

Chapitre 10. Analyse de la survie

D.M. Parkin¹ et T. Hakulinen²

¹*Centre International de Recherche sur le cancer,
150 cours Albert-Thomas, 69372 Lyon Cédex 08, France*

²*Registre de Cancer, Liisankatu 21B, 00170 Helsinki, Finlande*

Introduction

Les registres de cancers collectent des informations sur tous les cas de cancers survenant dans des zones déterminées. Les taux de survie pour les différents cancers sont donc calculés à partir de données représentant le pronostic moyen dans la population et fournissent, théoriquement, un index objectif de l'efficacité de la prise en charge des patients cancéreux dans les régions concernées. A l'opposé, les registres hospitaliers s'intéressent aux résultats des patients traités dans une seule institution, et peuvent être appelés à évaluer l'efficacité des différentes thérapeutiques.

Ce chapitre s'intéresse principalement à la description des méthodes de calcul de survie pour des données de population. Cependant, les méthodes analytiques s'appliquent également aux données hospitalières, et peuvent être utilisées pour décrire le devenir de tout groupe de patients cancéreux. Mais il faut souligner que l'analyse descriptive de la survie n'est pas suffisante pour évaluer l'efficacité des différents traitements, qui ne peut l'être que par un essai clinique *ad hoc*.

Définition de cas

La première étape d'une analyse de la survie consiste à déterminer clairement les groupes de patients à partir desquels les calculs seront faits. Ceux ci sont généralement définis par :

- Le type de cancer (site et/ou histologie),
- La date du diagnostic,
- Le sexe,
- Le stade de la maladie.

Le stade de la maladie est généralement présenté par de simples catégories, un maximum de quatre, qui sont dérivées de l'évaluation clinique (voir chap. 6, item 23), ou de l'évaluation chirurgicale et anatomopathologique (chap. 6, item 24). Les résultats peuvent être exprimés par groupes d'âge, d'origine ethnique, de modalité de traitement, etc....

Un registre de population devrait limiter ses analyses de la survie aux cas résidant dans la zone du registre, car on peut penser que les patients qui viennent spécialement dans un lieu donné pour être traités, constituent un groupe atypique avec une survie différente de la moyenne.

Il faut définir la nature des cas à inclure dans l'étude. Par exemple, il faut décider si on inclut les cas pour lesquels la base la plus valide de diagnostic est l'examen clinique. Les cas enregistrés sur la base du seul certificat de décès (CDS) posent un problème particulier. Pour de tels cas, la date d'incidence (chap. 6, item 16) est nécessairement la même que la date du décès, ils ont donc théoriquement une survie nulle. Un problème se pose aussi pour les cas diagnostiqués pour la première fois à l'autopsie.

Selon Hanai et Fujimoto (1985), quand des cas sont enregistrés sur la base du seul certificat de décès, il peut être admis qu'un nombre équivalent de cas encore vivants ait échappé à l'enregistrement au moment du diagnostic et ne figure toujours pas dans les données du registre. Si cette affirmation est vraie, l'inclusion des cas CDS aboutirait à sous-estimer le taux de survie par l'inclusion d'un excès de cas décédés dans le registre. De plus, comme la date d'incidence (chap 16, item 16) et la date du décès (item 32) sont les mêmes, la durée de la survie est égale à 0. Dans l'évaluation de la survie cumulée par la méthode actuarielle (voir plus loin), de tels individus sont inclus dans les personnes survivant moins d'une année et le taux de survie à un an est artificiellement réduit. Cependant, si ces cas forment une proportion substantielle de la totalité des cas enregistrés, leur exclusion biaisera aussi le calcul qui ne reflétera plus le vrai taux de survie des cancers dans la communauté.

Quand la durée de la maladie est notifiée sur le certificat de décès, cette information peut être utilisée pour fixer la date du diagnostic (ou la date d'incidence), alors les cas CDS peuvent être inclus, sinon, le choix est arbitraire. La pratique la plus courante est d'omettre les cas CDS, mais c'est probablement parce que la plupart des travaux publiés sur la survie proviennent de registres ayant une assez faible proportion de tels cas. Une solution alternative est de présenter deux taux de survie, un pour les cas incidents et les cas CDS, et un autre présentant les cas sans les CDS. La proportion de CDS devra toujours être spécifiée sur les publications traitant de la survie.

Définition de la date de début

Pour le registre de population, la date de début (à partir de laquelle la survie est calculée) est la date d'incidence (chap 6, item 16). Pour les registres hospitaliers, la date d'admission à l'hôpital est souvent utilisée. Quand la survie sert à mesurer le résultat final d'un traitement, la date du début de ce traitement est plus appropriée. Dans les essais cliniques, quand on compare les résultats de plusieurs traitements, on peut utiliser la date de randomisation (Peto et coll, 1976, 1977).

Suivi

Pour calculer la survie, il faut suivre les cas enregistrés pour savoir si les patients sont vivants ou morts.

Suivi passif

Il repose sur le signalement du décès des patients enregistrés grâce au fichier des certificats de décès de la région. La fusion des deux fichiers, le fichier des certificats de décès dépendant de l'état civil et le fichier des cas enregistrés au registre, est réalisé par le registre des cancers ou dans le service local ou national de l'état civil. Dans ce processus d'appariement, un numéro national d'identification (s'il existe) ou une combinaison de plusieurs indices tels que le nom, la date de naissance et l'adresse est utilisée pour l'identification du patient.

Dans le suivi passif, tout patient cancéreux enregistré pour lequel le décès n'a pas été notifié au registre par le service de l'état civil (en d'autres termes, tout cas non apparié) est considéré comme vivant. Le suivi passif peut, conduire à surestimer le vrai taux de survie : l'importance de l'erreur tient d'une part à l'exactitude du processus d'appariement et d'autre part aux cas des migrants. Occasionnellement, il est possible d'accéder à un fichier des migrants (cf. en Finlande), ainsi ces cas peuvent être exclus de la liste des cas suivis et ne pas être considérés comme étant toujours en vie.

Suivi actif

Certains registres recueillent des informations par l'intermédiaire de tous les registres hospitaliers qui existent dans leur zone d'étude. Ces derniers effectuent régulièrement un suivi de leurs malades grâce au médecin du patient. Ce type de suivi porte le nom de suivi médical et il permet généralement d'améliorer tant la qualité, que la durée des études de survie.

