Établissement et comparaison des courbes de survie

Jeudi 18 janvier 2024 – VetAgro Sup

Karine Chalvet-Monfray

Objectif de la formation

- Comprendre le principe d'**établissement** de courbes de survie
- Comprendre le principe de **comparaison** de courbes de survie

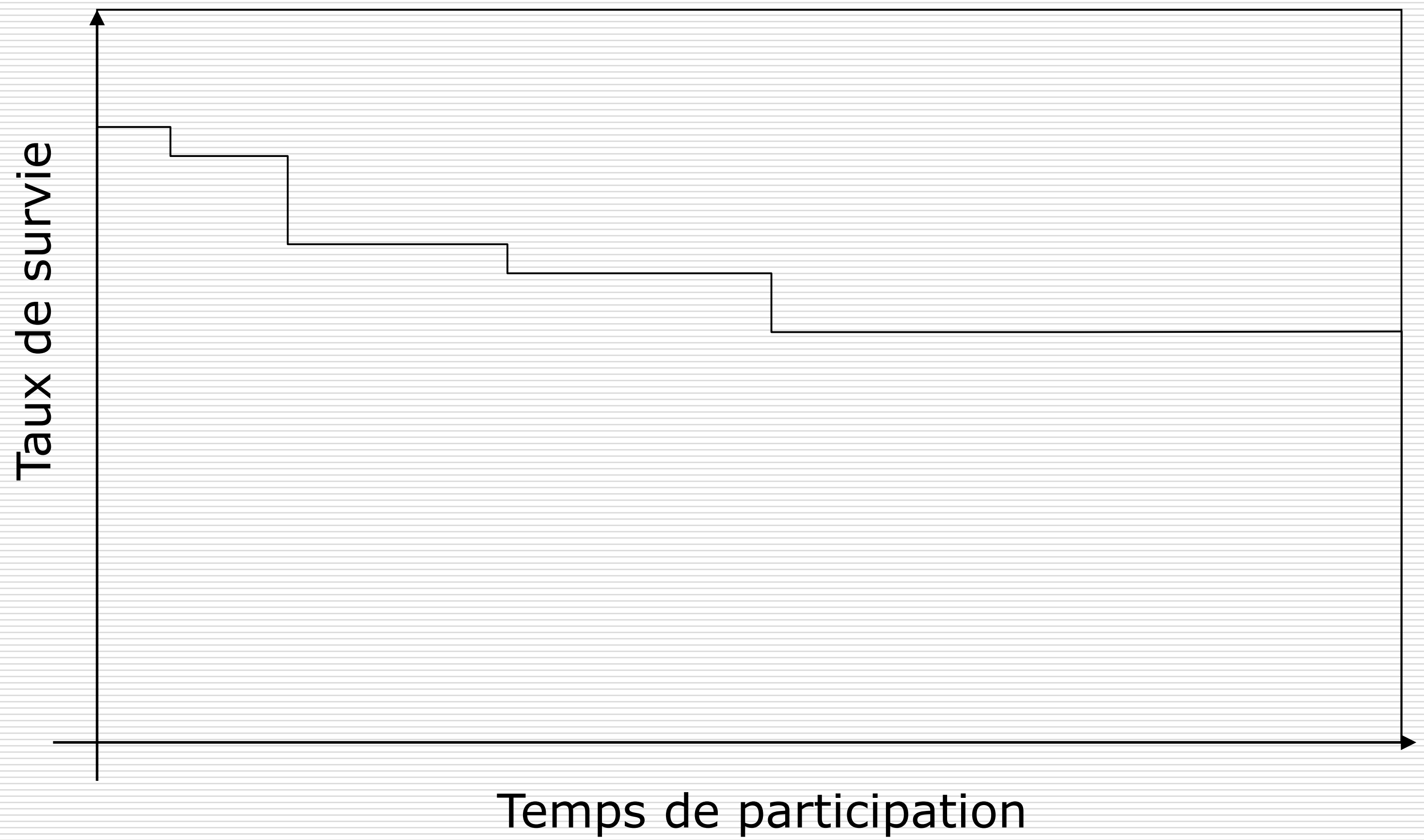
Intérêt des courbes de survie

- Décrire la survie dans l'échantillon ;
- Estimer la survie dans la population (prédictif) ;
- Comparer les survies selon les groupes (recherche thérapeutique et épidémiologique).

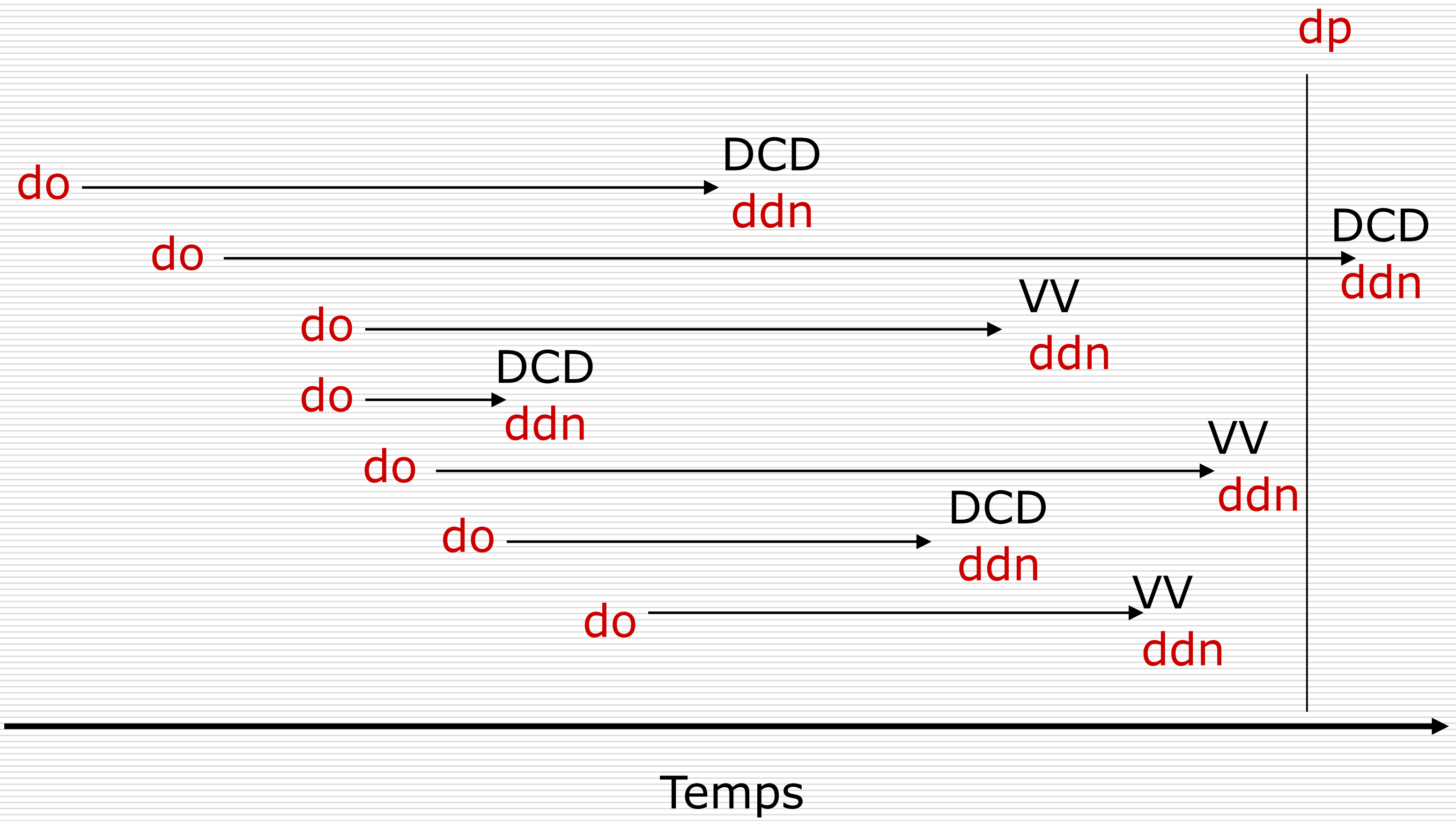
Données de survie

- Temps de participation : délai entre deux dates. La date finale n’est pas forcément la date de décès mais e.g., date de la première récidive, des complications,…
- Particularité de données de survie : possibilité de censure
- Taux de survie (Survie) : probabilité d’être vivant ≠1-Taux de mortalité
- La courbe de survie : Représentation du **taux de survie** en fonction du **temps de participation**

