

# Analyse de la survie

## Introduction :

L'analyse de la survie est l'estimation de la probabilité de survenue d'un événement (décès, complication post opératoire, rechute, guérison) dans le temps, en fonction de facteurs pronostiques (éléments influençant l'estimation).

Par convention, on nomme l'événement attendu : « décès ».

- Probabilité de survivre au moins un certain temps « t » à compter d'un instant de référence.
- Probabilité pour que l'évènement attendu survienne après un certain délai.

*Exemple : Probabilité pour que le décès d'un patient survienne après un certain délai (1 an par ex) sachant que le cancer dont il souffre est au stade 4.*

Une étude de survie est une étude :

- Longitudinale (suivi des personnes au cours du temps)
- Prospective (prise en compte des événements survenant dans la durée de l'étude)
- De cohorte (observation d'un groupe de personnes dans le temps)

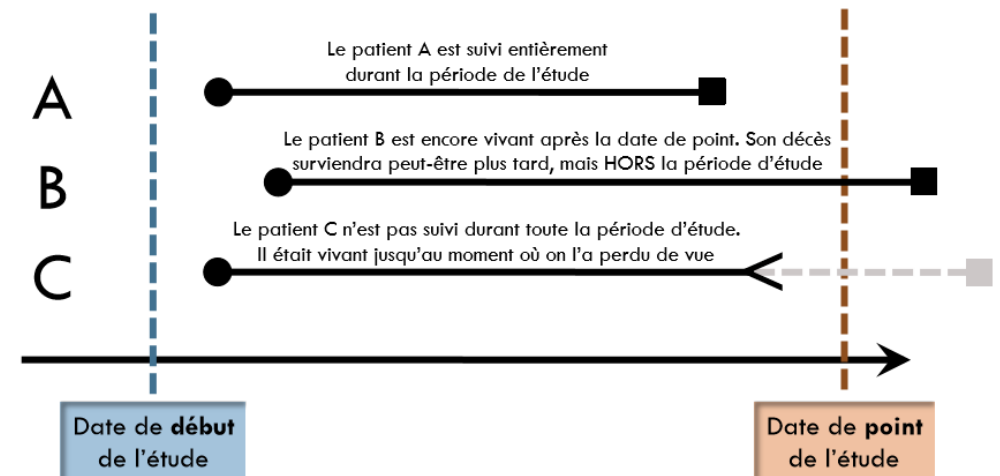
## I/ Définitions importantes :

- **Cohorte** : Ensemble de sujets inclus dans une étude au même moment, et suivis dans des conditions standardisées pendant une

durée prédéfinie.

- **Cohorte « incipiente »** : Dans ce cas, la cohorte des patients qui rentrent dans l'étude doit inclure des sujets observés au début de leur affection au même stade de leur maladie (« cas incidents »).
- **Événement d'intérêt** : événement auquel on s'intéresse au cours de l'étude : décès, décès lié à un AVC, complication, rechute, disparition de symptômes etc... On utilisera l'« analyse de la survie » dès qu'il y aura une notion de durée jusqu'à la survenue de l'évènement d'intérêt.
- **Date d'origine** : C'est la date correspondant au point de départ de la surveillance. Elle peut être différente pour chaque sujet selon les modalités d'inclusion du sujet. Dans certains cas la date d'origine peut être antérieure à l'inclusion dans l'étude : on parle alors de cohorte historique.
- **Date de point** : C'est la date choisie pour faire le bilan.
- **Date des dernières nouvelles** : C'est la date la plus récente à laquelle on a recueilli des informations sur le patient, notamment la survenue ou non de l'évènement d'intérêt.

- **Perdu de vue :** Un sujet est perdu de vue lorsque sa surveillance est interrompue avant la date de point et que l'événement d'intérêt ne s'est pas produit.
- **Censure :** Une durée de survie d'un individu est dite censurée lorsque l'événement d'intérêt n'a pas été observé. Elle concerne : les sujets perdus de vue (C) et ceux vivant à la date de point (B)
- **Temps de recul :** Délai entre la date d'origine et la date de point, c'est-à-dire le délai maximum potentiel de suivi pour un sujet.
- **Temps de participation :** Durée de surveillance utilisée pour chaque sujet dans l'estimation de la survie.