La majorité des registres de population ne disposent pas d'un système de suivi individuel, mais ils peuvent utiliser des listes établies à d'autres fins pour connaître indirectement la survie. Des sources d'information aussi diverses que, le recensement, les listes d'inscription dans un service de

santé, à la sécurité sociale, dans une mutuelle, ou encore les listes électorales peuvent être utilisées. Ces techniques peuvent évidemment venir en complément du suivi médical.

Le suivi actif met en évidence un certain nombre de patients pour lesquels il ne peut y avoir de suivi et dont le statut vital est inconnu. Quand la survie est calculée par la méthode actuarielle (voir plus loin), il faut considérer de tels patients comme vivants et résidants dans la région (et donc faisant partie de la population soumise au risque) durant exactement la moitié de la période pour laquelle on a eu de leurs nouvelles. Cependant, il est probable que la plupart d'entre eux sont encore vivants ; s'ils étaient morts, le registre l'aurait appris par leur certificat de décès. Dans ce cas, il en résulte un biais qui sous estime les taux de survie. Il est donc important d'avoir un minimum de patients perdus de vue.

Intervalle de survie

La survie peut s'exprimer en pourcentage des cas présents initialement, et encore vivants après un intervalle donné. Le choix de l'intervalle est arbitraire et dépend du pronostic du cancer étudié. Dans l'interprétation des taux de survie, on doit prendre en compte le nombre d'individus présents au début de l'intervalle. On ne devrait pas publier des taux calculés avec moins de dix patients présents initialement à cause de l'instabilité des résultats estimés avec si peu de sujets. Les méthodes exposées dans ce chapitre permettent de décrire la survie observée globalement dans un groupe de cancéreux. Dans une étude de survie, il est important de calculer la survie pour des délais successifs. Soulignons que le taux de survie à cinq ans est conventionnellement utilisé pour comparer la survie entre groupes de patients selon leur localisation cancéreuse, leur sexe..., et qu'il est souvent utilisé comme une mesure du taux de guérison ; mais il est évident que pour de nombreuses localisations, un délai de cinq ans est insuffisant (Hakulinen, 1981).

Calcul des taux de survie

La section suivante est issue du "Reporting of Cancer Survival and End Result", 1982, publié par l'American Joint Committee on Cancer.

Dans les registres de cancers, on calcule généralement, les taux de survie pour des cas enregistrés durant une période de plusieurs années avant la date donnée. Dans les exemples ci-dessous, ces principes sont illustrés pour un très petit groupe de patients (50) chez qui un mélanome a été diagnostiqué durant les quinze années précédant le 1er juin 1985. La survie de ces patients peut être évaluée grâce à une surveillance effectuée jusqu'à fin 1987, c'est à dire que la date de fin d'étude est fixée au 31 décembre 1987. Les données de base sont présentées dans le tableau 1.

Tableau 1. Données concernant 50 patients atteints de Mélanome

| Numéro de patient | Sexe | Age | Date de diagnostic mois/ année) | Dernières nouvelles | | | Années vécues depuis diagnostic |
|-------------------|------|-----|---------------------------------------|---------------------|--------|-------------------|---------------------------------------|
| | | | | Date (M/A) | Etat | Cause du décès | |
| 1 | M | 63 | 10/70 | 10/70 | Mort | Mélanome | 0 |
| 2 | M | 42 | 7/72 | 1/78 | M | Autre | 5 |
| 3 | M | 41 | 3/73 | 4/73 | M | M | 0 |
| 4 | F | 57 | 6/73 | 7/74 | M | M | 1 |
| 5 | M | 35 | 9/73 | 10/87 | Vivant | | 14 |
| 6 | F | 48 | 10/73 | 8/74 | M | M | 0 |
| 7 | M | 43 | 4/74 | 2/77 | M | M | 2 |
| 8 | F | 27 | 1/75 | 1/75 | M | M | 0 |
| 9 | F | 56 | 12/76 | 10/87 | V | | 10 |
| 10 | F | 33 | 1/77 | 11/87 | V | | 10 |
| 11 | F | 37 | 4/77 | 4/87 | V | | 10 |
| 12 | F | 58 | 9/77 | 8/87 | V | | 9 |
| 13 | M | 21 | 2/77 | 5/78 | M | A | 0 |
| 14 | M | 71 | 2/78 | 11/86 | V | | 8 |
| 15 | F | 66 | 6/79 | 8/79 | M | M | 0 |
| 16 | F | 35 | 7/79 | 12/87 | V | | 8 |
| 17 | F | 31 | 10/79 | 11/87 | V | | 8 |
| 18 | M | 35 | 3/80 | 6/87 | V | | 7 |
| 19 | F | 44 | 4/80 | 7/87 | V | | 7 |
| 20 | M | 26 | 4/80 | 10/87 | V | | 7 |
| 21 | M | 57 | 10/80 | 6/81 | M | M | 0 |
| 22 | M | 54 | 12/80 | 2/81 | M | M | 0 |
| 23 | M | 63 | 1/81 | 1/82 | M | M | 1 |
| 24 | F | 32 | 1/81 | 10/83 | M | M | 2 |
| 25 | F | 43 | 4/81 | 2/87 | V | | 5 |
| 26 | F | 76 | 7/81 | 2/87 | M | M | 4 |
| 27 | M | 31 | 9/81 | 11/87 | V | | 6 |
| 28 | M | 77 | 11/81 | 2/87 | V | | 5 |
| 29 | F | 59 | 11/81 | 4/87 | V | | 5 |
| 30 | F | 76 | 12/81 | 9/87 | V | | 5 |
| 31 | M | 39 | 3/82 | 8/85 | M | M | 3 |
| 32 | F | 50 | 7/82 | 4/87 | V | | 4 |
| 33 | F | 38 | 10/82 | 6/87 | M | M | 4 |
| 34 | F | 82 | 3/83 | 12/87 | V | | 4 |
| 35 | M | 65 | 4/83 | 7/83 | M | M | 0 |
| 36 | M | 40 | 4/83 | 10/87 | V | | 4 |
| 37 | M | 22 | 6/83 | 2/87 | V | | 3 |
| 38 | F | 25 | 1/84 | 11/87 | V | | 3 |
| 39 | M | 33 | 4/84 | 11/87 | V | | 3 |
| 40 | F | 51 | 5/84 | 10/87 | V | | 3 |
| 41 | F | 40 | 7/84 | 11/87 | V | | 3 |
| 42 | M | 70 | 9/84 | 9/85 | M | A | 1 |
| 43 | M | 47 | 9/84 | 12/85 | M | M | 1 |
| 44 | M | 67 | 10/84 | 4/86 | M | A | 1 |
| 45 | F | 58 | 1/85 | 8/87 | V | | 2 |
| 46 | M | 75 | 1/85 | 10/87 | V | | 2 |
| 47 | M | 40 | 4/85 | 7/87 | V | | 2 |
| 48 | F | 35 | 4/85 | 7/87 | V | | 2 |
| 49 | F | 49 | 5/85 | 12/86 | M | M | 1 |
| 50 | F | 21 | 6/85 | 3/87 | V | | 1 |