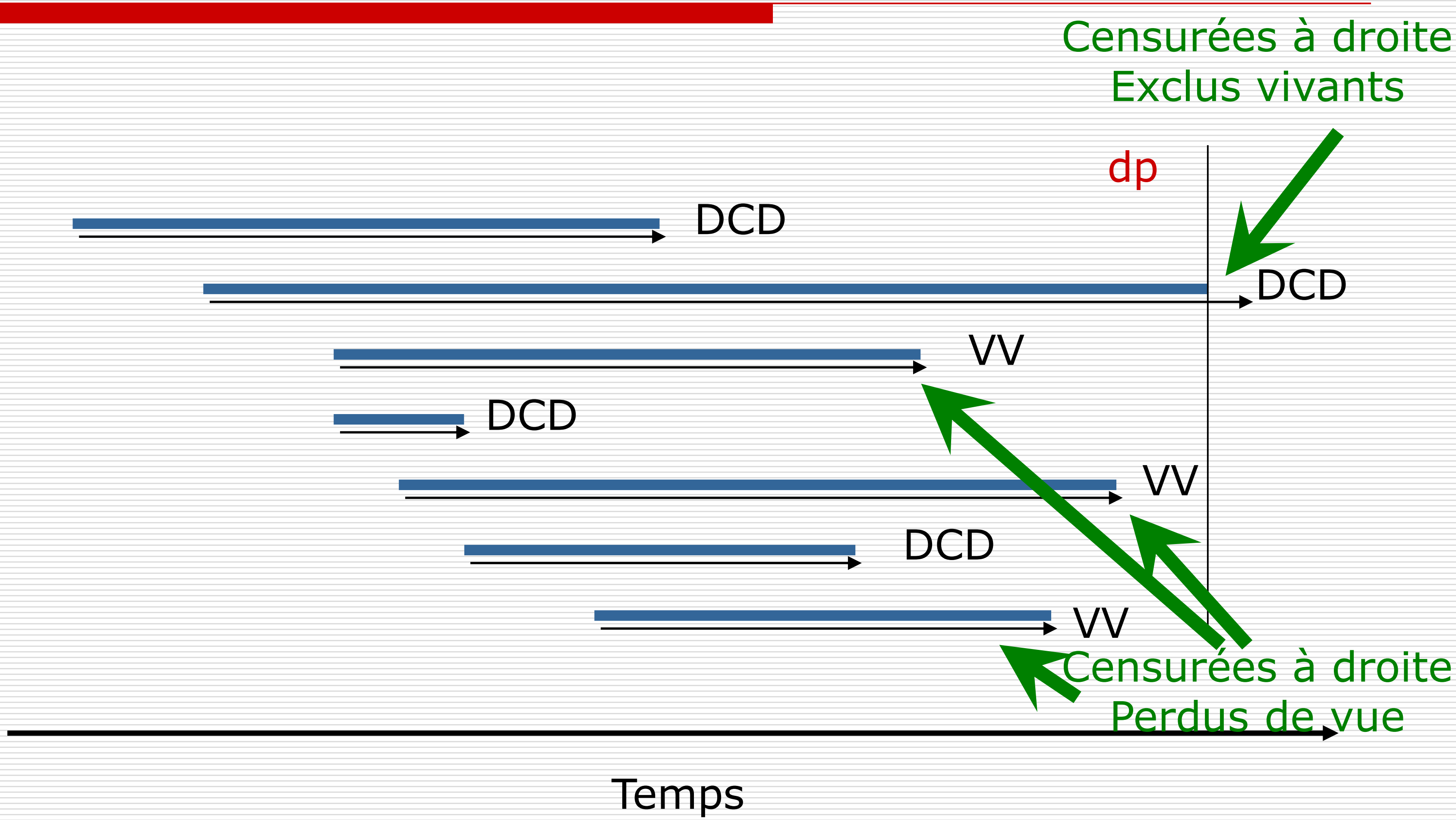
Courbe de survie



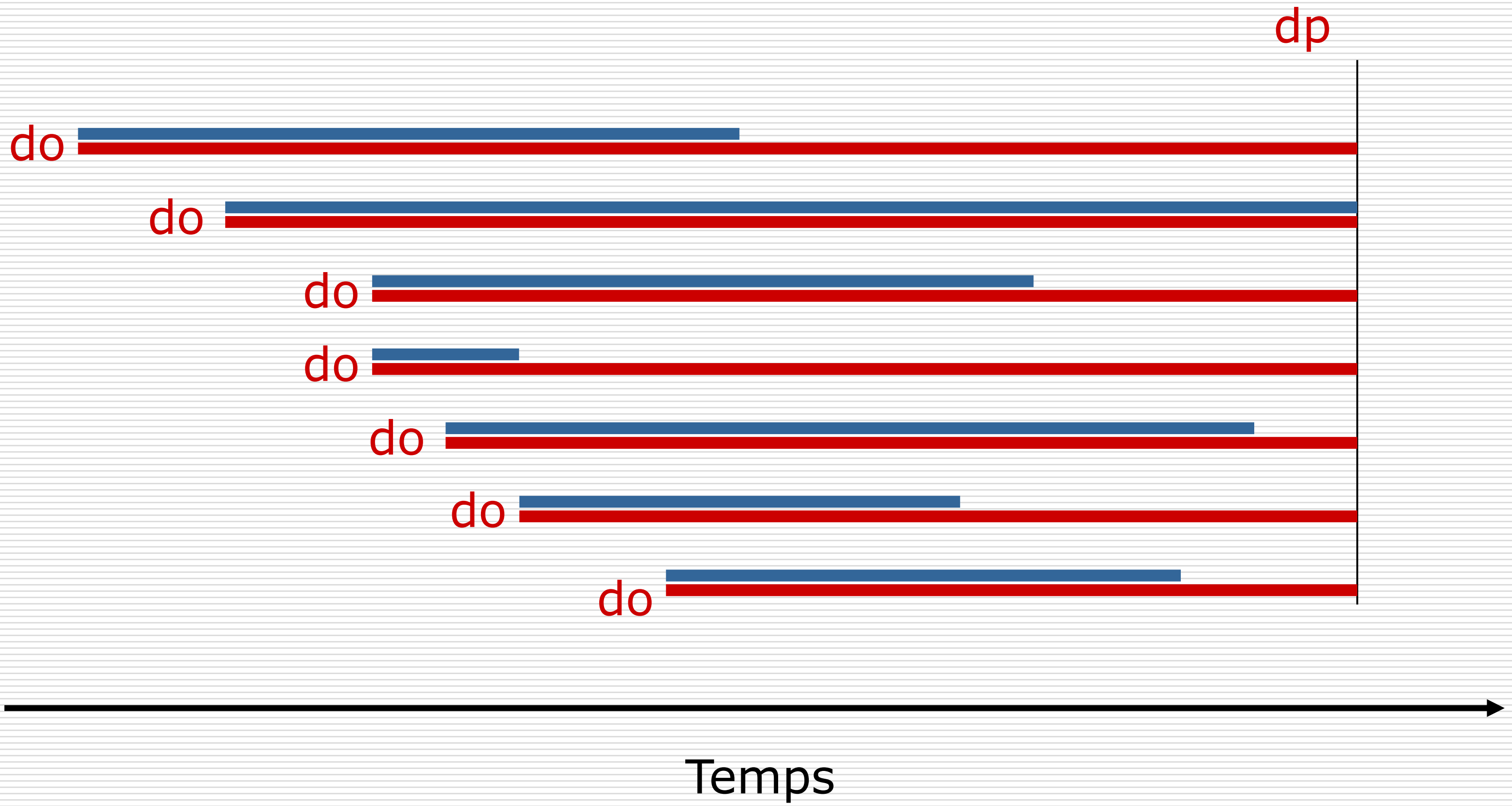
Temps de participation



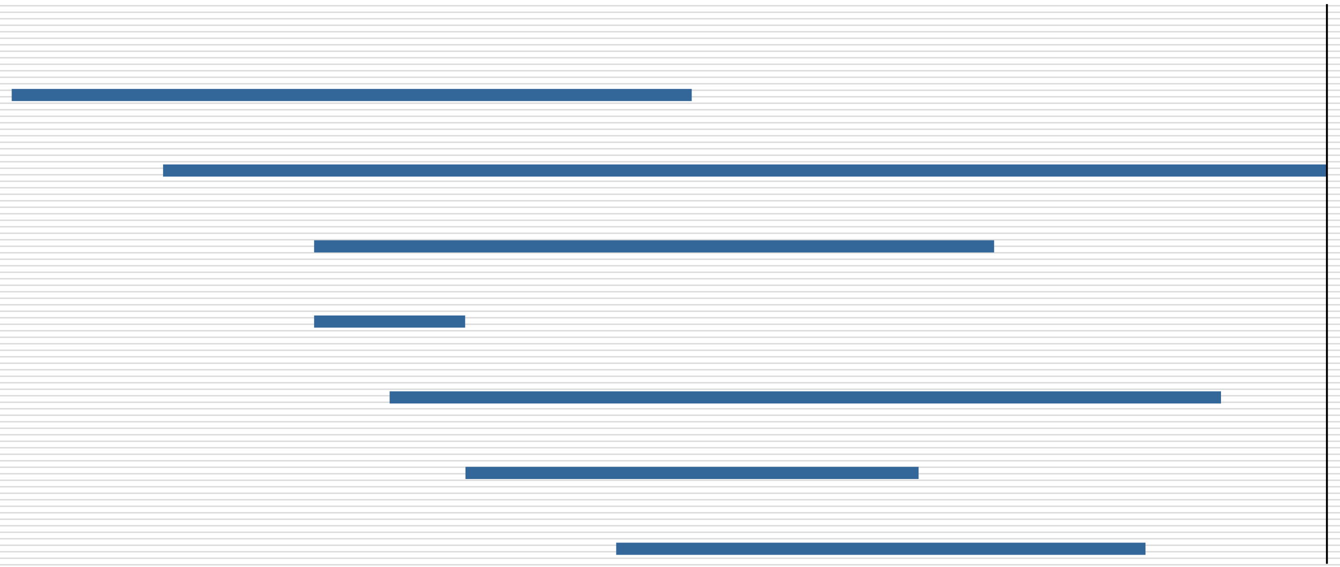
Temps de participation et censure



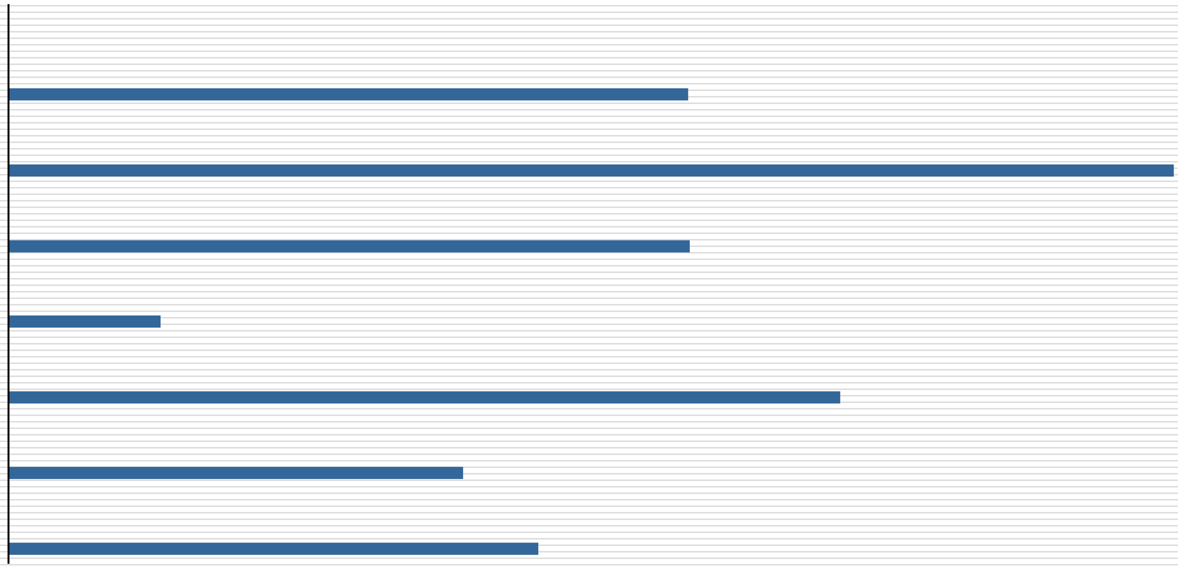
Temps de participation et délai



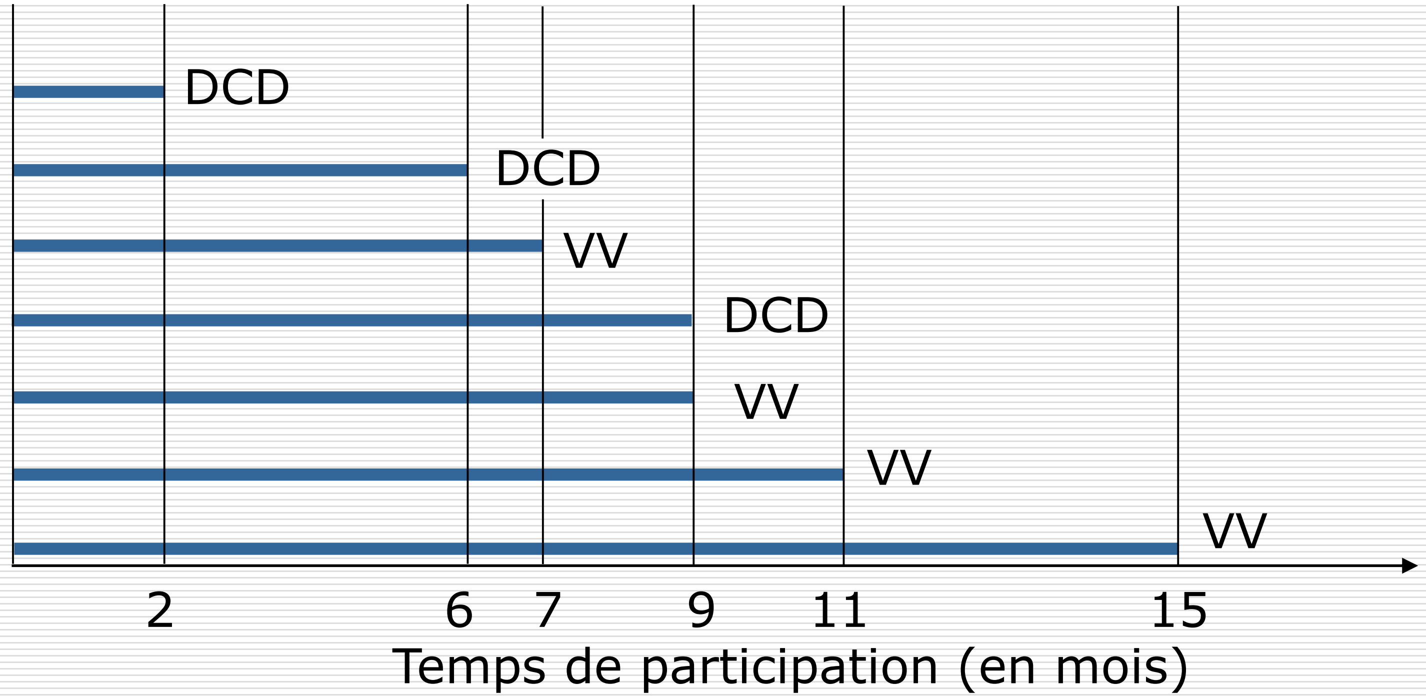
Temps de participation à partir de 0



Temps de participation classés



Récapitulatif des temps de participation



→ Courbe de survie

Taux de survie

Le taux de survie représente la probabilité d'être encore en vie à un instant t. Il existe plusieurs méthodes non paramétrique pour l'estimer :

- Méthode de **Kaplan-Meier**

Cette méthode est basée sur le principe des probabilités conditionnelles.

Méthode de Kaplan-Meier

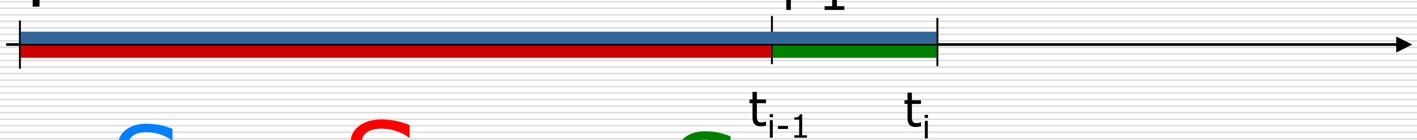