## II/ Fonction de survie :

Trois cas :

- 1er cas : L'événement a lieu au cours de la surveillance : Temps de participation = Date de survenue de l'événement - Date d'origine.
- 2e cas : Le sujet est vivant à la date de point : Temps de participation = Date de point - Date d'origine.
- 3e cas : Le sujet est perdu de vue : Temps de participation = Date des dernières nouvelles - Date d'origine

### 1. Loi exponentielle :

*Rappel : La loi exponentielle est utilisée couramment pour représenter la durée de vie de composants ou d'équipements pour lesquels on suppose que le taux de défaillance est constant au cours du temps (la durée de vie au-delà de « t » est indépendante de « t »).*

- Fonction de répartition de la loi exponentielle :  $F(t)$  représente la proportion d'équipement qui tombent en panne avant le temps « t ».

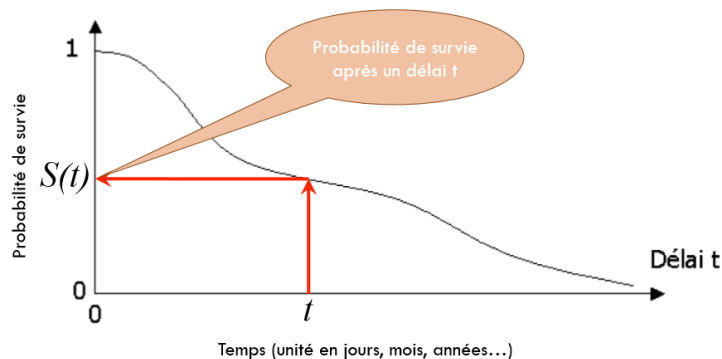
$$F(t) = P(X \leq t) = \int_0^t \lambda e^{-\lambda x} dx = 1 - e^{-\lambda t}$$

## 2. La fonction de survie :

La quantité  $1-F(t)$  représente la quantité d'équipement qui fonctionne pendant une durée au moins égale à « t ». Il s'agit de la fonction de survie  $S(t)$ .

C'est la probabilité pour que l'événement d'intérêt « T » (le décès par exemple) intervienne après un délai supérieur à « t ». Autrement dit, que l'événement d'intérêt « T » ne survienne pas avant la date « t ».

$$S(t) = 1 - F(t) = P(X > t) = e^{-\lambda t}$$



- Probabilités :

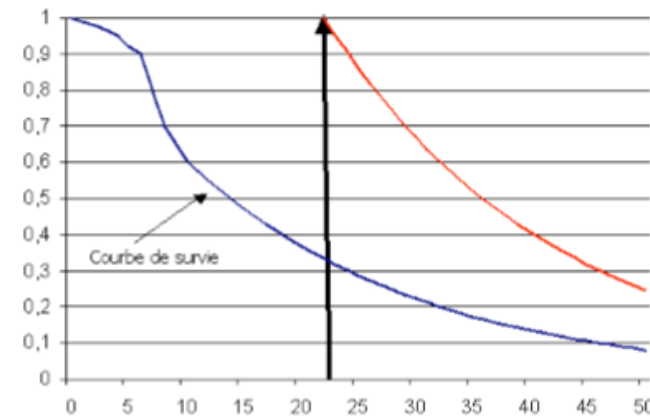
- Probabilité pour que le décès survienne après un délai  $t_1$  et avant un délai  $t_2$  ( $t_2 > t_1$ ) :

$$Pr(T \in ]t_1; t_2]) = F(t_2) - F(t_1) = S(t_1) - S(t_2)$$

- Probabilité de survivre encore après un délai « t » sachant que l'on est survivant après un délai «  $\tau$  » ( $\tau < t$ ), que l'on note  $S(t/\tau)$ .

$$S(t/\tau) = \frac{S(t)}{S(\tau)}$$

*Exemple du cours : Probabilité de vivre après 33 ans sachant que l'on est encore vivant à 23 ans.*



*D'après la courbe on voit que  $S(23\text{ans})=0.33$  et que  $S(33\text{ans})=0.2$ . Ainsi  $S(33\text{ans})/S(23\text{ans})=0.2/0.33=0.6$ .*

## III/ Estimation de la survie :

*Pour cela, connaître toutes les définitions précédentes (date d'origine, de point...) pour le recueil des données !*

Les 2 méthodes d'analyse **non paramétriques** utilisées sont :