Calcul par la méthode directe

La façon la plus simple de résumer la survie d'un patient est de calculer le pourcentage de patients vivants à la fin d'un intervalle déterminé (par exemple cinq ans), en ne prenant en compte que les patients exposés au risque de mourir durant tout l'intervalle étudié. Cette technique est connue sous le nom de "méthode directe".

L'ensemble des données du tableau 1 montre que l'on a eu des nouvelles des patients durant l'année 1987, mais ces nouvelles sont parvenues à différents mois de l'année. On sait que tous les patients contactés en 1987 étaient encore vivants le 31/12/86. Donc, le 31/12/86 est considéré comme la date effective de fin d'étude. Cela signifie qu'à la date de fin d'étude, tous les patients traités pour la première fois à partir du 1/1/82 n'avaient pas été potentiellement exposés durant au moins cinq ans au risque de mourir. Ainsi, 20 des 50 patients (n°31 à 50) doivent être exclus du calcul par la méthode directe.

L'examen des données de la colonne "Etat aux dernières nouvelles" du tableau 1 pour les 30 premiers patients exposés au risque durant au moins cinq ans, indique que 14 d'entre eux étaient décédés avant le 31/12/86. Cependant, un de ces patients (n°2) avait vécu cinq années entières avant de décéder. Donc, 17 patients, parmi les 30, étaient encore vivants au cinquième anniversaire de leur premier traitement : en conséquence, le taux de survie à cinq ans est de 57%.

Calcul par la méthode actuarielle

Le calcul par la méthode directe n'utilise pas toute l'information disponible. Par exemple, les données indiquent que le patient n°31 est décédé quatre ans après traitement, et que le patient n°32 a survécu plus de quatre ans. Ces informations méritent d'être utilisées mais ne peuvent l'être si l'on respecte les règles de la méthode directe, ces patients ayant été diagnostiqués après décembre 1981.

La méthode actuarielle (ou des tables de survie) permet d'utiliser toute l'information fournie par le suivi jusqu'à la date de clôture de celui-ci. La méthode actuarielle a aussi l'avantage de donner des informations sur l'allure de la survie, montrant comment le groupe des patients a diminué durant la période d'observation (Culter et Ederer, 1958; Ederer et coll., 1961).

Les méthodes décrites ici sont bien adaptées à l'étude fine de la survie d'une petite série de patients (Dans cet exemple, 50 patients). Leur utilisation dans une série plus importante n'est possible que si l'on utilise un ordinateur pour les calculs (cf. Hakulinen et Abeywickrama, 1985).

Tableau 2. Calcul des taux de survie et de leur écart type par la méthode actuarielle

| Années écoulées depuis le diagnostic | Nb de vivants au début de l'année | Nb de décédés durant l'année | Nb de sujets vus vivants pour la dernière fois dans l'année | Nb de sujets exposés au risque de décès | Proportion de décédés durant l'année | Proportion de survivants durant l'année | Proportion de survivants du diagnostic à la fin de l'année | Col 5- Col 3 | Col 6 / Col 9 |
|---|---|---------------------------------|---|--|--|---|--|-----------------------------------|---------------------------|
| (i) | (l _i) | (d _i) | (w _i) | (r _i) ^a | (q _i) ^a | (p _i) ^a | (Πp _i) | (r _i -d _i) | $\frac{q_i}{(r_i - m_i)}$ |
| 0 | 50 | 9 | 0 | 50.0 | 0.180 | 0.820 | 0.820 | 41.0 | 0.0044 |
| 1 | 41 | 6 | 1 | 40.5 | 0.148 | 0.852 | 0.699 | 34.5 | 0.0043 |
| 2 | 34 | 2 | 4 | 32.0 | 0.063 | 0.937 | 0.655 | 30.0 | 0.0021 |
| 3 | 28 | 1 | 5 | 25.5 | 0.039 | 0.961 | 0.629 | 24.5 | 0.0016 |
| 4 | 22 | 2 | 3 | 20.5 | 0.098 | 0.902 | 0.567 | 18.5 | 0.0053 |
| >5 | 17 | - | 17 | - | - | - | - | - | - |
| Total | | 20 | 30 | | | | | | 0.0177 |

^a $r_i = l_i - \frac{w_i}{2}$; $q_i = d_i/r_i$; $p_i = 1 - q_i$

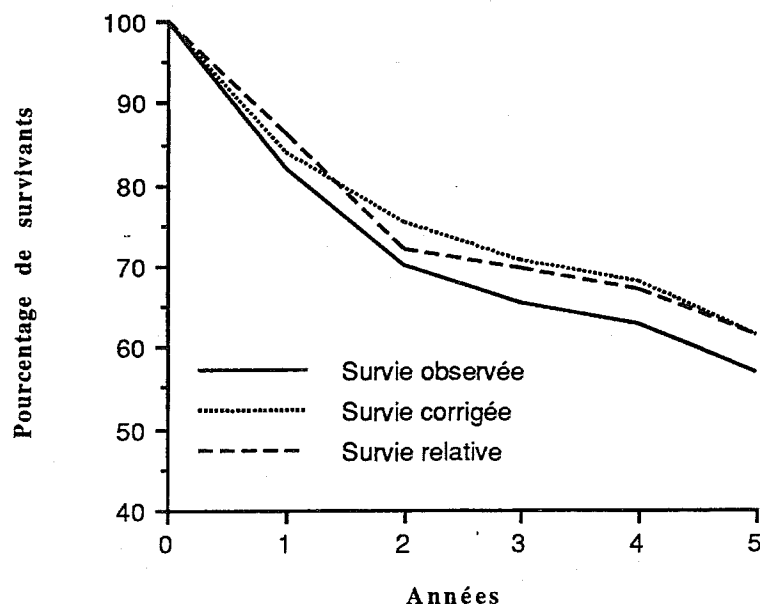


Figure 1. Taux de survie observée, corrigée et relative des mélanomes
Etablis à partir des données des tableaux 1 et 2.