Dans cette méthode, on ne s'intéresse qu'au temps t_i où il y a un temps de participation qui se termine.

- Si le temps de participation se termine par un décès alors le taux de survie chute.
- Dans le reste des cas, le taux de survie reste constant.

Principe de la méthode de Kaplan-Meier

Ainsi, la probabilité d'être en vie à un temps t_i est égale à :

- la probabilité d'être en vie à t_{i-1} multiplié par
- la probabilité d'être en vie à t_i sachant qu'on était en vie à t_{i-1}.



$$S_i = S_{i-1} \times S_{i|i-1}$$

Pour le temps compris entre t_i et t_{i+1} exclu, la probabilité d'être en vie reste constante.

Calcul des taux de survie avec la méthode de Kaplan-Meier

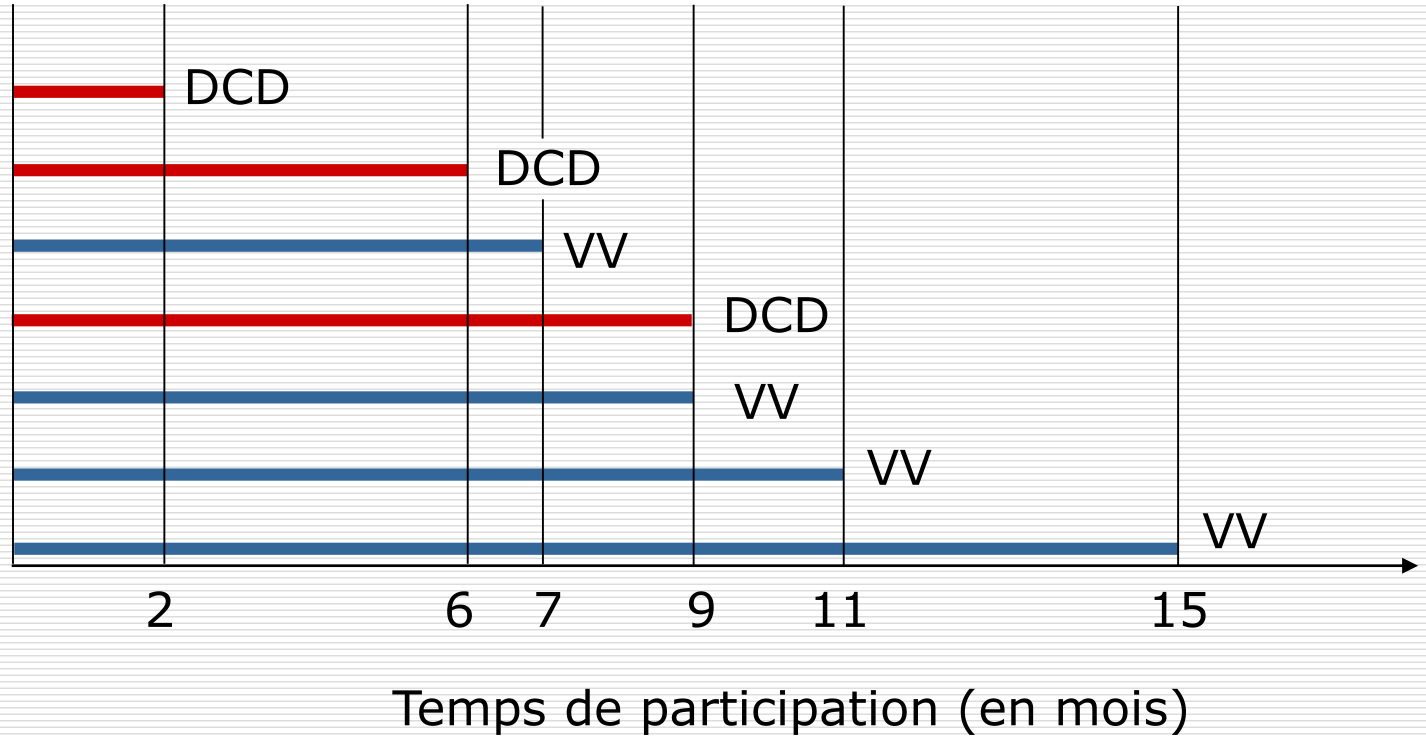
$$S_{i|i-1} = \frac{N_i - D_i}{N_i}$$

$$S_i = \frac{N_1 - D_1}{N_1} \times \frac{N_2 - D_2}{N_2} \times \ldots \times \frac{N_i - D_i}{N_i}$$