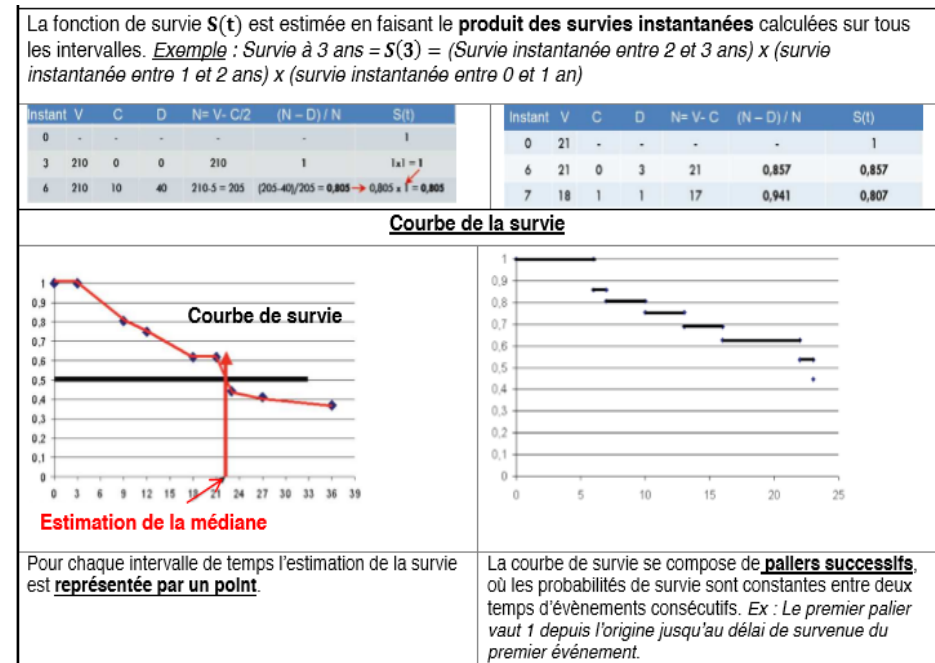
-**La méthode Actuarielle** : Utilisée lorsque les échantillons sont grands > 200 sujets.

-**La méthode de Kaplan-Meier** : Utilisée lorsque les échantillons sont < 200 sujets.

Ces deux méthodes supposent que les probabilités de survie sont indépendantes du calendrier.

*Exemple : la survie à 1 an d'un groupe de patients inclus en 1980 est identique à celle d'un groupe de patients inclus en 2000.*

<b>Méthode Actuarielle (n &gt; 200)</b>	<b>Méthode de Kaplan-Meier (n &lt; 200)</b>
La fonction de survie est calculée sur des intervalles de temps fixés à priori (mois, trimestre, année)	Les intervalles sont définis par les instants auxquels les événements sont observés → Ces intervalles sont donc inégaux, débutent à l'instant d'un événement et s'arrêtent juste avant l'événement suivant.
Pour chaque intervalle de temps on définit : ⇒ <b>V</b> : Nombre de <b>sujets vivants</b> au début de l'intervalle : ⇒ <b>D</b> : Nombre de <b>sujets décédés</b> dans l'intervalle : ⇒ <b>C</b> : Nombre de <b>sujets vivants aux dernières nouvelles</b> , dont le temps de participation s'arrête dans l'intervalle = <b>censure</b>	
<b>N</b> : Nombre de sujet exposés au risque d'événement sur l'intervalle	
$N = V - \frac{C}{2}$	$N = V - C$
Probabilité d'événement durant l'intervalle : $\frac{D}{N}$	
Probabilité de survie = survie instantanée : $\frac{N - D}{N}$	



(suite du tableau, méthode actuarielle à gauche et méthode de Kaplan-Meier à droite)

- Choix d'une valeur résumée :

**La médiane de survie** représente la durée  $t$  pour laquelle la probabilité de survie  $S(t)$  est de 50 %. En pratique, la médiane est estimée par la plus petite durée pour laquelle la survie est inférieure à 50 %. (La moyenne de survie n'est pas un bon indicateur)

*Voilà c'est fini ! ☺*

*J'espère que cette fiche vous a plu ! Je vous envoie beaucoup de courage avec ma team biostat' pour cette année difficile !*

**Minipousse**