

# UNE APPROCHE ORIGINALE POUR L'ÉVALUATION DES CAPACITÉS PRONOSTIQUES D'UN MARQUEUR À PARTIR DE COURBES DE SURVIE PUBLIÉES

Etienne Dantan <sup>1</sup>, Christophe Combescure <sup>2</sup>, Marine Lorent <sup>3</sup>, Joanna Ashton-Chess <sup>4</sup>, Pascal Daguin <sup>5</sup>, Jean-Marc Classe <sup>6</sup>, Magali Giral <sup>7</sup> & Yohann Foucher <sup>8</sup>

<sup>1</sup> EA4275 - SPHERE - Biostatistique, Pharmacoépidémiologie et Mesures Subjectives en Santé, Université de Nantes, 1 rue Gaston Veil, 44035 Nantes, France. E-mail : Etienne.Dantan@univ-nantes.fr

<sup>2</sup> CRC & Division of Clinical Epidemiology, Department of Health and Community Medicine, University of Geneva and University Hospitals of Geneva, rue Gabrielle Perret-Gentil 4, 1211 Geneva, Switzerland. E-mail : Christophe.Combescure@hcuge.ch

<sup>3</sup> EA4275 - SPHERE - Biostatistique, Pharmacoépidémiologie et Mesures Subjectives en Santé, Université de Nantes, 1 rue Gaston Veil, 44035 Nantes, France. E-mail : marine.lorent@etu.univ-nantes.fr

<sup>4</sup> TcLand Expression, 21 rue Lanoue Bras de Fer, 44200 Nantes, France. E-mail : ashton\_chess@yahoo.fr

<sup>5</sup> Institut de Transplantation, Urologie et Néphrologie (ITUN), CHU Nantes et INSERM CR1064, 30 bd Jean-Monnet, 44093 Nantes, France. E-mail : pascal.daguin@chu-nantes.fr

<sup>6</sup> Institut de Cancérologie de l'Ouest Centre René Gauducheau, Boulevard Jacques Monod, 44805 Saint-Herblain, France. E-mail : Jean-Marc.Classe@ico.unicancer.fr

<sup>7</sup> Institut de Transplantation, Urologie et Néphrologie (ITUN), INSERM CR1064 et CIC Biothérapie, CHU Nantes, 30 bd Jean-Monnet, 44093 Nantes, France. E-mail : magali.giral@chu-nantes.fr

<sup>8</sup> EA4275 - SPHERE - Biostatistique, Pharmacoépidémiologie et Mesures Subjectives en Santé, Université de Nantes, 1 rue Gaston Veil, 44035 Nantes, France. E-mail : Yohann.Foucher@univ-nantes.fr

**Résumé.** Prédire l'évolution d'une maladie chronique à partir d'un marqueur est un champ de recherche clé en épidémiologie clinique. Mais, les capacités pronostiques d'un marqueur sont rarement évaluées avec une méthodologie adéquate. A partir de courbes de survie publiées, nous proposons de recalculer la sensibilité et la spécificité pour un temps de pronostic ainsi que les valeurs prédictives dépendantes du temps. Nous introduisons de nouveaux indicateurs dépendants du temps comme les rapports de vraisemblance et les rapports de probabilités post-tests. Ce travail a récemment été publié dans le *Journal of Clinical Epidemiology*, une application en ligne est disponible pour le calcul de ces indicateurs (<http://www.divat.fr/en/online-calculators/evalbiom>). La méthodologie proposée est illustrée par le calcul a posteriori de ces indicateurs à partir d'un article publié présentant un marqueur génétique hautement corrélé avec le temps de survenue de métastases distantes chez des femmes atteintes d'un cancer du sein, concluant à la forte capacité pronostique de ce marqueur et présentant les courbes de Kaplan-Meier associées. Des conclusions publiées peuvent être trop optimistes et ainsi induire des décisions thérapeutiques erronées. Notre approche permet aux lecteurs de mieux évaluer les capacités de marqueurs pronostiques. La sensibilité et la spécificité dépendantes du temps informent sur les capacités inhérentes du marqueur pour un temps de pronostic défini. Les valeurs prédictives, les rapports de vraisemblance et les rapports de probabilités post-tests dépendants du temps fournissent des informations supplémentaires pour interpréter un marqueur.