N_i le nombre de sujet exposés au temps i

D_i le nombre de morts au temps i

Récapitulatif des temps de participation



Méthode de Kaplan-Meier

t_i	$[t_i,t_{i+1}[$	N_i	D_i	$S_{i\|\ i-1}$	S_i
0	[0,2[				
2	[2,6[				
6	[6,7[				
7	[7,9[				
9	[9,11[				
11	[11,15]				

Méthode de Kaplan-Meier

t_i	$[t_i,t_{i+1}[$	N_i	D_i	$S_{i\mid i-1}$	S_i
0	[0,2[	7	0		
2	[2,6[	7	1		
6	[6,7[	6	1		
7	[7,9[	5	0		
9	[9,11[	4	1		
11	[11,15]	2	0		

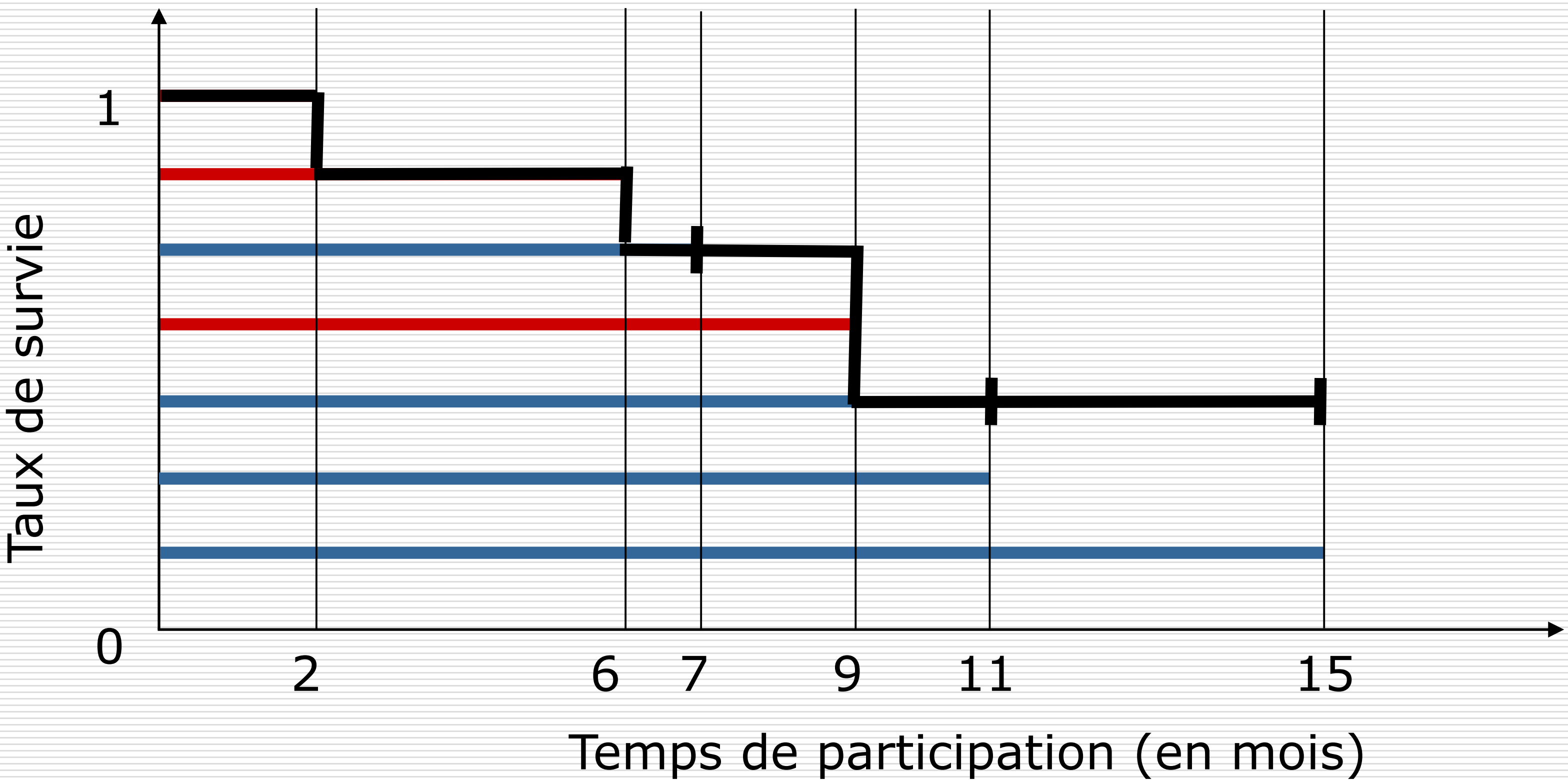
Méthode de Kaplan-Meier

t_i	$[t_i,t_{i+1}[$	N_i	D_i	$S_{i\|\ i-1}$	S_i
0	[0,2[	7	0	1	
2	[2,6[	7	1	6/7=0.857	
6	[6,7[	6	1	5/6=0.833	
7	[7,9[	5	0	1	
9	[9,11[	4	1	3/4=0.75	
11	[11,15]	2	0	1	

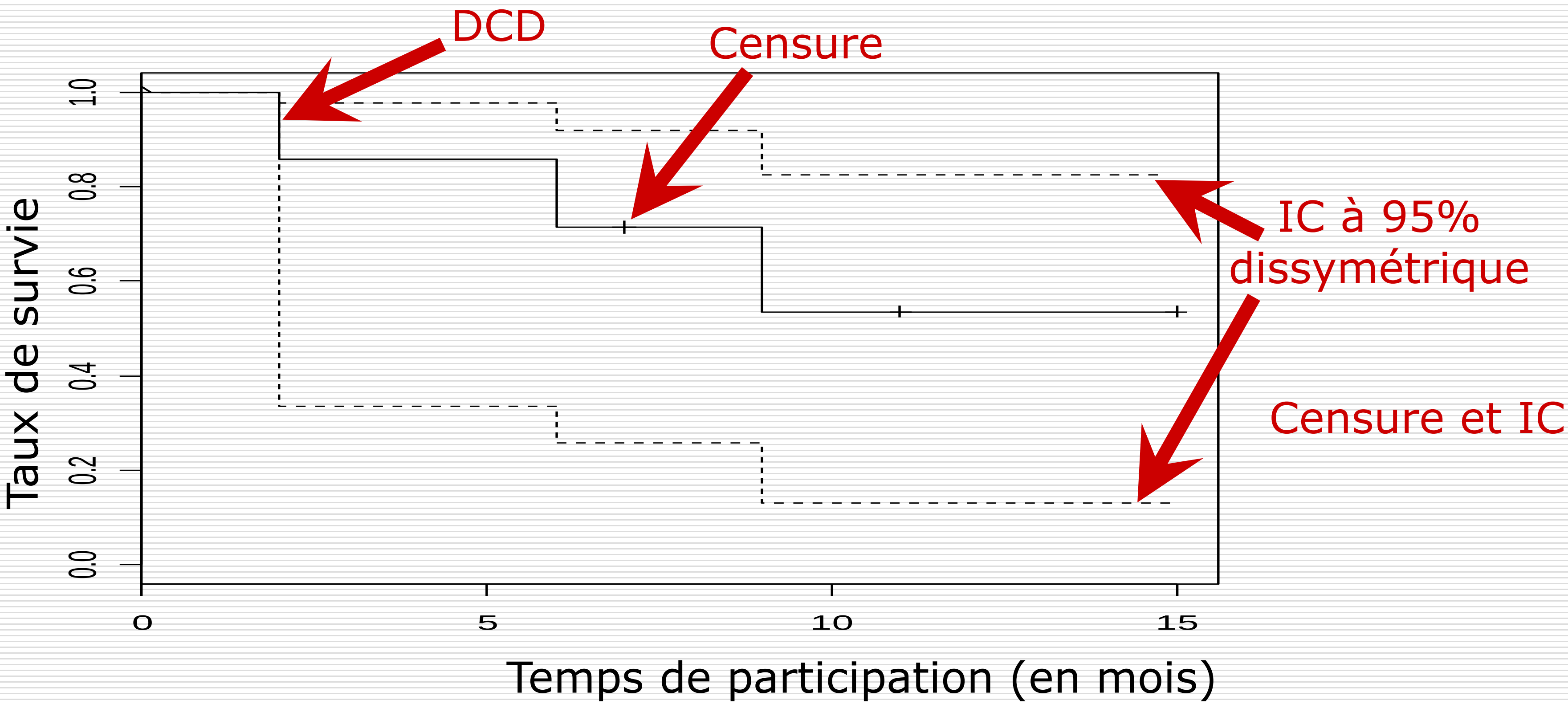
Méthode de Kaplan-Meier

t_i	$[t_i,t_{i+1}[$	N_i	D_i	$S_{i\| i-1}$	S_i
0	[0,2[	7	0	1	1
2	[2,6[	7	1	6/7=0.857	0.857
6	[6,7[	6	1	5/6=0.833	0.833x 0.857=0.714
7	[7,9[	5	0	1	0.714
9	[9,11[	4	1	3/4=0.75	0.714x 0.75=0.536
11	[11,15]	2	0	1	0.536

