ANALYSE DE LA SURVIE

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.

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*).

Exemple : On teste un traitement contre le cancer. A un groupe on donne le traitement, à l'autre on donne un placébo. On étudie dans les deux groupes la survenue dans le temps des décès. On compare les résultats afin d'établir l'efficacité éventuelle du traitement.

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

• **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 :** 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 ⇨ 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'événement 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 vus (C) et ceux vivant à la date de point (B)

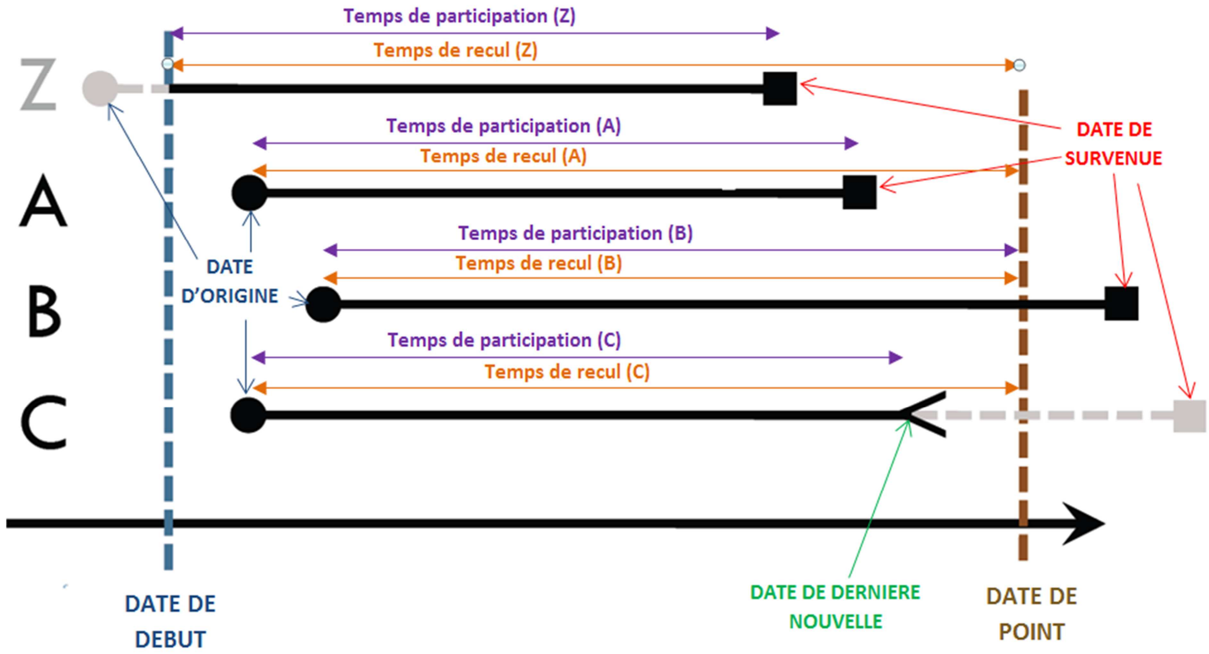
• **Temps de recul** : Délai entre la date d'origine et la date de point, c'est-à-dire le délai maximum potentiel de suivi pour un sujet. Les reculs minimum et maximum d'une série de sujets définissent donc l'ancienneté de cette série.

• **Temps de participation** : Durée de surveillance pour chaque sujet utilisée dans l'estimation de la survie. Trois cas :

1er cas: L'événement a lieu au cours de la surveillance ⇨**Temps de participation = Date de survenue de l'évènement - Date d'origine**.