**Mots-clés.** Facteur pronostique, Sensibilité, Spécificité, Valeurs prédictives, Analyse de survie, Rapports de vraisemblance

**Abstract.** Predicting chronic disease evolution from a marker is a key field of research in clinical epidemiology. But, the prognostic capacity of a marker is rarely evaluated using the appropriate methodology. Based on published survival curves, we proposed the use of simple equations to calculate time-dependent sensitivity and specificity and time-dependent predictive values to reappraise prognostic marker accuracy. We also introduced new time-dependent indicators as likelihood ratios, and post-test probability ratios. This work have recently been accepted for publication in the *Journal of Clinical Epidemiology*, the tools necessary to run these direct and simple computations are available online at <http://ww.divat.fr/en/online-calculators/evalbiom>. The methodology is illustrated by back calculating time-dependent indicators from one published article presenting a genetic marker as highly correlated with the time to distant metastasis in patients with breast cancer, concluding on the high prognostic capacity of the marker, and presenting the Kaplan-Meier survival curves. Published conclusions about prognostic marker accuracy may be overoptimistic, thus giving potential for major mistakes in therapeutic decisions. Our approach should help readers better evaluate clinical articles reporting on prognostic markers. Time-dependent sensitivity and specificity inform on the inherent prognostic capacity of a marker for a defined prognostic time. Time-dependent predictive values, likelihood ratios, and post-test probability ratios may additionally contribute to interpret the marker's prognostic capacity.

**Keywords.** Prognostic factor, Sensitivity, Specificity, Predictive values, Survival analysis, Likelihood ratios

## 1 Introduction

Dans beaucoup de champs thérapeutiques, prédire la survenue d'événements de santé est un véritable challenge pour améliorer la prise en charge de patients atteints de maladies chroniques. Un marqueur est souvent présenté comme pronostique s'il est significativement corrélé avec la distribution des temps d'événement. Toutefois, la pertinence clinique devrait porter sur l'aptitude du marqueur à prédire l'évolution future du patient. Un marqueur pronostique doit permettre d'identifier les patients à haut risque et à bas risque de connaître l'événement et ce afin d'adapter leur suivi thérapeutique (Steffen and Steffen, 2013).

Dans la littérature en épidémiologie clinique, une erreur répandue concerne la confusion entre les notions de corrélation et de pronostic (Lachenbruch et al., 2004). La distance entre deux courbes de survie définies par un marqueur binaire, le risque relatif correspondant et la probabilité critique associée sont souvent interprétés comme des indicateurs de capacité pronostique alors qu'ils ne démontrent qu'exclusivement une relation qui ne serait pas due au hasard. Ware (2006) poursuit les travaux de Pepe et al. (2004) et conclut au fait qu'un risque relatif obtenu à partir d'un modèle de survie n'est pas synonyme de capacités pronostiques.

Des approches statistiques comme les courbes ROC dépendantes du temps (Heagerty et al., 2000) et des indices de concordance (Uno et al., 2011) ont été développées pour l'évaluation de capacités pronostiques, mais elles sont peu utilisées dans les publications médicales. Heagerty et al. (2000) a notamment défini la sensibilité et la spécificité dépendantes du temps. Certains auteurs estiment la sensibilité et la spécificité à partir des patients ayant un suivi au moins jusqu'au temps de pronostic, c'est-à-dire en excluant les patients censurés avant le temps de pronostic (Kaplan et al., 2003; ?). Un biais de sélection important est ainsi introduit.

L'évaluation a posteriori des aptitudes d'un marqueur est une solution pour réinterpréter des résultats déjà publiés (Simel et al., 2013). L'objectif de ce travail est d'évaluer a posteriori les capacités pronostiques d'un marqueur à partir de courbes de survie publiées, en calculant la sensibilité, la spécificité, les valeurs prédictives, les rapports de vraisemblance et les rapports de probabilités post-tests pour un temps de pronostic. Nous illustrons ici l'intérêt de notre approche à partir d'un article étudiant un marqueur

généétique comme pronostique de la survenue de métastases chez des femmes atteintes d'un cancer du sein.

## 2 Méthodes

La grande majorité des articles de recherche clinique portant sur un marqueur pronostique montre les courbes de survie de Kaplan-Meier. La figure 1 illustre le cas de deux courbes de survie pour les sujets à Haut Risque (HR) et à Bas Risque (BR) de connaître l'événement étudié et définis à partir d'un marqueur binaire. On note  $NHR$  et  $NBR$  les nombres de sujets à l'inclusion classés respectivement dans les groupe HR et BR.  $SHR(t)$  et  $SBR(t)$  sont les probabilités de survie au temps  $t$  correspondantes. A partir de ces informations, une application disponible en ligne (<http://ww.divat.fr/en/online-calculators/evalbiom>) a été développée pour calculer les indicateurs dépendants du temps présentés dans les sections suivantes.