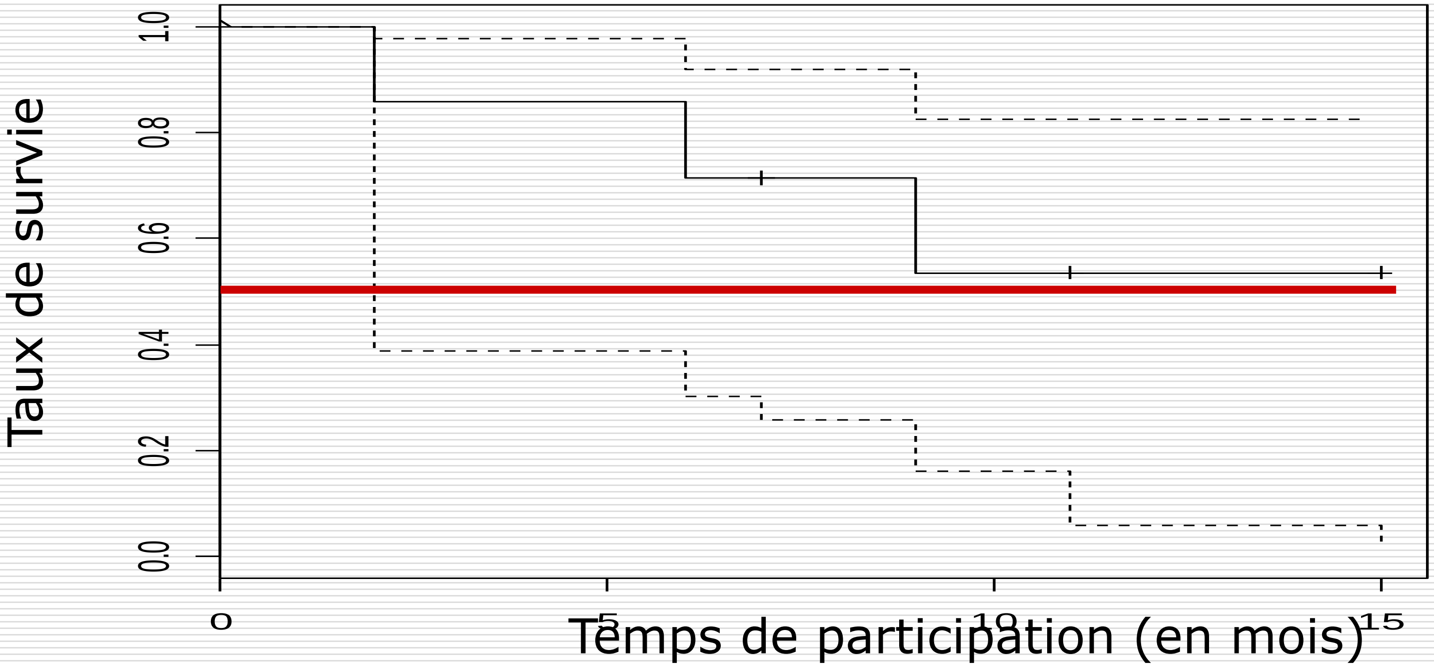
Représentation graphique d'une courbe de survie (Kaplan-Meier)



Représentation graphique d'une courbe de survie (Kaplan-Meier)



Commentaires sur les courbes et le délai-médian



Commentaires sur les courbes et le délai-médian

Les distributions de survie sont en générale dissymétriques dans le temps et le délai médian est très instable si le risque instantané de décès n'est pas élevé autour de la valeur de 50 %. Donc, une courbe représentera mieux la survie et sa variabilité que le délai médian.

Conclusion sur la méthode Kaplan-Meyer

Cette méthode permet d’obtenir une taux de survie qui est modifié à chaque décès et un intervalle de confiance qui change à chaque décès et qui peut changer à chaque censure. Il est à noter que si l’échantillon est trop petit, l’intervalle de confiance est très grand et la courbe de survie n’apporte pas d’information.

Remarque sur les données censurées

- Soit des exclus-vivants car leur histoire n’est pas suffisamment longue pour participer à l’estimation de tous les taux,
- Soit des perdus de vue
  - 1ère hypothèse même survie que les autres assimilés aux les exclus-vivants
  - 2ème hypothèse survie augmentée ou diminuée -> considérés soit vivants soit morts
  - Cas de l’oncologie pédiatrique

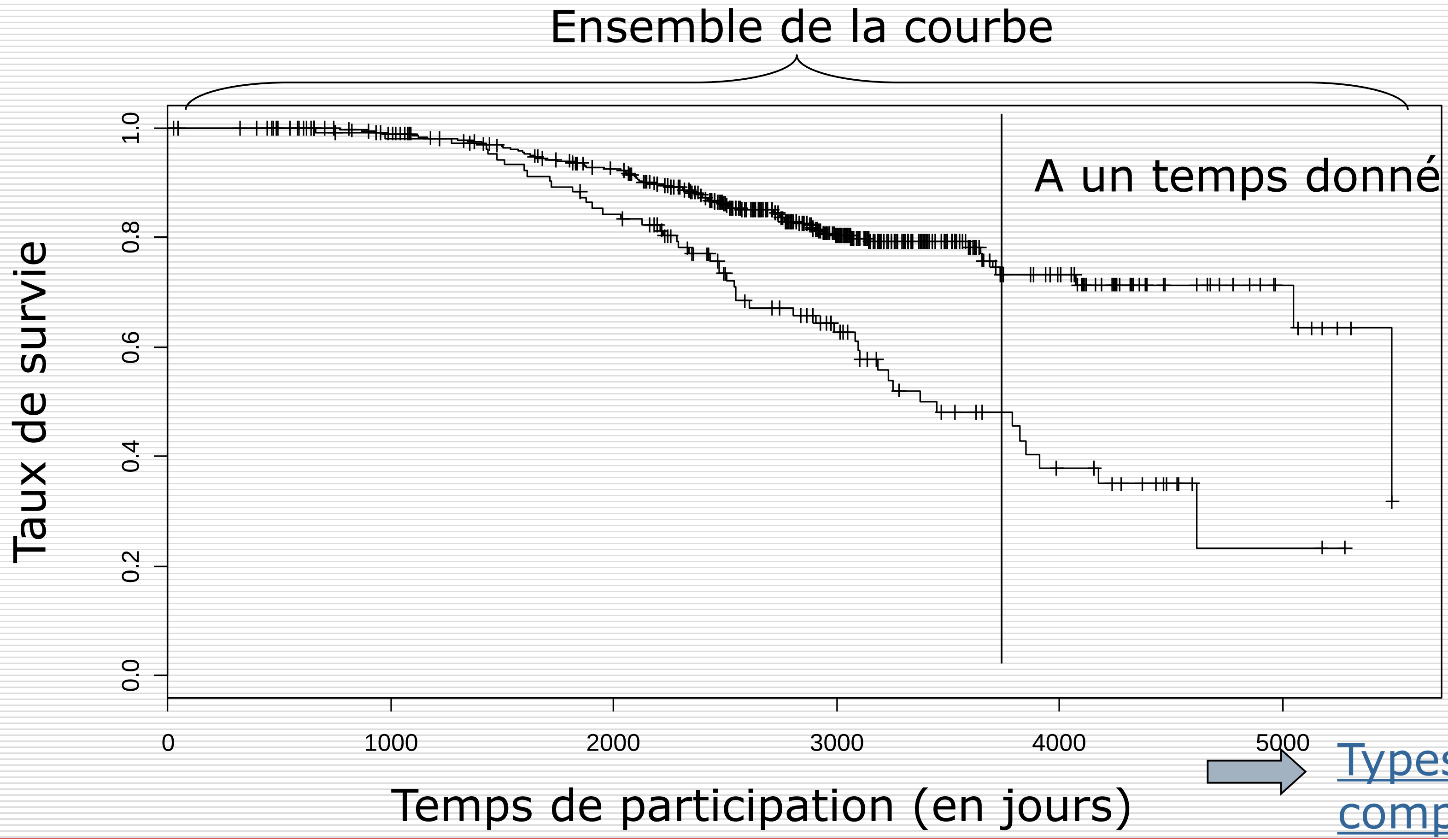
Il est temps de pratiquer

Comparaison de courbes de survie

Types de comparaison

- Comparaison de plusieurs Taux de survie
  - A temps donné
  - Ensemble de la courbe
    - Analyse univariée
    - Analyse multivariée

Comparaison de plusieurs courbe de survie



Types de comparaison

A un temps donné t_i

- Si les deux IC (pour le groupe A et le groupe B) sont a peu près symétrique (on supposera l'hypothèse des distribution normales des taux de survie de moyenne respectivement S_{Ai} et S_{Bi} et de variances $\mathrm{var}S_{Ai}$ et $\mathrm{var}\ S_{Bi}$).

$$u_{obs} = \frac{S_{Ai} - S_{Bi}}{\sqrt{\mathrm{var}\,S_{Ai} + \mathrm{var}\,S_{Bi}}}$$

