df,  p=8e-06
## Score (logrank) test = 24.47 on 2 df,  p=5e-06
```

A première vue, les test d'adéquations semblent satisfaisants et les résultats indiquent toujours que la composante d'âge n'est pas significative pour le modèle. Les  $p$  valeurs des tests d'hypothèse initiale  $H_0 : \beta_{trt} = \beta_{age} = 0$  sont en effet quasi-nulles. On peut à présent de nouveau regarder les courbes de survie, mais cette fois-ci par extrapolation du modèle de Cox.

```
new.df <- expand.grid(trt = c(0, 1), age = c(0,1))
new.df
```

```
##   trt age
## 1   0   0
## 2   1   0
## 3   0   1
```

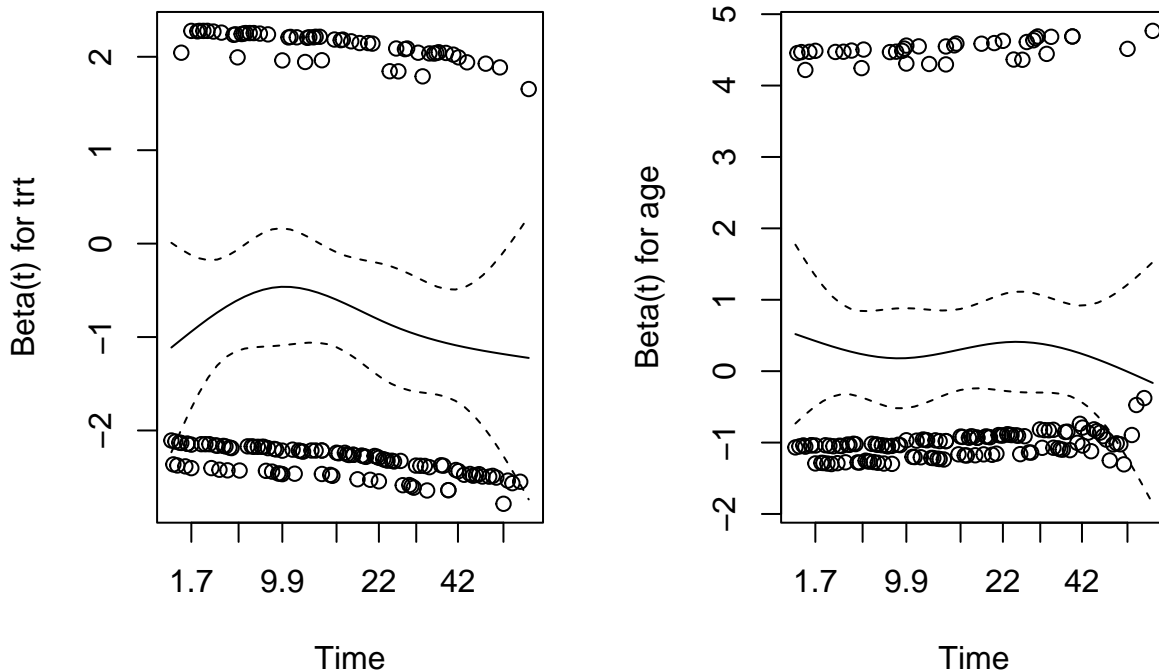
```
## 4 1 1
new.fit.cox <- survfit(fit.cox, data = diabetic, newdata = new.df)
ggsurvplot(new.fit.cox, ggtheme = theme_minimal(), conf.int=FALSE)
```



De nouveau, seule le traitement semble pouvoir influencer sur la survie des patients mais on peut s'interroger sur la validité du modèle de Cox, en particulier sur la variable traitement. Une méthode peut donc consister à calculer les résidus de Schoenfeld (voir cours)

