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)

**on considère que l'événement d'intérêt est le décès*

On s'intéresse donc à :

- ⇨ La probabilité de survivre **au moins un certain temps t** à compter d'un instant de référence.
- ⇨ La probabilité pour que **l'évènement attendu survienne après un certain délai**.

  Exemple : Probabilité pour que le décès d'un patient survienne après un 1 an sachant que le cancer dont il souffre est au stade 4.

L'analyse de la survie c'est aussi **l'étude comparative de la survenue dans le temps d'un événement** dans différents groupes (*Test du log-rank*).

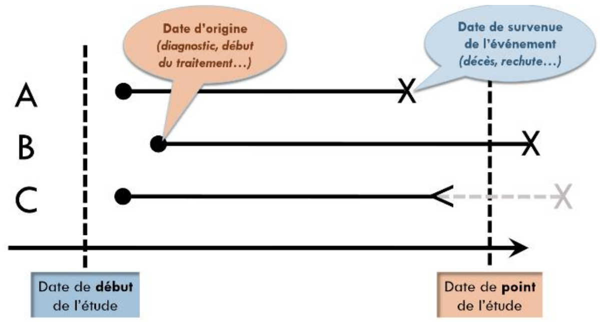
Une étude de survie est une étude :

- **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. DEFINITIONS

- **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 »)
- **Evénement d'intérêt** : événement auquel on s'intéresse au cours de l'étude → Décès, décès lié à un AVC, complication, rechute, disparition de symptômes. On utilisera l'« analyse de survie » dès qu'il y aura une notion de durée jusqu'à la survenue de l'évènement d'intérêt (qu'on nommera « décès »).
- **Durée de survie :** Délai entre la date d'origine et la date de survenue ou la date des dernières nouvelles.
- **Date d'origine :** elle correspond 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 → **cohorte historique** (Z)

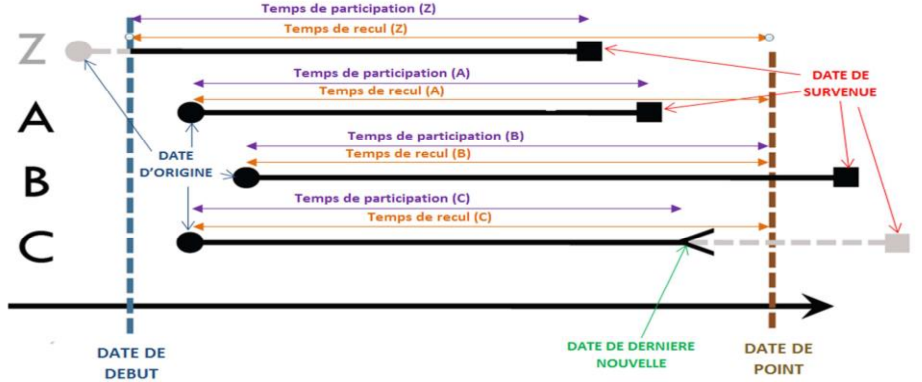


- **Date de point :** C'est la date choisie pour faire le bilan. Au-delà de cette date, les informations recueillies ne sont plus considérées dans l'analyse.
- **Date des dernières nouvelles :** C'est la date la plus récente à laquelle on a **recueilli des informations** sur le patient, notamment la survenue ou non de l'événement d'intérêt.
- **Perdu de vue** (C) **:** 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 vus** 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**. Les reculs minimum et maximum d'une série de sujets définissent donc l'ancienneté de cette série.

- **Temps de participation** : Durée de surveillance pour chaque sujet, utilisée dans l'estimation de la survie.