Cf. présentation sur la survie avec R

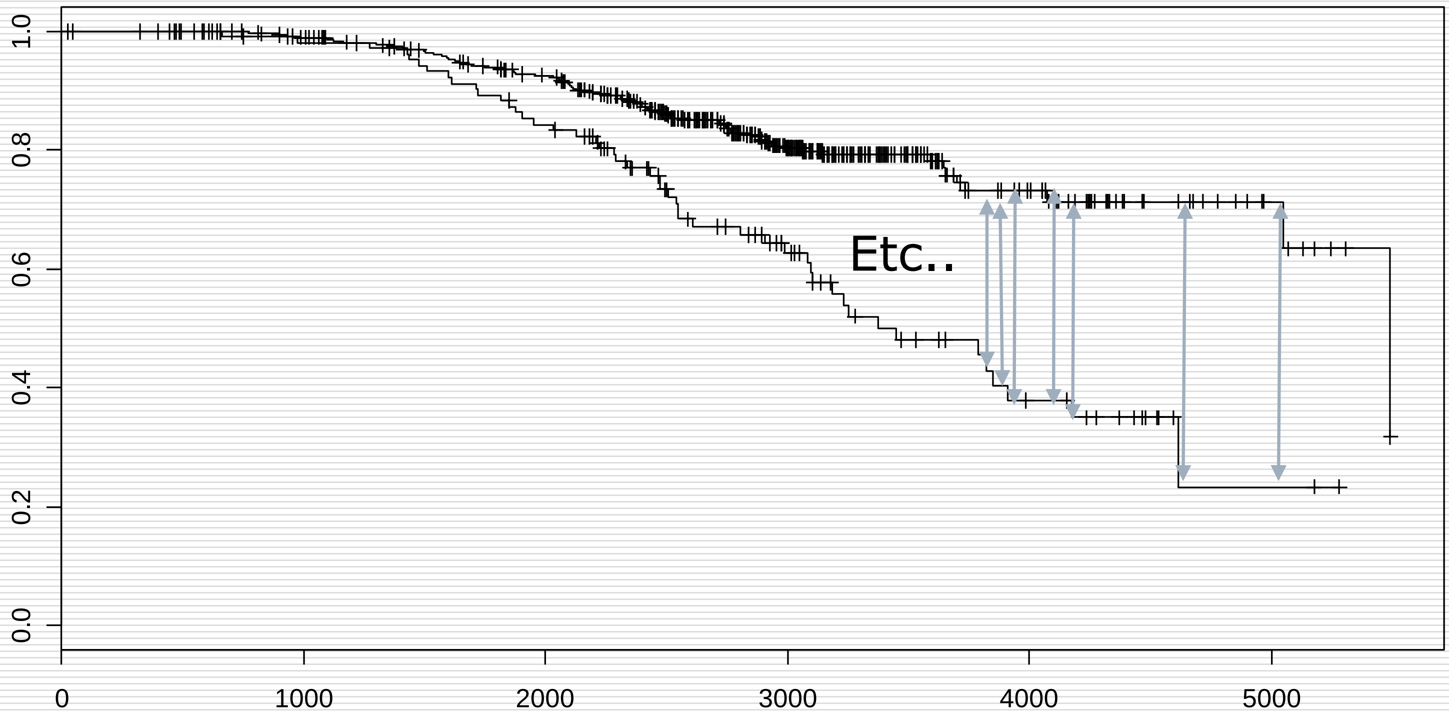
à comparer à : $u_{1-\alpha/2} = 1.96$

Types de comparaison

Analyse univariée sur l'ensemble des courbes

- Test du logrank
  - Ce test s'appelle aussi test de Mantel-Haenszel ou test de Cox.
  - C'est le test le plus souvent employé.
  - Ce test est non paramétrique, c'est-à-dire ne supposant pas une distribution particulière du temps de survie.

Analyse univariée sur l'ensemble des courbes



Principe du test du logrank

Comme de le cas de la méthode de Kaplan-Meier ne s'intéresse qu'aux k temps t_i où il y a eu au moins un décès. Dans le cas de 2 groupes :

t_i	Groupe A	Groupe B	Tot
DC	O_{Ai}	O_{Bi}	D_i
VV	$N_{Ai}-O_{Ai}$	$N_{Bi}-O_{Bi}$	N_i-D_i
Tot	N_{Ai}	N_{Bi}	N_i

$$E_{Ai} = N_{Ai}\frac{D_i}{N_i}$$

$$E_{Bi} = N_{Bi}\frac{D_i}{N_i}$$

O_{ai}=Nombre de DC au temps i pour le groupe A

N_{ai}=Nombre de sujet au temps i pour le groupe A

Principe du test du logrank (suite)

Puis, on somme pour tous les k temps t_i :

$$O_A = \sum_{i=1}^{k} O_{Ai} \text{ et } O_B = \sum_{i=1}^{k} O_{Bi}$$

$$E_A = \sum_{i=1}^{k} E_{Ai} \text{ et } E_B = \sum_{i=1}^{k} E_{Bi}$$

$$\chi^2_{MH} = \frac{(O_A - E_A)^2}{var(O_A - E_A)}$$ à comparer à : $\chi^2_{1-\alpha/2}(1) = 3.84$

$var(O_A - E_A)$ = formule compliquée

Formule simplifiée

$$\chi_S^2 = \frac{(O_A - E_A)^2}{E_A} + \frac{(O_B - E_B)^2}{E_B} < \chi_{MH}^2$$

à comparer à : $\chi^2_{1-\alpha/2}(1) = 3.84$

Cas où il y a 3 groupes :

$$\chi_S^2 = \frac{(O_A - E_A)^2}{E_A} + \frac{(O_B - E_B)^2}{E_B} + \frac{(O_C - E_C)^2}{E_C} < \chi_{MH}^2$$

à comparer à : $\chi^2_{1-\alpha/2}(2) = 5.99$

Taux de survie S(t) et risque instantané de mourir h(t)

$S(t)=P(T>t)$ = probabilité de survie au temps t

$F(t)=P(T\leq t)$ = probabilité d'être mort au temps t

$f(t)=F'(t)$= risque instantané de mourir au temps t * probabilité d'avoir vécu jusqu'à t

$= h(t)*S(t)$

$h(t)=f(t)/S(t)$

Hazard ratio HR

$$HR = \frac{\text{risque instantané de décès dans le groupe B}}{\text{risque instantané de décès dans le groupe A}}$$

$$HR_{\text{ajusté}} = \frac{O_B / E_B}{O_A / E_A}$$

Savoir si HR est significativement différent de 1 revient à effectuer le test du logrank.

Application du test du logrank

t_i	Groupe	Etat	O_{Ai}	O_{Bi}	D_i	N_{Ai}	N_{Bi}	N_i	E_{Ai}	E_{Bi}
2	B	DCD								
3	A	DCD								
4	B	DCD								
6	A	DCD								
7	A	VV								
7	B	VV								
8	B	DCD								
9	A	DCD								
9	B	DCD								
9	A	VV								
11	B	DCD								
11	A	VV								
12	B	VV								
15	A	VV								
Total										

Application du test du logrank

t_i	Groupe	Etat	O_{Ai}	O_{Bi}	D_i	N_{Ai}	N_{Bi}	N_i	E_{Ai}	E_{Bi}
2	B	DCD	0	1	1					
3	A	DCD	1	0	1					
4	B	DCD	0	1	1					
6	A	DCD	1	0	1					
7	A	VV								
7	B	VV								
8	B	DCD	0	1	1					
9	A	DCD	1	1	2					
9	B	DCD								
9	A	VV								
11	B	DCD	0	1	1					
11	A	VV								
12	B	VV								
15	A	VV								
Total			3	5						

Application du test du logrank

t_i	Groupe	Etat	O_{Ai}	O_{Bi}	D_i	N_{Ai}	N_{Bi}	N_i	E_{Ai}	E_{Bi}
2	B	DCD	0	1	1	7	7	14	7/14 =0.5	7/14 =0.5
3	A	DCD	1	0	1	7	6	13	0.538	0.462
4	B	DCD	0	1	1					
6	A	DCD	1	0	1					
7	A	VV								
7	B	VV								
8	B	DCD	0	1	1					
9	A	DCD	1	1	2					
9	B	DCD								
9	A	VV								
11	B	DCD	0	1	1					
11	A	VV								
12	B	VV								
15	A	VV								
Total			3	5						

Application du test du logrank

t_i	Groupe	Etat	O_{Ai}	O_{Bi}	D_i	N_{Ai}	N_{Bi}	N_i	E_{Ai}	E_{Bi}
2	B	DCD	0	1	1	7	7	14	7/14 =0.5	7/14 =0.5
3	A	DCD	1	0	1	7	6	13	0.538	0.462
4	B	DCD	0	1	1	6	6	12	0.5	0.5
6	A	DCD	1	0	1	6	5	11	0.545	0.455
7	A	VV								
7	B	VV								
8	B	DCD	0	1	1	4	4	8	0.5	0.5
9	A	DCD	1	1	2	4	3	7	1.143	0.857
9	B	DCD								
9	A	VV								
11	B	DCD	0	1	1	2	2	4	0.5	0.5
11	A	VV								
12	B	VV								
15	A	VV								
Total			3	5					4.226	3.774

Application du test du logrank

t_i	Groupe	Etat	O_{Ai}	O_{Bi}	D_i	N_{Ai}	N_{Bi}	N_i	E_{Ai}	E_{Bi}	$V(O_{Ai}-E_{Ai})$
2	B	DCD	0	1	1	7	7	14	7/14 =0.5	7/14 =0.5	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Total			3	5					4.226	3.774	1.905