2e cas : Le sujet est vivant à la date de point ⇨**Temps de participation = Date de point - Date d'origine**

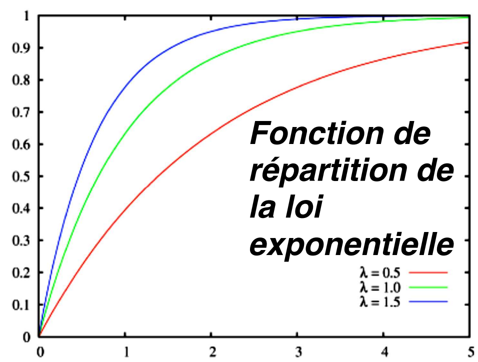
3e cas : Le sujet est perdu de vue ⇨ **Temps de participation = Date de dernière nouvelle - Date d'origine**



- ⇨ Le **patient A** est suivi entièrement durant la période de l'étude
- ⇨ Le **patient B** est encore vivant après la date de point. Son décès surviendra peut-être plus tard, mais HORS la période d'étude
- ⇨ Le **patient C** n'est pas suivi durant toute la période d'étude. Il était vivant jusqu'au moment où on l'a perdu de vue (date de dernière nouvelle)
- ⇨ Le **patient Z**, sa maladie est apparue avant la date de début d'étude (cohorte historique)

II. Fonction de survie

A. Loi exponentielle



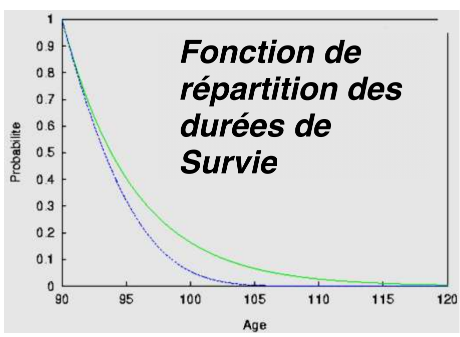
Nota : La loi exponentielle est utilisée couramment pour représenter la durée de vie de composants ou d'équipements pour lesquels on suppose que le taux de défaillance "λ" est constant au cours du temps (la durée de vie au-delà de « t » est indépendante de « t »)

- **Fonction de répartition de la loi exponentielle :**

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

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

B. 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 $S(t)$.

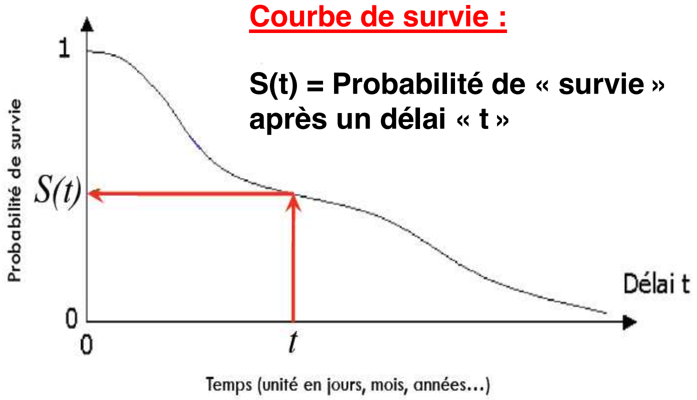
- **La fonction de survie S(t) :** C'est la probabilité pour que l'événement d'intérêt « T » (le décès par exemple) intervienne après un délai supérieur à « t ». Autrement dit, que l'événement d'intérêt « T » ne survienne pas avant la date « t ».

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

Exemple : L'événement d'intérêt est le décès : → Probabilité de survivre au moins jusqu'à la date « t »
→ Proportion « vraie » des survivants après un délai « t »

- **La courbe de survie :**

La courbe de survie est une fonction :
- Décroissante
- S(t) ϵ [0 ;1]



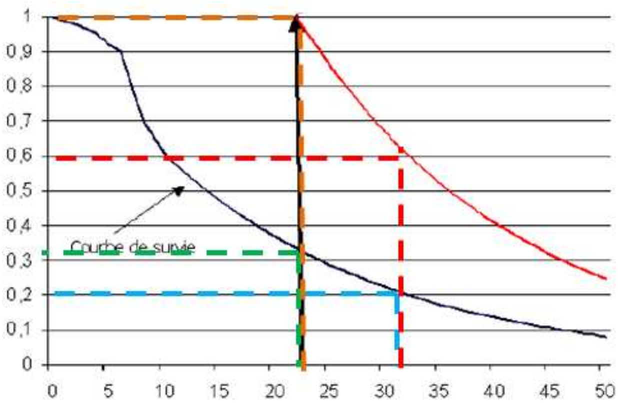
- **Probabilités :**

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

$$\mathbf{P(T\epsilon[\,t1\,;t2\,]) = S(t1) - S(t2)}$$

⇨ Probabilité de survivre encore après un délai « t » sachant que l'on est survivant après un délai « τ » (τ < t), que l'on note $S(t/\tau)$. On a donc $S(t) = P(T > t)$ et $S(\tau) = P(T > \tau)$