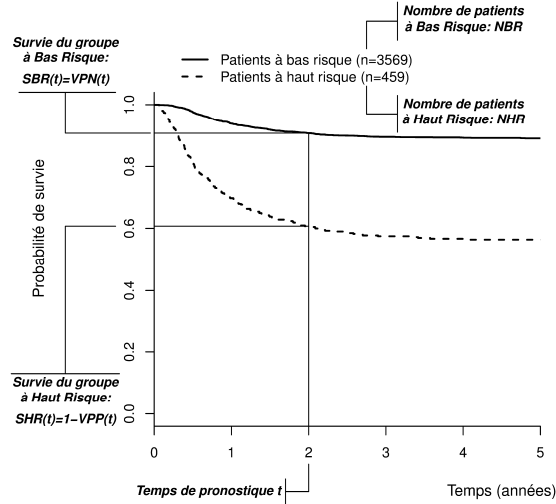


FIGURE 1 – Information disponible sur la plupart des courbes de survie

### 2.1 Sensibilité, spécificité et valeurs prédictives dépendantes du temps

Soit  $t$  un temps de pronostic défini a priori. On note  $D(t)$  la fonction indicatrice dépendante du temps de l'événement, i.e.  $D(t) = 1$  si l'événement d'intérêt survient avant le temps de pronostic  $t$  et  $D(t) = 0$  sinon. D'après Heagerty et al. (2000), la sensibilité au temps  $t$  représente la proportion de patients correctement considérés à haut risque chez les patients subissant l'événement avant  $t$ , i.e.  $Se(t) = P[HR|D(t) = 1]$ . La spécificité au temps  $t$  représente la proportion de patients correctement considérés à bas risque chez les patients ne connaissant pas l'événement avant  $t$ , i.e.  $Sp(t) = P[BR|D(t) = 0]$ . A l'aide du théorème de Bayes, nous démontrons facilement que ces indicateurs peuvent être estimés en fonction des effectifs et des probabilités de survie de chacun de deux groupes (Dantan et al., 2014) :

$$Se(t) = \frac{[1 - SHR(t)] \times NHR}{[1 - SHR(t)] \times NHR + [1 - SBR(t)] \times NBR}$$

$$Sp(t) = \frac{SBR(t) \times NBR}{SHR(t) \times NHR + SBR(t) \times NBR}$$

La sensibilité et la spécificité dépendantes du temps représentent les capacités inhérentes du marqueur. Les valeurs prédictives dépendantes du temps (Heagerty et al., 2000) sont utiles pour l'interprétation clinique et la prise de décision thérapeutique. La valeur prédictive positive au temps  $t$  est la probabilité qu'un patient HR connaisse l'événement avant  $t$ , i.e.  $VPP(t) = P[D(t) = 1|HR] = 1 - SHR(t)$ . La valeur prédictive négative au temps  $t$  est la probabilité qu'un patient BR ne connaisse pas l'événement avant  $t$ , i.e.  $VPN(t) = P[D(t) = 0|BR] = SBR(t)$ . Ces deux indicateurs peuvent donc directement être obtenus à partir des deux courbes de survie publiées. Nous montrons également que les estimateurs des valeurs prédictives sont fonction de la sensibilité et de la spécificité dépendante du temps ainsi que de la probabilité de connaître l'événement (Dantan et al., 2014) :

$$VPP(t) = \frac{Se(t) \times P[D(t) = 1]}{Se(t) \times P[D(t) = 1] + [1 - Sp(t)] \times \{1 - P[D(t) = 1]\}}$$

$$VPN(t) = \frac{Sp(t) \times \{1 - P[D(t) = 1]\}}{Sp(t) \times \{1 - P[D(t) = 1]\} + [1 - Se(t)] \times P[D(t) = 1]}$$

Les conclusions qui en découlent sur les capacités pronostiques d'un marqueur ne peuvent donc pas être directement généralisées dans une population où la probabilité de survenue d'événement est différente de celle dans laquelle le marqueur a été défini initialement. Cette limite est largement acceptée en médecine diagnostique traditionnelle mais souvent ignorée en analyse pronostique.

## 2.2 Rapports de vraisemblance et rapports de probabilités post-tests dépendants du temps

Dans un contexte diagnostique, les rapports de vraisemblance et les rapports de probabilités post-tests sont bien définis. A notre connaissance, leur définition n'a jamais été adaptée au contexte pronostique. Nous définissons les rapports de vraisemblance positif  $RV^+(t)$  et négatif  $RV^-(t)$  pour un temps de pronostic  $t$  :

$$RV^+(t) = \frac{P[HR|D(t) = 1]}{P[HR|D(t) = 0]} = \frac{Se(t)}{1 - Sp(t)}$$

$$RV^-(t) = \frac{P[BR|D(t) = 1]}{P[BR|D(t) = 0]} = \frac{1 - Se(t)}{Sp(t)}$$

Un rapport de vraisemblance proche de 1 indique une règle de classification non informative sur la survenue de l'événement. Plus un rapport de vraisemblance positif est élevé, plus la probabilité d'être considéré à haut risque est associée à la survenue de l'événement avant  $t$ . Plus un rapport de vraisemblance négatif est bas, plus la probabilité d'être considéré à bas risque est associée à l'absence de l'événement avant  $t$ .