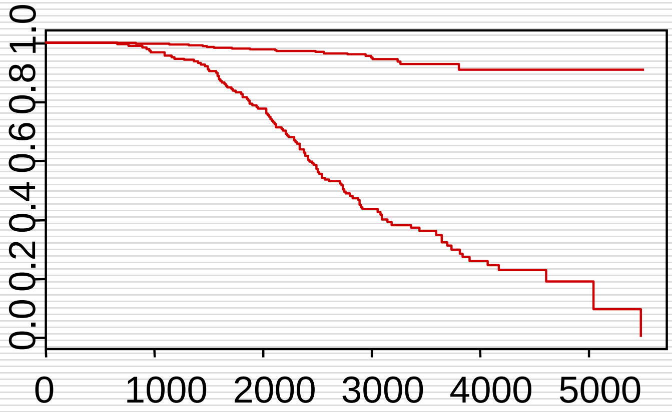
$$\chi^2_{MH} = \frac{(3-4.226)^2}{1.905} = 0.789$$ **NS**

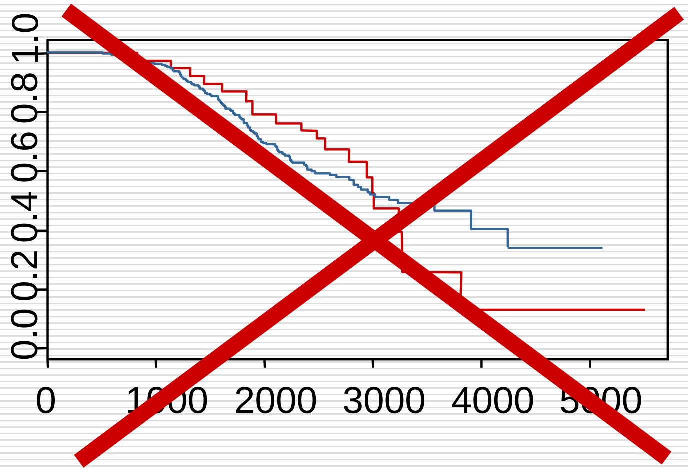
$$\chi^2_S = \frac{(3-4.226)^2}{4.226} + \frac{(5-3.774)^2}{3.774} = 0.754$$ **NS**

$$HR = \frac{\frac{5}{3.774}}{\frac{3}{4.227}} = 1.9$$

Commentaire sur le test du logrank

Le test n'est pas adapté si les courbes se croisent.





=> Risques proportionnels

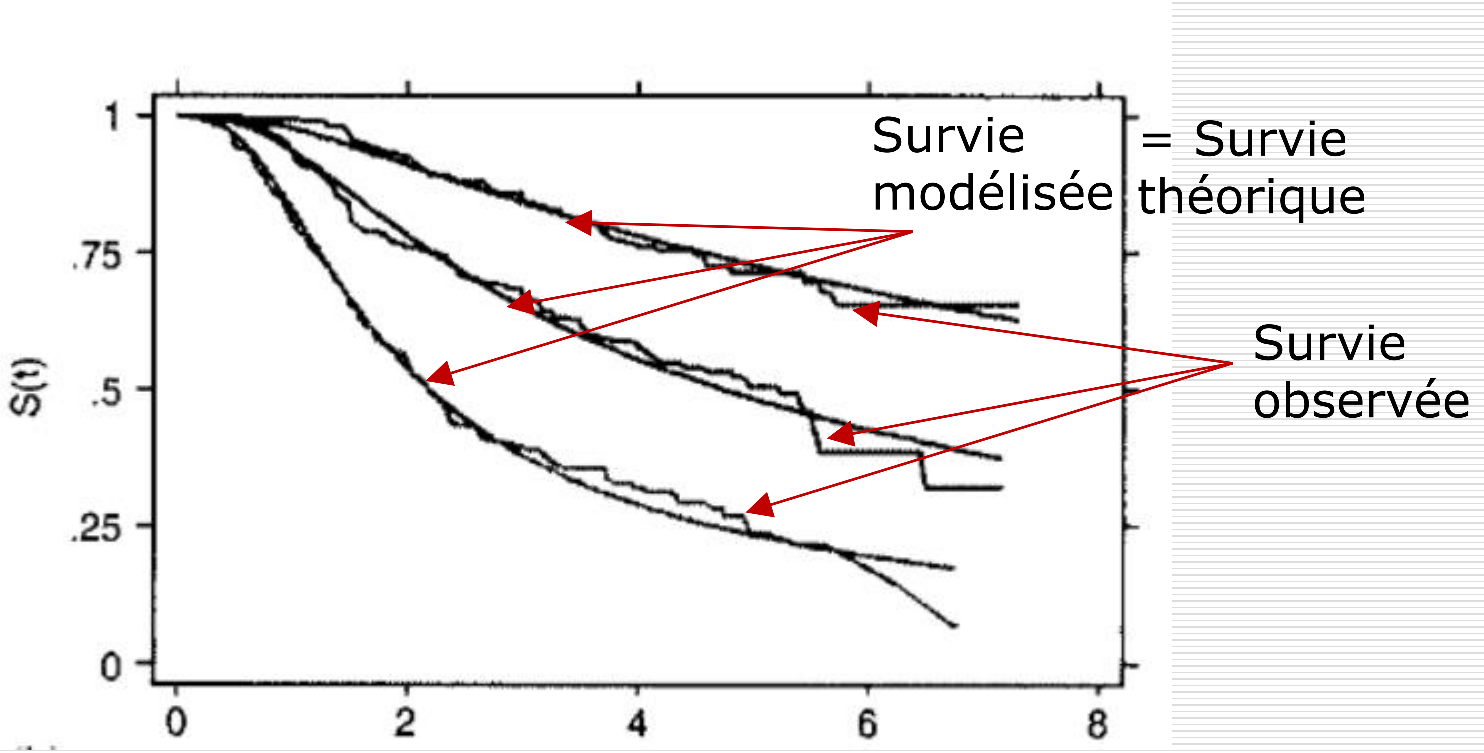
[Types de comparaison]

Il est temps de pratiquer

Analyse multivariée sur l'ensemble des courbes

- Test du logrank ajusté sur un facteur de confusion de manière de similaire à ce qui est fait avec les OR ajusté (Cf Epidémio S7)
- Modèle de survie (prise en compte de plusieurs facteurs de confusion simultanément)

Qu'est ce qu'un modèle de survie ?



Qu’est ce qu’un modèle de survie ?

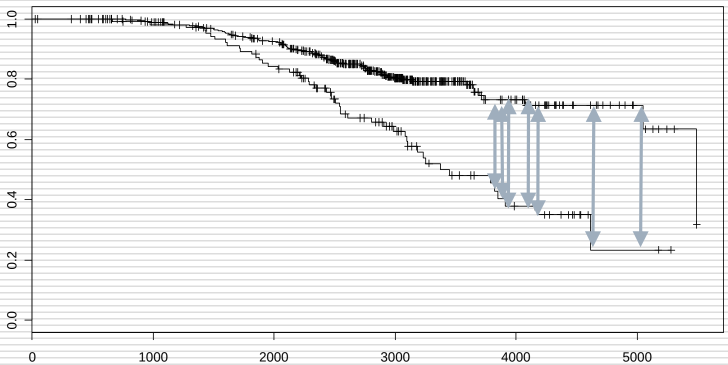
Décrire a minima l’augmentation ou la diminution du risque de mourir en fonction des variables (sexe, age, stade de maladie…) :

Modèle semi paramétrique (Modèle de cox)

Décrire la survie en fonction des variables (sexe, age, stade de maladie…) :

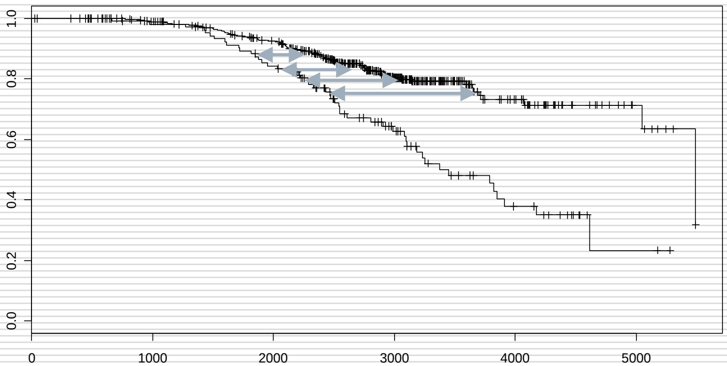
Modèle paramétrique (Modèle exponentiel / Modèle de Weibull / Modèle lognormal…)

Analyse multivariée sur l'ensemble des courbes



Risques proportionnels (PH)

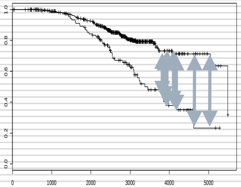
- Test du logrank ajusté
- Modèle non paramètrique de Cox
- Modèle paramétrique :
  - Exponentielle
  - Weibull



Accelerated Failure Time (AFT)

- Modèle paramétrique :
  - Exponentielle
  - Weibull
  - Loglogistique..

Risques proportionnels (PH) :



c'est-à-dire que le rapport des risques pour deux individus est constant et indépendant du temps.