$$\mathbf{S(t\ \text{sachant}\ \tau) = S(t|\tau) = \frac{S(t)}{S(\tau)}}$$



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

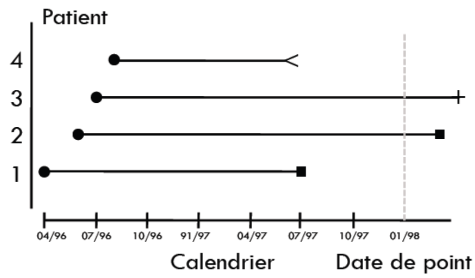
On lit sur la courbe bleue (celle de gauche) qui décrit la probabilité de survivre après « t »:
- ⇨ *Proportion de survivant à 23 ans =* ***33%*** *→ Il s'agit de la probabilité de survivre après 23 ans =* $S(t) = P(T > 23)$
- ⇨ *Proportion de survivant à 33 ans =* ***20%*** *→ Il s'agit de la probabilité de survivre après 33 ans =* $S(t) = P(T > 33)$

On lit sur la courbe rouge (celle de droite) qui décrit la probabilité de survivre après « t » ***sachant*** *que l'on est vivant à 23 ans (celle de droite) :*
- ⇨ *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 =* $S(23|23) = \frac{S(23)}{S(23)} = \frac{0{,}33}{0{,}33} = \mathbf{1}$
- ⇨ *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 =* $S(33|23) = \frac{S(33)}{S(23)} = \frac{0{,}2}{0{,}33} = \mathbf{0,6}$

III. Estimation de la survie

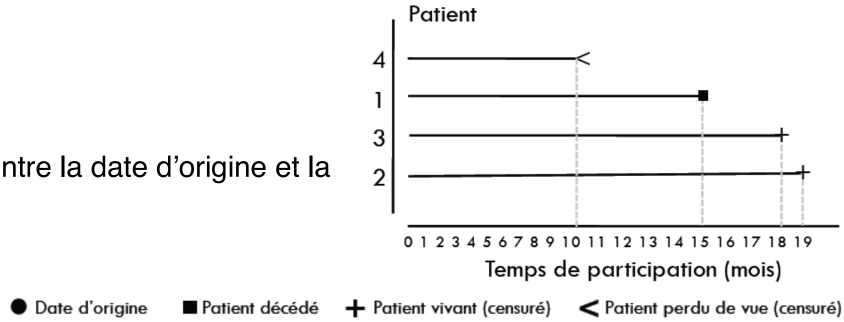
A. Recueil des données

Dans le cadre de l'estimation de la survie, on s'intéresse à :
⇨ La **date d'origine**, ex : diagnostic du cancer du sein
⇨ La **date des dernières nouvelles**, ex : date de décès ou date des dernières données dont on dispose
⇨ La **date de point :**
→ Censure des patients chez qui l'événement d'intérêt n'a pas été observé
→ Censure à la date de dernière nouvelle des patients « perdus de vue »
⇨ Un **événement binaire** « Tout ou rien » :



B. Calcul des durées de suivi

Les durées de suivi correspondent au délai entre la date d'origine et la date des dernières nouvelles qui peut être :
- ⇨ La date de décès
- ⇨ La date de point
- ⇨ La date de perte de vue



C. Calcul de la survie

Les 2 méthodes d'analyse non paramétriques de préférence 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 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 évènements 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 :

- ⇨ **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 sujet exposés au risque d'événement sur l'intervalle