Trois cas :



A. L'événement a lieu au cours de la surveillance ➔ **Temps de participation = Date de survenue de l'évènement - Date d'origine.**

B. Le sujet est vivant à la date de point ➔ **Temps de participation = Date de point - Date d'origine**

C. Le sujet est perdu de vue ➔ **Temps de participation = Date de dernière nouvelle - Date d'origine**

II. FONCTION DE SURVIE

1. Loi exponentielle

La loi exponentielle est utilisée couramment pour représenter la durée de vie de composants ou d'équipements pour lesquels on suppose que **le <u>taux de défaillance</u> λ est constant au cours du temps** (la durée de vie au-delà de « t » est indépendante de « t »). Les défaillances sont donc uniquement dues au **hasard**.

La probabilité de vivre encore après 80 ans est indépendante du fait que j'ai vécu jusque mes 80 ans.

➢ <u>Fonction de **densité** de la loi exponentielle :</u>

Pour tout x≥0 ;

$$F(x) = \lambda e^{-\lambda x}$$

➢ <u>Fonction de **répartition** de la loi exponentielle :</u>

F(t) représente la proportion d'équipement qui tombent en panne avant le temps « t ».

$$F(t) = P(X \leq t) = \int \left(\lambda e^{-\lambda x}\right) dx$$

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

2. 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)***

La fonction de survie 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 ». Elle est donc une fonction de **répartition**.

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

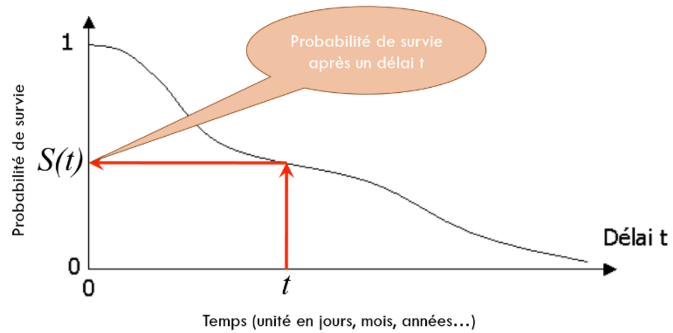
On s'intéresse donc à :

⇨ la probabilité pour qu'un **patient soit encore vivant après un délai t**

⇨ la **proportion « vraie » des survivants** après un délai t.

➢ <u>**Courbe de survie**</u>

La fonction de survie est représentée graphiquement par une courbe de survie. Elle est décroissante et S(t) ϵ [0 ;1]



- **Probabilités**

  - Probabilité que le décès survienne après un délai t1 et avant un délai t2 (t2 > t1) :

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

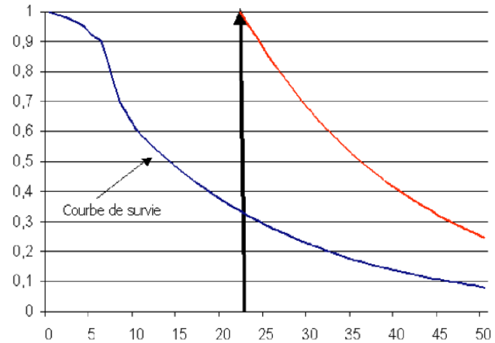
  - Probabilité de survivre encore après un délai « t » sachant que l'on est survivant après un délai « τ » (T < t) que l'on notera S(t/ τ)

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

__Exemple :__ Probabilité de survivre après 33 ans sachant que l'on est vivant à t = 23 ans.

On lit sur __la courbe bleue__ qui décrit la probabilité de survivre après «t»:

- *Proportion de survivant à 23 ans = **33%** → Il s'agit de la probabilité de survivre après 23 ans*
  S(t) = P(T>23)
- *Proportion de survivant à 33 ans = **20%** → Il s'agit de la probabilité de survivre après 33 ans*
  S(t) = P(T>33)



*On lit sur __la courbe rouge__ qui décrit la probabilité de survivre après « t » __**sachant**__ que l'on est vivant à 23 ans :*