```
test.cox <- cox.zph(fit.cox)
par(mfrow=c(1,2))
```

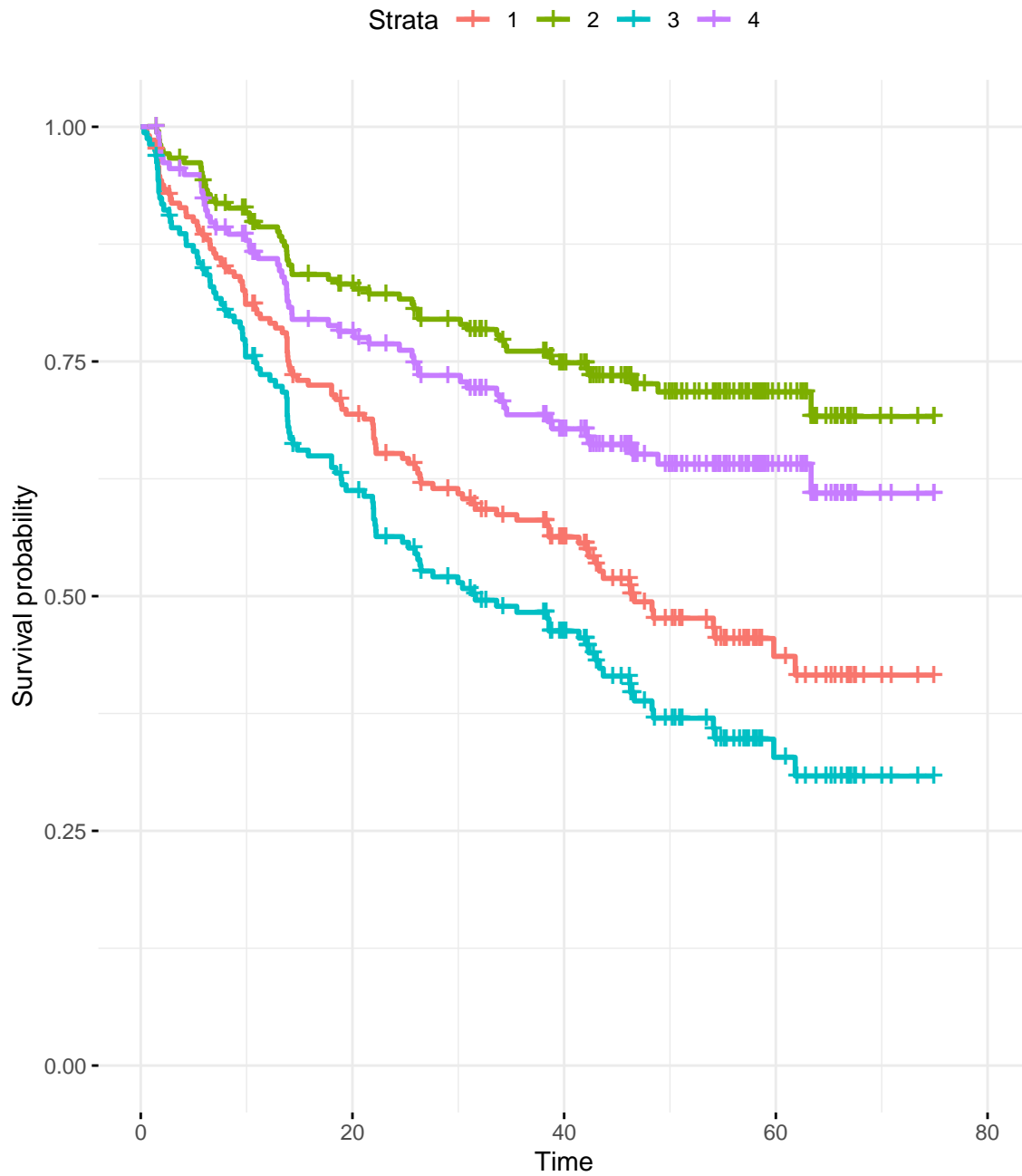
```
plot(test.cox)
```



Là où la variable d'âge ne semble pas contraindre l'hypothèse de proportionnalité, la courbe pour la variable traitement est loin de l'abscisse. Une solution serait d'étendre le modèle, en proposant par exemple, un modèle de Cox stratifié par rapport à la variable traitement.

```
fit.cox.strat <- coxph(Surv(time,status) ~ strata(trt) + age, data = diabetic)
summary(fit.cox.strat)
```

```
## Call:
## coxph(formula = Surv(time, status) ~ strata(trt) + age, data = diabetic)
##
##      n= 394, number of events= 155
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## age 0.2935      1.3412   0.1891 1.553   0.121
##
##      exp(coef) exp(-coef) lower .95 upper .95
## age      1.341      0.7456   0.9259   1.943
##
## Concordance= 0.524  (se = 0.018 )
## Likelihood ratio test= 2.29  on 1 df,  p=0.1
## Wald test               = 2.41  on 1 df,  p=0.1
## Score (logrank) test = 2.43  on 1 df,  p=0.1
new.fit.cox.strat <- survfit(fit.cox.strat, data = diabetic, newdata = new.df)
ggsurvplot(new.fit.cox.strat, ggtheme = theme_minimal(), conf.int=FALSE)
```



Note : on remarquera que le passage à un modèle stratifié change les  $p$  valeurs pour les tests avec hypothèse initiale  $H_0 : \beta_{age} = 0$  qui sont à présent supérieurs à 0.5.