Taux de survie observée

Un exemple d'utilisation de la méthode actuarielle utilisant l'ensemble de l'information connue sur les 50 patients vus précédemment, est donné dans le Tableau 2. La préparation de ce tableau compte sept étapes.

(1) Le statut vital des patients (vivant ou mort) et les exclusions survenant au cours de chacune des années suivant le diagnostic (voir Tableau 1), sont utilisés comme entrées des colonnes 3 et 4. Notons que 17 patients vivants au début de la dernière période après le diagnostic (5 ans et plus) figurant en colonne 2, ont aussi été reportés en colonne 4 (nombre de sujets vus vivants pour la dernière fois durant l'année).

(2) Le nombre de sujets vivants au début de chaque année est entré dans la colonne 2. Il est obtenu par des soustractions successives. Donc, sur les 50 patients diagnostiqués, 9 sont décédés durant la première année et 41 étaient encore vivants un an après le diagnostic. Dans le deuxième intervalle, 6 sont morts et 1 a disparu tout en étant encore vivant. Il restait donc 34 patients observables au début du troisième intervalle (deux ans après le diagnostic).

(3) Le calcul du nombre réel de sujets exposés au risque de décès (colonne 5) est basé sur l'hypothèse que les sujets perdus de vue ont été observés au moins durant la moitié de l'année. Donc, le nombre de sujets

exposés au risque durant la troisième année est de : $34 - (1/2 \times 4) = 32,0$, et pour la quatrième année de : $28 - (1/2 \times 5) = 25,5$.

(4) La proportion de sujets décédés durant une année quelle qu'elle soit (colonne 6) est calculée en divisant le nombre figurant dans la colonne 3 par celui figurant en colonne 5. Donc, pour la première année, cette proportion est de : $9/50,0 = 0,18$, et pour la deuxième année de : $6/40,5 = 0,148$.

(5) La proportion de survivants durant l'année, qui est le taux de survie observée dans l'année est égale à : $1 - (\text{Proportion de décédés})$, (colonne 6).

(6) La proportion de survivants entre le diagnostic et la fin de chaque année, qui est le taux de survie cumulé observé, est le produit des taux annuels de l'année et des années précédentes. Par exemple, pour la cinquième année, la proportion 0,567 est le produit des proportions calculées dans la colonne 7 de la première à la cinquième année.

Tableau 3. Calcul de la survie spécifique

| (1) Année après le diagnostic | (2) Nb de vivants au début de l'année | (3) Nb de morts au début de l'année | | (4) Nb de sujets vus pour la dernière fois dans l'année mais vivants | (5) Nb de sujets exposés au risque de décès | (6) Taux de décès dans l'année ^a | (7) Taux de survie en fin d'année ^a | (8) Taux de survie cumulée ^a |
|--|--|---|---------------------------|--|---|--|--|--|
| | | (a) Par la maladie | (b) Par autre cause | | | | | |
| (i) | (li) | (d(m)i) | (d(o)i) | (wi) | (ri) ^b | (qi) ^b | (pi) ^b | (Iipi) |
| 0 | 50 | 8 | 1 | 0 | 49.5 | 0.162 | 0.838 | 0.838 |
| 1 | 41 | 4 | 2 | 1 | 39.5 | 0.101 | 0.899 | 0.754 |
| 2 | 34 | 2 | 0 | 4 | 32.0 | 0.063 | 0.937 | 0.706 |
| 3 | 28 | 1 | 0 | 5 | 25.5 | 0.039 | 0.961 | 0.679 |
| 4 | 22 | 2 | 0 | 3 | 20.5 | 0.098 | 0.902 | 0.613 |
| >5 | 17 | - | - | 17 | | | | |
| Total | | 17 | 3 | 30 | | | | |

^a Note "décédés et survivants" des colonnes 5 à 8 font références aux morts (et non morts) de la maladie étudiée.

$$^b \text{ ou } r_i = l_i - \frac{(w_i + d(o)_i)}{2} \quad q_i = d(m)_i / r_i ; \quad p_i = 1 - q_i$$

Le taux de survie à 5 ans, calculé par la méthode actuarielle est de 0,567 environ 57%. Dans cet exemple, le résultat obtenu en utilisant l'information disponible sur les 50 patients est très proche de celui obtenu avec 30 patients par la méthode directe. Cette convergence des deux méthodes n'existe généralement pas lorsque beaucoup de patients ont été exclus du calcul par la méthode directe. Dans ce cas, la méthode actuarielle est plus fiable car basée sur davantage d'informations.

Un des avantages de la méthode actuarielle est de calculer le risque de décès pour chaque intervalle de temps. Ainsi, dans la colonne 6 (qi), on voit que la proportion de malades qui meurent passe de 18% la première année, à 4% la quatrième. L'augmentation à 10% la cinquième année est probablement attribuable au hasard car les effectifs étaient faibles, il ne restait que 22 patients vivants au début de la cinquième année.

Les taux cumulés de la colonne 8 peuvent être utilisés pour dessiner la courbe de survie, ce qui permet d'avoir une description graphique de la survie (figure 1).

Survie spécifique

Les taux de survie observée décrits plus haut prennent en compte tous les décès qu'elle qu'en soit la cause. Il s'agit d'une image exacte de la mortalité du groupe de malades, mais on souhaite généralement connaître la mortalité due à la maladie étudiée. On observe au tableau 1 que dans 4 cas le mélanome n'est pas la cause du décès (patient 2, 13, 42 et 44). Trois de ces décès étant survenus au début des cinq ans de suivi, cela joue sur le taux de survie à cinq ans, calculé dans le tableau 2.

Lorsque l'on dispose d'une information sur la cause du décès, il est possible de faire une correction pour les décès dus à une cause différente de celle étudiée.