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

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

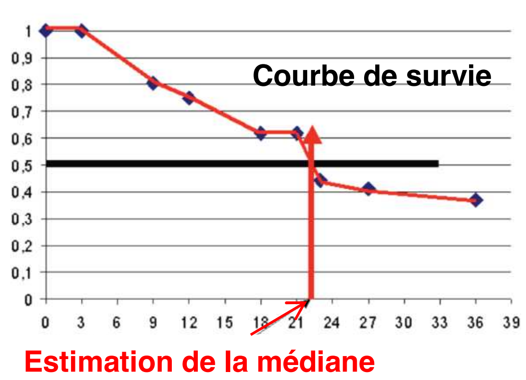
Probabilité de survie = survie instantanée : $\frac{N-D}{N}$

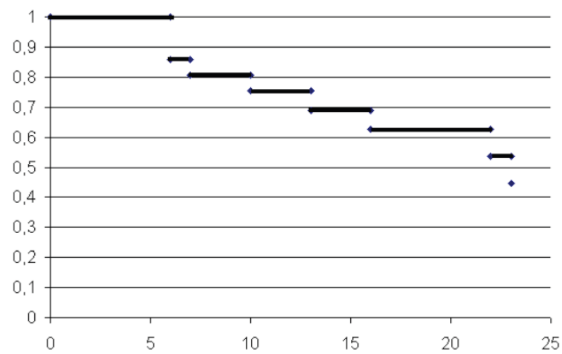
La fonction de survie $S(t)$ est estimée en faisant le **produit des survies instantanées** calculées sur tous les intervalles. *Exemple : Survie à 3 ans =* $S(3)$ = *(Survie instantanée entre 2 et 3 ans) x (survie instantanée entre 1 et 2 ans) x (survie instantanée entre 0 et 1 an)*

Instant	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**

Instant	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 la survie





Pour chaque intervalle de temps l'estimation de la survie est **représentée par un point**.

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

D. 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)

• **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*

E. Estimation de la fonction de survie

On cherche à estimer localement la fonction de survie $\boldsymbol{S(t_i)}$ pour « $\boldsymbol{i}$ » fixé :

⇨ $\boldsymbol{S(t_i)}$ **:** Probabilité pour que le décès survienne après $t_i \rightarrow P(T > t_i)$

⇨ $\boldsymbol{S(t_{i-1})}$ **:** Probabilité pour que le décès survienne après $t_{i-1} \rightarrow P(T > t_{i-1})$

⇨ $\boldsymbol{S(t_i|t_{i-1})}$ **:** Probabilité de survie = survie instantanée (= $\frac{N-D}{N}$ dans le cas de la méthode actuarielle ou Kaplan-Meier)

$$\boldsymbol{S(t_i) = S(t_i|t_{i-1}) \times S(t_{i-1})}$$

Exemple précèdent p3: $S(33_{ans}) = S(33_{ans}|23_{ans}) \times S(23_{ans}) = 0{,}6 \times 0{,}33 = 0{,}2$

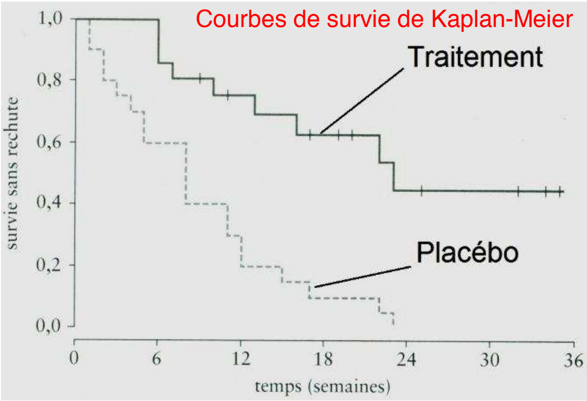
IV. Comparaison de 2 fonctions de survie

On compare la survie de 2 groupes de personnes lorsqu'on souhaite démontrer qu'une action (opération, traitement) a un lien avec leur survie.

A. Méthode du log-rank

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.