- *Proportion de survivant à 23 ans = **100%** → Il s'agit de la probabilité de survivre après 23 ans sachant que l'on est vivant à 23 ans*

$$S(23 / 23) = \frac{S(23)}{S(23)} = \frac{0{,}33}{0{,}33} = 1$$

- *Proportion de survivant à 33 ans = **60%** → Il s'agit de la probabilité de survivre après 33 ans sachant que l'on est vivant à 23 ans*

$$S(33 / 23) = \frac{S(33)}{S(23)} = \frac{0{,}2}{0{,}33} = 0{,}6$$

III. ESTIMATION DE LA SURVIE

1. Calcul des durées de suivi

À partir du recueil de données, les durées de suivi (ou temps de participation) de chaque patient sont calculées par différence. Elles correspondent au délai entre la **date d'origine** et la **date des dernières nouvelles** qui sera : la date de décès, de point ou de perte de vue.

2. Les 2 méthodes d'analyse de survie

- La **méthode actuarielle** : utilisée lorsque les échantillons sont grands **> 200** (moins utilisée)
- La **méthode de Kaplan-Meïer** : utilisée lorsque les échantillons sont **< 200**

Ce sont deux méthodes **non paramétriques** : elles ne nécessitent aucune hypothèse sur la distribution des temps de survie.

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 1970 est identique à celle d'un groupe de patients inclus en 1990

Méthode Actuarielle n > 200	Méthode de Kaplan-Meier n < 200
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.

Pour chaque intervalle de temps on définit

- **V** : Nombre de **sujets vivants** au début de l'intervalle
- **D** : Nombre de **sujets décédés** dans l'intervalle
- **C** : Nombre de **sujets vivants aux dernières nouvelles**, dont le temps de participation s'arrête dans l'intervalle = **censure**

N : Nombre de **sujets exposés** au risque d'événement sur l'intervalle

Méthode Actuarielle	Méthode de Kaplan-Meier
$N = V - \frac{C}{2}$	$N = V - C$

Probabilité d'événements durant l'intervalle : $\frac{D}{N}$

Survie sur cet intervalle = survie instantanée : $\frac{N-D}{N}$

Méthode Actuarielle n > 200	Méthode de Kaplan-Meier n < 200

La fonction de survie

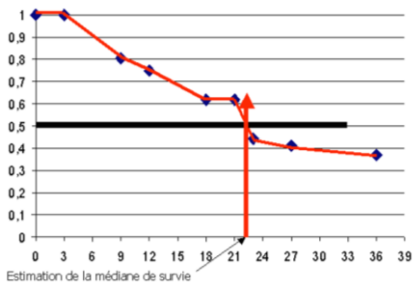
La fonction de survie S(t) est obtenue en faisant le **produit des survies instantanées** sur l'ensemble des intervalles

Exemple : la survie à 3 ans = (survie instantanée entre 2 et 3 ans) x (survie instantanée entre 1 et 2 ans) x (survie instantanée entre 0 et 1 an)

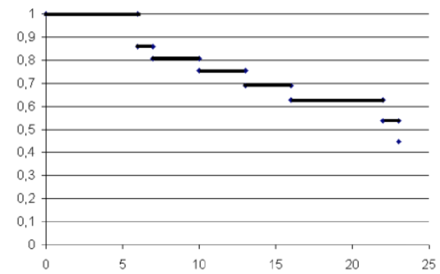
Instants	V	C	D	N = V – C/2	(N - D) / N	S(t)
0	-	-	-	-	-	1
3	210	0	0	210	**1**	1x1 = **1**
6	210	10	40	210 - 5 = 205	(205-40)/205 = **0,805** →	0,805 x 1 = **0,805**

Instants	V	C	D	N = V - C	(N - D) / N	S(t)
0	21	-	-	-	-	1
6	21	0	3	21	**0,857**	**0,857**
7	18	1	1	17	**0,941**	**0,807**

Courbe de survie



Pour chaque intervalle de temps, on représente l'estimation de la survie S(t) par un **point**.