Un exemple de calcul figure tableau 3. Les décès observés sont classés en 2 catégories : ceux dus à la maladie étudiée (colonne 3a), ceux dus à une autre cause (colonne 3b). Les patients morts d'une autre cause sont traités comme s'ils avaient été vus vivants pour la dernière fois durant l'année (colonne 4), car ces 2 groupes échappent au risque de mourir du mélanome. Donc, le nombre de sujets exposés au risque de décès (par mélanome) (colonne 5) durant la seconde année d'observation est de : $41 - (2+1)/2 = 39,5$.

Le taux de survie spécifique à 5 ans est de : 0,613, soit 61% comparé au taux observé de 57%. Le taux de survie spécifique indique que 61% des sujets atteints d'un mélanome n'en meurent pas dans les 5 premières années qui suivent le diagnostic.

L'utilisation de la survie spécifique est particulièrement importante lorsque l'on compare des groupes qui diffèrent par des caractéristiques comme l'âge, le sexe, la race, le statut socio-économique ou d'autres qui ont une influence importante sur le risque de mourir d'une autre cause que le cancer étudié. La figure 1 compare la survie observée et spécifique pour les 50 patients. L'écart entre les 2 courbes correspond à la mortalité normale (hors mélanome).

Taux de survie relative

La cause de décès est souvent inconnue ou peu fiable. Il est donc impossible de calculer la survie spécifique. Cependant, il est possible de tenir compte des différences qui existent dans la mortalité normalement attendue pour 2 groupes de sujets : elles correspondent aux différences de risque de mourir d'une autre cause que la maladie étudiée. Ainsi on calcule le taux de survie relative qui est le rapport de la survie observée sur la survie attendue pour un groupe de sujets de la population générale ayant les mêmes caractéristiques de race, d'âge et de sexe à la même période.

On obtient les probabilités de survie attendue en utilisant les tables de survie publiées chaque année pour la population générale. Ces probabilités dépendent de l'âge, du sexe et de l'année durant laquelle se déroule l'étude. A partir de l'exemple Finlandais, le tableau 4 donne les informations nécessaires au calcul de la survie pour le patient n°1, qui est un homme de 63 ans en 1970. En Finlande, les taux de mortalité annuels sont publiés par groupe d'âge de 1 an tous les 5 ans. Ils donnent une valeur moyenne pour 5 années calendaires. Le patient n°1 avait 63 ans dans la période 1966-1971 (en 1970 exactement) et a vécu durant les 5 années suivantes (correspondant à la période 1971-1975). Le taux de mortalité de la population correspondant à l'âge du patient est fourni par les tables de mortalité publiées sous forme de probabilités de décès (statistiques officielles de Finlande 1974-1980). La probabilité de survie est égale à : $1 - (\text{probabilité de décès})$.

Pour tenir compte du fait que le patient n'avait pas exactement 63 ans mais plutôt 63,5 ans, au début du suivi, une moyenne mobile est calculée avec les probabilités annuelles de survie normale. Les 5 probabilités correspondant aux âges 63,5 à 67,5 sont multipliées pour obtenir la probabilité de survie à 5 ans que l'on peut attendre normalement. Dans notre exemple, le résultat est de 0,812.

Pour l'ensemble du groupe du tableau 1, la moyenne des survies attendues est la somme des probabilités individuelles divisée par le nombre des individus (50). Dans ce cas, on trouve 0,94 ou 94%.

$$\text{Taux de survie relative} = \frac{\text{Taux de survie observée}}{\text{Taux de survie attendue}} \times 100 = \frac{0,57}{0,94} = 61\%$$

Dans ce cas, le taux est égal au taux de survie spécifique trouvé précédemment.

Tableau 4. Calcul de la probabilité de survie attendue en utilisant la mortalité générale de la population (en Finlande)

| Age | Période calendaire | Probabilité annuelle de décès ^a | Probabilité annuelle de survie | Moyenne mobile sur 2 ans |
|-----|-----------------------|---|-----------------------------------|-----------------------------|
| 63 | 1966-70 | 36.08 | 0.96392 | |
| 64 | 1971-75 | 35.28 | 0.96472 | 0.964320 |
| 65 | 1971-75 | 38.75 | 0.96125 | 0.962985 |
| 66 | 1971-75 | 41.40 | 0.95860 | 0.959925 |
| 67 | 1971-75 | 46.24 | 0.95376 | 0.956180 |
| 68 | 1971-75 | 48.52 | 0.95148 | 0.952620 |

^a Probabilité annuelle de décès pour 100 (Statistiques officielles de Finlande, 1974, 1980)

Il est d'usage, en pratique, de calculer le taux de survie relative pour chaque intervalle et cumulativement pour chaque intervalle successif du suivi (voir Ederer et coll., 1961).

La survie relative ne demande pas de connaître la cause du décès (ni de savoir si le cancer a causé la mort, ou y a contribué), ce qui est un avantage important (Hakulinen, 1977). Cependant, cela suppose que la population observée est soumise à la même force de mortalité que celle qui a servi à construire la table de mortalité. Quand il n'existe pas de table de mortalité appropriée (par exemple pour une race ou un groupe socio-économique), il vaut mieux calculer une mortalité spécifique. Si l'on veut comparer des groupes, il faut évidemment utiliser la même méthode de calcul pour tous les groupes.

Si l'on doit utiliser la survie relative pour des suivis de 10 ans ou plus, il faut lire l'article d'Hakulinen (1982) qui traite des biais liés au vieillissement de la population et aux différences dans les tendances évolutives de l'incidence du cancer pour chaque âge.

Calcul par la méthode de Kaplan-Meier

Très fréquemment, les logiciels proposent un calcul de la survie par la méthode de Kaplan-Meier (Kaplan-Meier, 1958). Cette méthode est semblable à une méthode actuarielle où les taux de survie cumulés ne sont pas calculés à la fin de chaque année, mais à des intervalles de temps beaucoup plus courts si tant est que la précision des informations sur la date de décès le permette.

L'exemple donné au tableau 5 reprend les données du tableau 1 où les durées de suivi sont données au mois près. Les calculs sont presque identiques à ceux de la méthode actuarielle, excepté le fait que l'on utilise des intervalles de temps d'un mois et que l'on considère que les patients exclus ont survécu durant tout l'intervalle (un mois).