Attention aux biais !: Ne pas comparer la survie des patients répondant au traitement à la survie de ceux n'y répondant pas (dans ce cas les 2 groupes reçoivent bien le même traitement) ⇨ Il faut comparer la survie des patients recevant un traitement, à la survie de ceux qui n'en reçoivent aucun, ou un différent.



1. *Le Test du log-rank*

• Test du χ^2 à 1 degré de liberté (1 ddl)
• **H0** : Les fonctions de survie sont les mêmes dans les 2 groupes
• **H1** : Les fonctions de survie sont différentes
• **Test généralisable** à plusieurs groupes

2. *Principe du log-rank*

a) Estimation des décès : 1ère étape

⇨ Utilisation de la méthode de Kaplan-Meier

⇨ $\boldsymbol{1 - S(t_i|t_{i-1}) = 1 - \frac{N-D}{N}}$: Estimation, tous groupes confondus, de **la probabilité de décéder** à $\boldsymbol{t_i}$ sachant que l'on est vivant à $\boldsymbol{t_{i-1}}$, c'est-à-dire estimer $\boldsymbol{1 - S(t_i|t_{i-1})}$ pour chacun des temps de décès observés $\boldsymbol{t_i}$

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

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

On étudie la probabilité de décéder à t_i sachant que l'on est vivant à t_{i-1} ⇨ $1 - S(t_i|t_{i-1})$ ***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

b) Calcul du nombre « E » de décès attendus : 2ème étape

⇨ 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) **$D = D_A + D_B$**

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

D_B : Nombre de décès **observés** dans le groupe **B** à t_i (D_B 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)

$$\mathbf{E_A = D \times \frac{N_A}{N} \qquad E_B = D \times \frac{N_B}{N} \qquad E_A + E_B = D}$$

<u>Exemple</u> : Pour t_2 *:* ***$N_A = 19$ et $N_B = 21$***. $\mathbf{E_A = 2 \times \frac{19}{40} = 0,950}$ et $\mathbf{E_B = 2 \times \frac{21}{40} = 1,050}$

<u>Nota</u> : j'ai un nombre plus grand de personnes « à risque » dans le groupe B, je devrais donc avoir un nombre de décès attendus plus grand. C'est le cas, $E_B > E_A$ et $(0,950 + 1,050 = 2)$

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

On calcule pour finir le **nombre TOTAL de décès attendus** pour les groupes A et B, ainsi le **nombre TOTAL de décès observés** pour les groupes A et B.

$D_{A\ total} = \sum D_A \qquad D_{B\ total} = \sum D_B \qquad E_{A\ total} = \sum E_A \qquad E_{B\ total} = \sum E_B$

c) Test du χ^2

⇨ On calcule la valeur du test du χ^2: $Q_c = \frac{(D_{A\ total} - E_{A\ total})^2}{E_{A\ total}} + \frac{(D_{B\ total} - E_{B\ total})^2}{E_{B\ total}}$

⇨ Conditions de validités : $E_{A\ total} \geq 5$ et $E_{B\ total} \geq 5$

⇨ On compare Q_c à la valeur seuil observée dans la **<u>table du χ^2 à 1 degré de liberté</u>** (1 ddl) au risque de 5% (la valeur lue est 3,84)

⇨ Si $Q_c \in [0 ; 3,84]$ appelé **intervalle de Paris** → On ne rejette pas H0, on ne peut pas conclure à une différence entre les fonctions de survie dans deux groupes A et B. Si $Q_c > 3,84$, on rejette H0, on peut conclure à une différence.

d) Meilleure survie

Si H0 est rejetée, la survie est la meilleure dans le groupe dont le nombre de décès observé est inférieur au nombre de décès attendus → $D_{total} < E_{total}$

<u>Exemple :</u> $D_{A\ total} = 21$, $D_{B\ total} = 9$, $E_{A\ total} = 10,74$, $E_{B\ total} = 19,26$ → $Q_c = 15,26 > 3,84$ → *Rejet de H0*

$D_{A\ total} > E_{A\ total}$ *et* $D_{B\ total} < E_{B\ total}$ *→ Le groupe B est donc celui qui a la meilleure survie au risque de 5%.*