La courbe de survie se compose de **paliers successifs**, où les probabilités de survie sont constantes entre deux temps d'événements consécutifs. Le premier palier vaut 1 depuis l'origine jusqu'au délai de survenue du premier événement et s'abaisse ensuite.

3. Choix d'une valeur résumée

- **Médiane de survie** : Elle 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 %.**
  (Nota : La moyenne de survie n'est pas un bon indicateur)

Il arrive que la fonction de survie soit toujours supérieure à 50 %. Dans ce cas, la médiane ne peut être estimée. On estime alors les quantiles

- **Quantiles de survie** : Pour le $p^{ième}$ quantile on estime la **durée pour laquelle la probabilité de survie est de 100 - p.**
  Exemple : le 25e quantile (ou 1er quartile) correspond à la plus petite durée pour laquelle la survie est inférieure à 75 %

- **Survie à date fixée** : Estimation de la survie à un **temps donné**.
  Exemple : survie à 5 ans

IV. COMPARAISON DE DEUX FONCTIONS DE SURVIE : Méthode du log-rank

On compare la survie de 2 groupes de personnes lorsqu'on souhaite démontrer qu'une **action** (opération, traitement) a un **lien avec leur survie**. On met donc en œuvre un test d'hypothèses.

1. Le test du log-rank

- On utilise un test du X^2 avec ddl de 1
- Ce test est **généralisable au cas de k groupes** et permet de tester si globalement la survie est différente entre les groupes.
- **H0** : les **fonctions de survie sont les mêmes dans les deux populations** d'où sont issus les groupes A et B. SA(t)=SB(t)
- **H1** : les deux fonctions de survie diffèrent

2. Le principe du test

Le principe du test du log-rank est de **comparer**, dans chaque groupe, le **nombre observé** et le **nombre attendu** d'événements **si la survie était identique dans les deux groupes**, sur l'ensemble de la période étudiée.

On obtient le nombre attendu d'événements en appliquant au **nombre de sujets exposés au risque d'événements**, la **proportion d'événements observés sur l'ensemble des deux groupes.** On le fait pour chaque intervalle.

Attention aux biais :

1) Il ne faut pas assimiler l'efficacité du traitement à la réponse des patients à ce traitement
2) Il ne faut pas comparer la survie entre les sujets qui répondent au traitement et les sujets qui ne répondent pas *(parce qu'ils prennent tous le TTT)* mais entre les patients traités et les patients non traités

3. Estimation des décès

On estime la **probabilité de décéder à t_i sachant que l'on est vivant à t_{i-1}** : c'est-à-dire estimer **1- S (t_i / t_{i-1})**. On fait ceci pour **chacun des temps de décès observés** t_i. → On utilise ici l'estimateur de Kaplan-Meïer

Nota : S (t_i / t_{i-1}) = probabilité de décéder après 1 sachant que l'on est vivant à 1

Exemple : On souhaite comparer 2 groupes patients : Groupe A = 21 patients et Groupe B : 21 patients.

On étudie la probabilité de décéder à ti sachant que l'on est vivant à ti-1 (dernière colonne du tableau) pour les 42 patients, sans distinction de groupe.

Survie instantanée — Probabilité de décès

t_i	V	C	N = V - C	D	$S(t_i / t_{i-1}) = (N - D) / N$	$1 - S(t_i / t_{i-1})$
1	42		42	2	0,952	0,048
2	40		40	2	0,950	0,050

4. Calcul des décès attendus

On estime ensuite le nombre de décès « E » que l'on attend dans chacun des groupes A et B, à chaque t_i, en supposant que la probabilité conditionnelle de décès estimée s'applique identiquement à chacun des deux groupes. Pour cela on évalue à chaque t_i l'effectif à risque à cette date. On obtient les deux dernières colonnes du tableau suivant :

t_i	V	C	N = V - C	D	$S(t_i / t_{i-1}) = (N - D) / N$	$1 - S(t_i / t_{i-1})$	N_A	N_B	E_A	E_B
1	42		42	2	0,952	0,048	21	21	1,000	1,000
2	40		40	2	0,950	0,050	19	21	0,950	1,050