La figure 2 montre une courbe de survie calculée par la méthode de Kaplan-Meier. Elle présente des paliers successifs et chaque changement

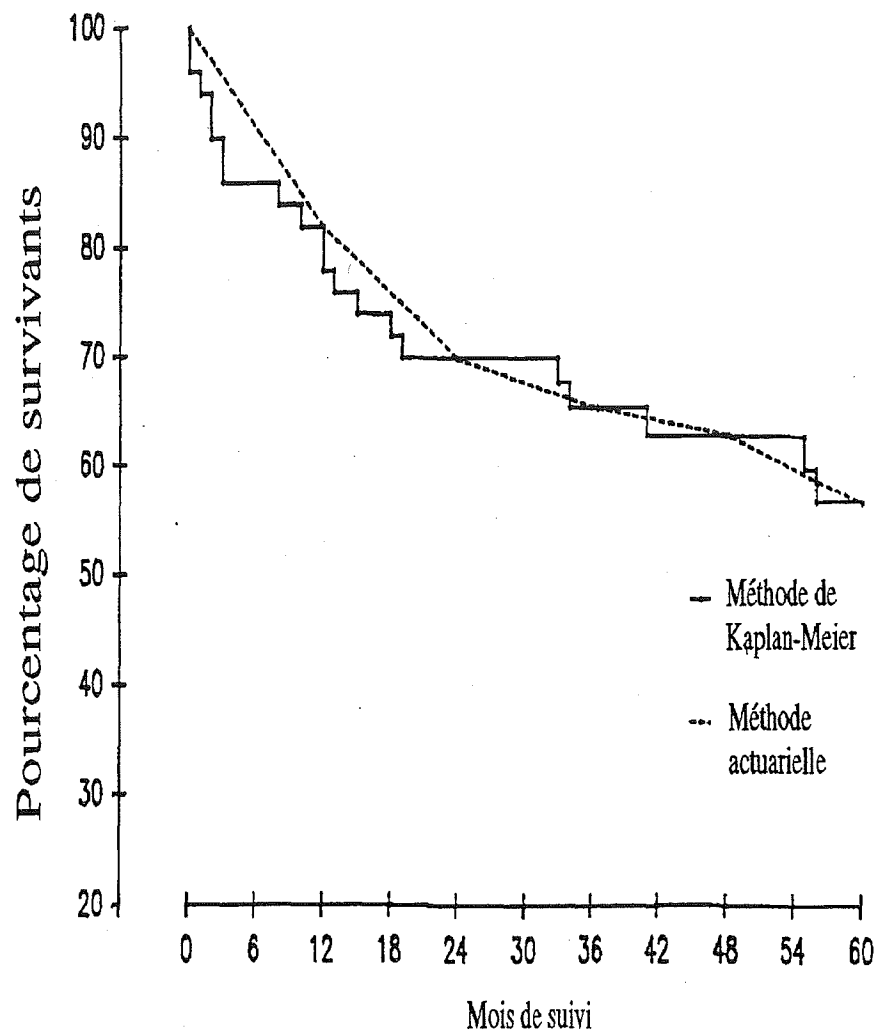


Figure 2. Courbe de survie de Kaplan-Meier pour les patients présentant un mélanome (comparée avec la survie observée, calculée par la méthode actuarielle).

de niveau correspond à un ou des décès, alors que le graphique issu de la méthode actuarielle est une succession de pentes.

Avec cette méthode, on peut calculer un taux corrigé en considérant comme exclus les 3 décès qui ne sont pas imputables au mélanome durant les 5 premières années de survie (signalés par un astérisque dans le tableau 5). Les taux de survie relative se calculent en divisant le taux observé par le taux attendu comme dans la méthode actuarielle.

Tableau 5. Calcul du taux de survie calculé par la méthode de Kaplan-Meier

| Mois après le diagnostic | Nombre de vivants au début du mois | Morts | Exclus | Proportion de morts | Proportion de survivants | Survie cumulée |
|--------------------------|------------------------------------|-------------------|-------------------|---------------------|--------------------------|--------------------|
| (i) | (v _i) | (m _i) | (w _i) | (q _i) | (p _i) | (P _p i) |
| 0 | 50 | 2 | 0 | 0,040 | 0,960 | 0,960 |
| 1 | 48 | 1 | 0 | 0,021 | 0,979 | 0,940 |
| 2 | 47 | 2 | 0 | 0,043 | 0,957 | 0,900 |
| 3 | 45 | 2* | 0 | 0,044 | 0,956 | 0,860 |
| 8 | 43 | 1 | 0 | 0,023 | 0,977 | 0,840 |
| 10 | 42 | 1 | 0 | 0,24 | 0,976 | 0,820 |
| 12 | 41 | 2* | 0 | 0,049 | 0,951 | 0,780 |
| 13 | 39 | 1 | 0 | 0,026 | 0,974 | 0,760 |
| 15 | 38 | 1 | 0 | 0,026 | 0,974 | 0,740 |
| 18 | 37 | 1* | 0 | 0,027 | 0,973 | 0,720 |
| 19 | 36 | 1 | 0 | 0,028 | 0,972 | 0,700 |
| 21 | 35 | 0 | 1 | | | |
| 27 | 34 | 0 | 2 | | | |
| 30 | 32 | 0 | 1 | | | |
| 33 | 31 | 1 | 1 | 0,032 | 0,968 | 0,677 |
| 34 | 29 | 1 | 0 | 0,034 | 0,966 | 0,654 |
| 38 | 28 | 0 | 1 | | | |
| 40 | 27 | 0 | 1 | | | |
| 41 | 26 | 1 | 0 | 0,038 | 0,962 | 0,628 |
| 43 | 25 | 0 | 1 | | | |
| 44 | 24 | 0 | 1 | | | |
| 46 | 23 | 0 | 1 | | | |
| 54 | 22 | 0 | 1 | | | |
| 55 | 21 | 1 | 0 | 0,048 | 0,952 | 0,598 |
| 56 | 20 | 1 | 0 | 0,050 | 0,950 | 0,568 |
| 57 | 19 | 0 | 2 | | | |
| >=60 | 17 | 1* | 16 | | | |

Taux de survie ajusté sur l'âge

En utilisant les taux de survie corrigée ou relative, on prend en partie en compte l'effet de l'âge dans la mesure où on élimine les décès liés à d'autres causes que le cancer dont la fréquence est évidemment liée à l'âge. Cependant, il peut exister une association entre risque de mourir du cancer et âge ; dans ce cas, comme pour les taux d'incidence, soit on limite les comparaisons aux taux de survie spécifique par âge, soit on utilise une standardisation sur l'âge.