Ex. Modèle de Cox : le risque mourir h au temps t avec $X_1,\ldots,X_n$

$$h(t, X_1, \ldots, X_n) = h_0(t) e^{\sum_{i=1}^{n} \beta_i X_i}$$

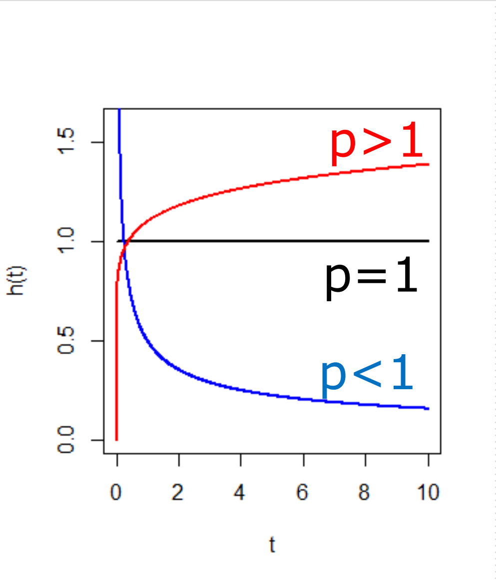
la partie paramétrique est $e^{\sum_{i=1}^{n} \beta_i X_i}$ $\beta_i = \log(\mathrm{HR}_i)$

et la partie non paramétrique $h_0(t)$

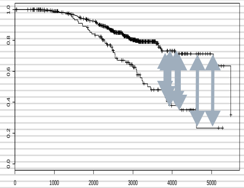
=>Modèle de Cox est modèle semi-paramétrique

Ex. Modèle de Weibull (modèle paramétrique) : le risque mourir h au temps t avec $X_1,\ldots,X_n$

$$\mathrm{h}(\mathrm{t}, \mathrm{X}_1, \ldots, \mathrm{X}_\mathrm{n}) = \lambda p t^{\mathrm{p}-1} e^{(\beta_0 + \sum_{i=1}^{n} \beta_i X_i)}$$



Risques proportionnels (PH) :



c'est-à-dire que le rapport des risques pour deux individus est constant et indépendant du temps.

Ex. Modèle de Weibull (modèle paramétrique) : le risque mourir h au temps t avec $X_1,\ldots,X_n$

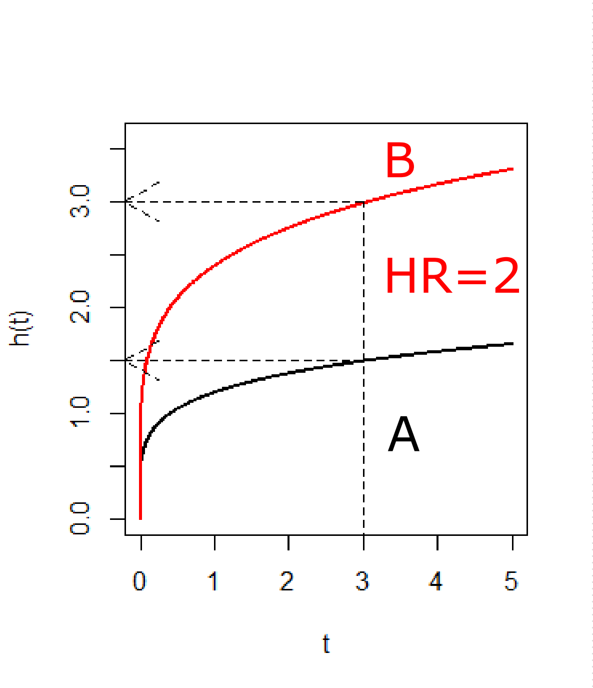
$$h(t, X_1, \ldots, X_n) = \lambda p t^{p-1} e^{(\beta_0 + \sum_{i=1}^{n} \beta_i X_i)}$$

Hazard Ratio (HR)

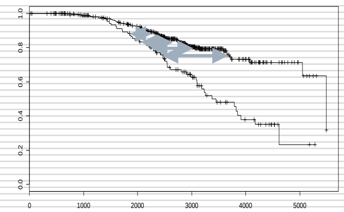
$$h_B(t) = h_A(t).\ \mathrm{HR}$$

$$\beta_i = \ln(\mathrm{HR}_i)$$

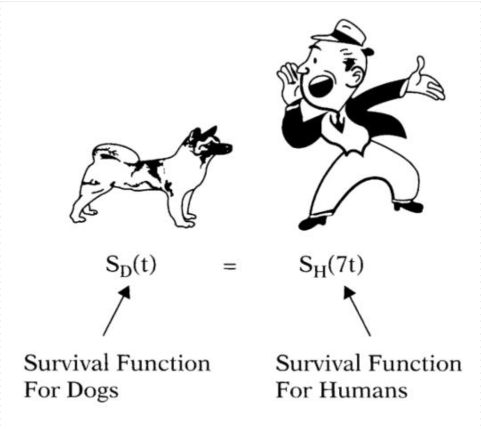
$\beta_i > 0$ ☹



l'effet des variables explicatives est multiplicatif par rapport au risque de mourir



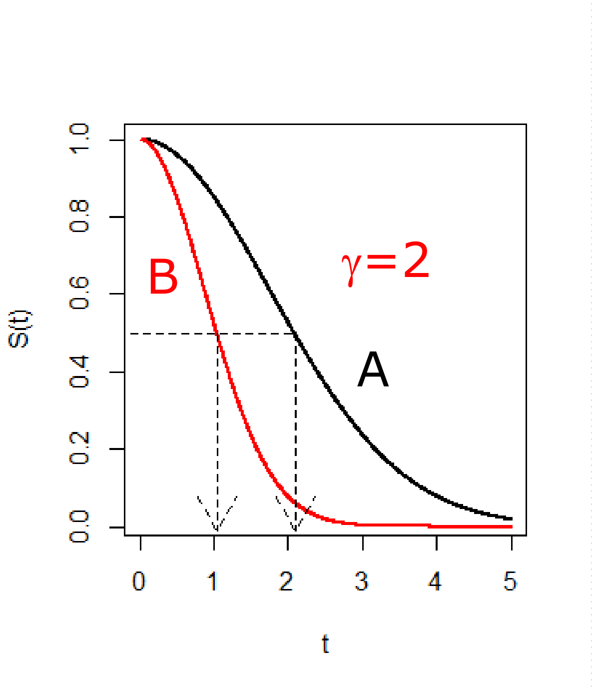
Accelerated Failure Time (AFT)

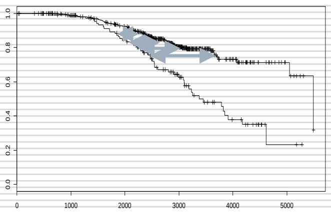


l'effet des variables explicatives est multiplicatif par rapport au temps de survie.

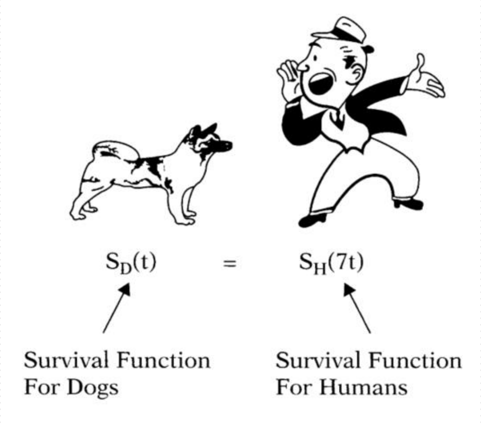
$$S_B(t\ ) = S_A(t.\gamma)$$

γ = facteur d'accélération





Accelerated Failure Time (AFT)



Ex. Modèle de Weibull (modèle paramétrique) : au temps t où la probabilité de survie est q avec $X_1,\ldots,X_n$

$$t(X_1, \ldots, X_n) = [-\ln(q)]^{1/p} e^{(\alpha_0 + \sum_{i=1}^{n} \alpha_i X_i)}$$

$$\alpha_i = \ln(\gamma_i) = -\beta_i p$$

α_i=log-ratio du temps de survie

$\alpha_i > 0$ ☺

Modèle de Cox

L'intérêt du modèle de Cox est de pouvoir prendre en compte plusieurs facteurs simultanément toute en ne faisant d'hypothèse sur la distribution de survie. Il suppose que le risque de base dépendent du temps et que le risque liés au facteur étudié ne dépendent pas du temps (hypothèse des risques proportionnels).

Il est actuellement l'un des modèles les plus utilisés pour comparer les courbes de survie.

Lorsque l'on prend qu'**un facteur** dans le modèle de Cox cela revient à faire le **test du logrank** et **2 facteurs** le **test du logrank ajusté**.

Modèle de survie

$\lambda(t)$ la probabilité de survenue d'un événement connaissant les valeurs des variables X_i,

c'est aussi ce que l'on appelle l'incidence instantanée.

On cherche à estimer $\lambda(t)$.

Modèle de cox

$\lambda(t)$ la probabilité de survenue d'un événement connaissant les valeurs des variables X_i,

c'est aussi ce que l'on appelle l'incidence instantanée.

On cherche à estimer $\lambda(t)$.

$\lambda(t)$ est relié à la probabilité S(t) d'être indemne de l'événement à l'instant t.

$$S(t) = e^{-\int_0^t \lambda(u)du}$$

Modèle de Cox

Dans le modèle de Cox, $\lambda(t)$ s'exprime ainsi :

$$\lambda(t, X_1, \ldots, X_p) = \lambda_0(t) e^{\sum_{i=p}^{p} \beta_i X_i}$$

Le modèle de Cox est souvent qualifié de modèle semi paramétrique : la partie paramétrée est $e^{\sum_{i=p}^{p} \beta_i X_i}$

et la partie « non paramétré » est $\lambda_0(t)$

Modèle de Cox

- Dans le cas le plus fréquent où Xi est dichotomique (valeur 0 et valeur 1).

$$e^{\beta_i} = HR$$

de la variable i ajusté sur les autres variables.

Ceci permet de mesurer l'effet de la variable i en tenant compte des autres variables afin d'éviter les facteurs de confusion.

Choisir le modèle le plus approprié

- Le principe de parcimonie (ne conserver que les paramètres significatifs)
- Vérifier l'hypothèse des risques proportionnels

Il est temps de pratiquer