Ces nombres sont notés E_{Ai} et E_{Bi}. On remarque que l'on utilise ici, comme toujours, la justesse supposée de l'hypothèse nulle puisque les probabilités de décès, et donc la survie, sont supposées ne pas dépendre du groupe. (*Rappel : l'hypothèse nulle c'est que les deux groupes sont identiques*). Sous l'hypothèse nulle ces nombres doivent être voisins des nombres de décès réellement observés. En particulier le total de ces nombres de décès au cours du temps (noté EA et EB selon le groupe) doit être voisin du nombre total de décès observés (noté DA et DB selon le groupe), et ceci dans chacun des groupes. Dans l'exemple, on obtient : EA=10,74 ; EB=19,26 ; DA=21 ; DB=9.

Petite explication : (*de la fiche de nos très très vieux cc Robin)*

Estimation du nombre de décès (théorique) E que l'on attend dans chacun des groupes A et B, à chaque t_i. Le nombre de décès attendus au temps t_i dans chaque groupe, est proportionnel aux effectifs à risque de chaque groupe.

N_A : Effectifs du groupe A à t_i. juste avant que ne se produisent le ou les décès

N_B : Effectifs du groupe B à t_i. juste avant que ne se produisent le ou les décès.

N : Effectifs global à t_i.avant que ne se produisent le ou les décès (- les censuré). **N = N_A + N_B**

D : Nombre de décès global à t_i. (groupe A et B confondus). **D = DA + DB**

DA : Nombre de décès observés dans le groupe A à t_i. (DA n'apparait pas dans le tableau).

DB : Nombre de décès observés dans le groupe B à t_i. (DB n'apparait pas dans le tableau).

E_A : Nombre de décès attendus à t_i.dans le groupe A, dans le cas où les fonctions de survie sont les même dans les 2 groupes (H0 accepté = pas de différence entre les groupes A et B).

E_B : Nombre de décès attendus à t_i.dans le groupe B, dans le cas où les fonctions de survie sont les mêmes dans les 2 groupes (H0 accepté = pas de différence entre les groupes A et B).

$E_A = D \times \frac{NA}{N}$ $\quad$ $E_B = D \times \frac{NB}{N}$ $\quad$ $E_A + E_B = D$

Exemple : Pour t_2 : N_A=19 et N_B=21 $\quad$ $E_A = 2 \times \frac{19}{40} = 0{,}950$ $\quad$ $E_B = 2 \times \frac{21}{40} = 1{,}050$

t_i	V	C	N = V - C	D	$S(t_i / t_{i-1}) = (N - D) / N$	$1 - S(t_i / t_{i-1})$	N_A	N_B	E_A	E_B
1	42		42	2	0,952	0,048	21	21	1,000	1,000
2	40		40	2	0,950	0,050	19	21	0,950	1,050

NB : Il y a un nombre plus grand de personnes « à risque » dans le groupe B, il doit donc avoir un nombre de décès attendus plus grand. C'est le cas, EB > EA et (0,950 + 1,050 = 2).

5. Le test du Chi-2

On va calculer le paramètre du test à partir de quatre valeurs :

$$\boxed{Q_c = \frac{(D_A - E_A)^2}{E_A} + \frac{(D_B - E_B)^2}{E_B}}$$

On a 1 degré de liberté.

Conditions de validité du test : **E_A** et **E_B** >5

➔ On compare Qc à la valeur seuil observée dans la table du X2 à 1 degré de liberté (1 ddl) au risque de 5%.

Globalement le prof veut que vous connaissiez le principe des analyses de survies, il passe parfois un peu vite en cours sur certaines parties : c'est qu'il s'en fout ! Bon courage pour la fin et n'oubliez pas : vous vous êtes engagés dans des études difficiles, passionnantes mais avant tout humaines ! (je paraphrase un ancien tut de biocell<3)