Écart type du taux de survie calculé

L'écart type et l'intervalle de confiance sont utilisés pour mesurer la précision des taux de survie, comme pour l'incidence.

Écart type des taux de survie calculé par la méthode directe

$$\text{e.t.}(P) = \sqrt{\frac{P(1-P)}{N}}$$

où : P=Taux de survie; N=Nombre de sujets

Dans le calcul du taux de survie par la méthode directe, le nombre total des patients observés durant 5 ans étant de 30, d'où :

$$\text{e.t.}(P) = \sqrt{\frac{0,57 \times (1-0,57)}{30}} = 0,090$$

L'intervalle de confiance à 95% est :

$$\begin{aligned} &P \pm 1,96 \times \text{e.t.}(P) \\ &= 0,57 \pm 1,96 \times 0,09 \\ &= 0,39 \text{ à } 0,75 \end{aligned}$$

Écart type des taux de survie actuarielle

Ce calcul utilise les 2 dernières colonnes du tableau 2. On obtient la colonne 9 en soustrayant les valeurs de la colonne 3 à celle de la colonne 5. La colonne 10 provient de la division des valeurs de la colonne 6 par celles obtenues dans la 9. La somme des chiffres de la colonne 10 est égale à 0,0177. L'écart type du taux de survie actuarielle est égal à ce taux de survie multiplié par la racine carrée du total des valeurs de la colonne 10, soit :

$$\begin{aligned} &= 0,567 \times \sqrt{0,0177} \\ &= 0,075 \end{aligned}$$

En utilisant les définitions du tableau 2 :

$$\begin{aligned} \text{e.t.}(P) &= P \sqrt{\sum \frac{q_i}{n-m_i}} \\ &= P \sqrt{\sum \frac{m_i}{n(n-m_i)}} \end{aligned}$$

Cette expression est connue sous le nom de formule de Greenwood.

De même, l'intervalle de confiance à 95% pour le taux de survie à 5 ans est :

$$0,567 \pm 1,96 \times 0,075 \text{ donc } 0,42 \text{ à } 0,72$$

En pratique, grâce aux tables publiées par Ederer (1960), on obtient immédiatement une approximation de l'écart type du taux de survie actuarielle.

On observera que l'écart type du taux de survie actuarielle est plus petit que celui calculé par la méthode directe (0,075 par rapport à 0,090). Cette différence montre l'avantage que l'on a, en terme de précision statistique, à utiliser toute l'information, c'est à dire l'information fournie par les patients observés moins de 5 ans.

Pour plus de détails, on se référera à Merrell et Shulman (1955) et Culter et Ederer (1958).

Écart type des taux de survie relative

L'écart type du taux de survie relative s'obtient facilement en divisant l'écart type du taux de survie observée (calculé par la méthode directe ou actuarielle) par le taux de survie attendue. Ainsi, pour le taux de survie à 5 ans, calculé par la méthode actuarielle qui est de 57% avec un taux de survie attendue à 94%, la survie relative est de 61%. L'écart type de ce taux de survie est de :

$$\frac{\text{Ecart type du taux observé}}{\text{Taux de survie attendue}} = \frac{0,075}{0,94} = 0,080$$

L'intervalle de confiance à 95% pour ce taux est de :

$$0,61\% \pm 1,96 \times 0,080 = 0,45 \text{ à } 0,77$$

Comparaison des taux de survie

On souhaite souvent comparer des taux de survie. Si les intervalles de confiance de 2 taux ne se chevauchent pas, on peut être tenté de dire que la différence est statistiquement significative, c'est à dire qu'il est peu probable qu'elle soit le fait du hasard. Cette technique n'est pas vraiment recommandable, nous décrirons donc d'autres procédés plus orthodoxes.

La littérature statistique décrit le test Z qui donne une estimation numérique de la probabilité pour qu'une différence "aussi grande que" ou "plus grande que" celle observée se produise uniquement par hasard. Il se calcule avec la formule :

$$Z = \frac{|P_1 - P_2|}{\sqrt{(e.t.(P_1))^2 + (e.t.(P_2))^2}}$$

où P_1 = Taux de survie du groupe 1

P_2 = Taux de survie du groupe 2

$|P_1 - P_2|$ = Valeur absolue de la différence

C'est à dire l'amplitude de la différence qu'elle soit positive ou négative

e.t.(P_1) = Ecart type de P_1

e.t.(P_2) = Ecart type de P_2

Dans le test Z, si la valeur de $Z > 1,96$, la probabilité que la différence observée soit due au hasard est $< 5\%$ et si $Z > 2,56$, cette probabilité est $< 1\%$.

L'exemple du tableau 6 montre le calcul de la survie observée à 5 ans par la méthode actuarielle pour les 24 hommes ($P_1 = 0,485$) et les 26 femmes ($P_2 = 0,646$). En utilisant la formule de Greenwood, l'erreur standard de P_1 est de 0,105 et celle de P_2 : 0,105.

Tableau 6. Calcul des taux de survie relative observés et du nombre de survivants par années, hommes et femmes

| Année | v_i | m_i | w_i | r_i | q_i | p_i | Pp_i | Décès attendus ($r_i \times Q_i$) ^a |
|---------------|-------|----------|-------|-------|-------|-------|--------|---|
| Hommes | | | | | | | | |
| 0 | 24 | 6 | 0 | 24,0 | 0,250 | 0,750 | 0,750 | 4,32 |
| 1 | 18 | 4 | 0 | 18,0 | 0,222 | 0,778 | 0,584 | 2,66 |
| 2 | 14 | 1 | 2 | 13,0 | 0,077 | 0,923 | 0,539 | 0,82 |
| 3 | 11 | 1 | 2 | 10,0 | 0,100 | 0,900 | 0,485 | 0,39 |
| 4 | 8 | 0 | 1 | 7,5 | 0,000 | 1,000 | 0,485 | 0,74 |
| 5 | 7 | 1 | 1 | 6,5 | 0,154 | | | 0,43 |
| | | <hr/> 13 | | | | | | <hr/> 9,36 |
| Femmes | | | | | | | | |
| 0 | 26 | 3 | 0 | 26,0 | 0,115 | 0,885 | 0,885 | 4,68 |
| 1 | 23 | 2 | 1 | 22,5 | 0,089 | 0,911 | 0,806 | 3,33 |
| 2 | 20 | 1 | 2 | 19,0 | 0,053 | 0,947 | 0,764 | 1,20 |
| 3 | 17 | 0 | 3 | 15,5 | 0,000 | 1,000 | 0,764 | 0,60 |
| 4 | 14 | 2 | 2 | 13,0 | 0,154 | 0,846 | 0,646 | 1,27 |
| 5 | 10 | 0 | 3 | 8,5 | 0,000 | | | 0,57 |
| | | <hr/> 8 | | | | | | <hr/> 11,65 |

^a Q_i est la proportion de la série complète (hommes et femmes) qui sont morts durant l'année (colonne 6 du tableau 2).

$$\text{Aussi } Z = \frac{|0,485 - 0,646|}{\sqrt{0,105^2 + 0,105^2}} = \frac{0,161}{0,148} = 1,09$$

La valeur calculée est inférieure à 1,96. La différence n'est donc pas statistiquement significative au risque 5%. Pour que cette différence soit significative, il aurait été nécessaire d'inclure davantage de sujets dans l'étude afin que les écarts types soient plus petits.

$$\chi^2 = \frac{|13 - 9,36|^2}{9,36} + \frac{|8 - 11,65|^2}{11,65} = 2,56$$

avec $ddl = 1$ $p > 0,10$, différence non significative.

Comme nous l'avons dit plus haut, le test de logrank se trouve dans de nombreux logiciels de statistiques. Il existe d'autres logiciels qui proposent des modèles de survie relative - Hakulinen et Abeywickrama (1985).

Dans la comparaison des survies entre différents groupes de patients, comme dans toute étude épidémiologique, il est souvent nécessaire de contrôler les facteurs de confusion. Par exemple, on peut vouloir étudier la survie d'un groupe de patients traités dans les hôpitaux par rapport à celle de patients traités ailleurs en prenant en compte les différences entre les 2 groupes pouvant influencer sur le pronostic (par exemple l'âge, l'ethnie, le statut social, le stade évolutif de la maladie).

On peut résoudre ce problème en stratifiant l'analyse sur les facteurs de confusion (Mantel, 1966), mais actuellement, on utilise de plus en plus souvent des techniques basées sur le modèle du risque proportionnel (Cox, 1972).

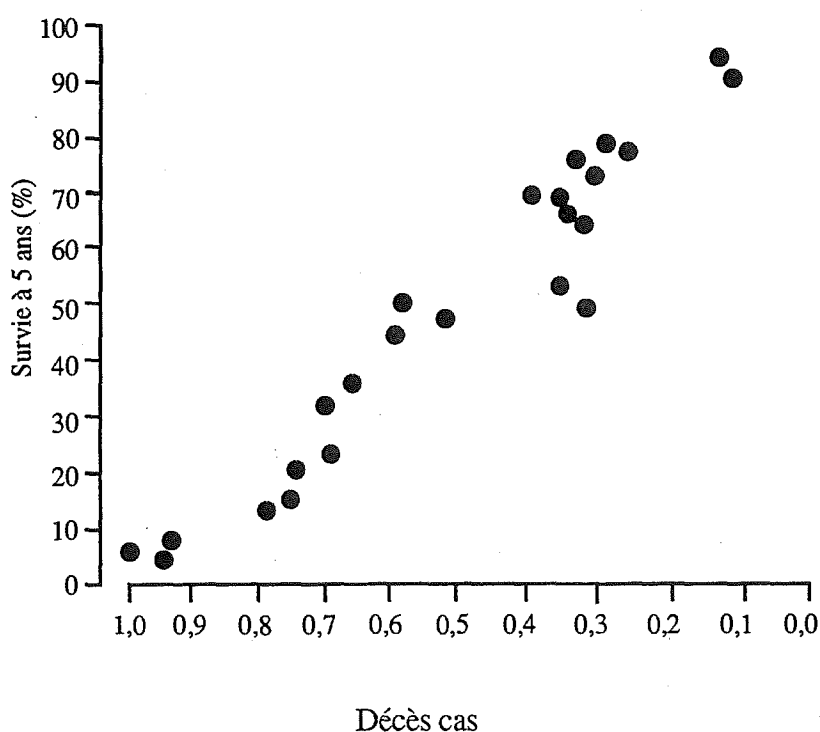


Figure 3. Relation entre les taux de survie relative à 5 ans des cas recensés pendant la période 1973-76 et le rapport décès/cas de 1973 à 77 pour les 24 principales localisations de cancer.
Source : SEER program

De la même façon, les logiciels proposent ces méthodes appliquées à la survie relative (Pocok et coll., 1982; Hakulinen et Tenkanen, 1987). Ce dernier est basé sur GLIM (Baker et Nelder, 1978) et admet aussi des risques non proportionnels¹.

Ratio de létalité

Pour beaucoup de registres, il est impossible de réaliser un suivi correct des cas enregistrés pour calculer des taux de survie. Il est alors possible de présenter à titre indicatif, le ratio de létalité qui est le rapport entre le nombre de nouveaux cas et le nombre de morts pour cette pathologie dans une période de temps donnée. Ce rapport a été utilisé sous le nom de "morts dans la période" (Muir et Waterhouse, 1976) et plus récemment "rapport mortalité/incidence" (Muir et coll., 1987). dans la série *cancer incidence in five continents* pour évaluer l'exhaustivité des registres. Évidemment, les cas incidents et les décès ne concernent pas les mêmes sujets, mais seulement les mêmes diagnostics, et le rapport est uniquement une description indirecte de la survie générale. Cependant, comme le montre la figure 3, la relation entre le taux de survie à 5 ans et le rapport mortalité/incidence pour différents cancers, est très étroite dans un même registre. En revanche, rien ne permet de comparer la survie entre différents registres en utilisant le rapport mortalité/morbidité.

Remerciements

Les auteurs tiennent à remercier le Dr M. Myers du *National cancer institute : Division of cancer Prevention and Control, biometry branch*, pour son aide, ses commentaires et ses suggestions dans la rédaction de ce chapitre.

¹NdT : Depuis 1996 est aussi disponible sur PC le logiciel SURVREL développé par Estève et Hédelin