Nous introduisons également les rapports de probabilités post-tests positif  $PT^+(t)$  et négatif  $PT^-(t)$  pour un temps de pronostic  $t$  :

$$PT^+(t) = \frac{P[D(t) = 1|HR]}{P[D(t) = 0|HR]} = \frac{P[D(t) = 1]}{P[D(t) = 0]} \times RV^+(t)$$

$$PT^-(t) = \frac{P[D(t) = 1|BR]}{P[D(t) = 0|BR]} = \frac{P[D(t) = 1]}{P[D(t) = 0]} \times RV^-(t)$$

Un patient HR a  $PT^+(t)$  fois plus de risque de subir l'événement avant  $t$  qu'après. Un patient BR a  $1/PT^-(t)$  fois plus de risque de ne pas connaître l'événement avant  $t$  que de le connaître. Ces rapports doivent néanmoins être interprétés avec attention car ils dépendent de la probabilité de connaître l'événement. Les rapports de vraisemblance sont des coefficients de proportionnalité entre les rapports de probabilités pré-tests et post-tests.

### 3 Application au pronostic de survenue de métastases chez des patients atteintes d'un cancer du sein

Dans un article publié par Mook et al. (2010), une signature génétique est présentée comme un facteur pronostique de survenue de métastases après une chirurgie chez des femmes atteintes d'un cancer du sein. Une signature génétique positive caractérise le groupe HR constitué de 439 patientes, le groupe BR est constitué de 525 patientes. Dans leur étude, la probabilité cumulée de voir apparaître des métastases avant 5 ans est de 12% et les probabilités de survie à 5 ans sont significativement différentes entre les deux groupes HR et BR ( $p < 0.001$ ) et égales à 80% et 95% respectivement.

Notre approche nous permet de réévaluer la signature génétique qui semble en réalité relativement sensible mais peu spécifique comme prédicteur de métastases. En effet, nous estimons a posteriori une sensibilité à 77% qui semble raisonnable pour le contrôle du nombre de faux négatifs pour un pronostic à 5 ans, mais au prix d'une faible spécificité de 59%. Pour un pronostic à 5 ans, la VPP et la VPN sont respectivement estimées à 20% et 95%. A partir de cette signature génétique, les patientes considérées à bas risque ont seulement 5% de risque voir apparaître des métastases avant 5 ans. En revanche, la précision d'une signature génétique positive est très faible avec 80% des patientes à haut risque qui ne verront pas apparaître de métastases avant 5 ans.

Ces valeurs prédictives sont bien différentes lorsque l'on imagine utiliser cette signature génétique dans une autre population. Par exemple, en France, la probabilité de voir apparaître des métastases avant 5 ans est de 4.5% dans une population similaire à celle étudiée par Mook et al. (données issues de la cohorte BERENIS, Institut de Cancérologie de l'Ouest, Nantes). Les VPP et VPN à 5 ans sont alors égales à 8% et 98% respectivement. Le risque de considérer une patiente HR à tort est alors augmenté.

Calculés à partir des informations disponibles dans Mook et al. (2010), les rapports de vraisemblance positif et négatif sont respectivement égaux à 1.86 et 0.39 et caractérisent une règle de classification déterminée par la signature génétique qui paraît non informative sur la survenue de métastases. De plus, les rapports de probabilités post-tests positif et négatif à 5 ans étaient respectivement de 0.25 et 0.05. Pour une patiente HR, le risque de survenue de métastases avant 5 ans est moindre que celui de le voir apparaître après. En revanche, pour une patiente BR, le risque de ne pas voir apparaître de métastases avant 5 ans est 20 fois plus élevé que celui d'en avoir avant. La signature génétique étudiée par Mook et al. permet d'identifier précisément les patientes sans métastase à 5 ans post-chirurgie mais ne permet aucunement d'identifier des sujets avec de futures métastases.

### 4 Discussion

Nous proposons de recalculer des indicateurs dépendants du temps tels que la sensibilité, la spécificité, les valeurs prédictives, les rapports de vraisemblance et les rapports de probabilités post-tests à partir de courbes de survie déjà publiées. Calculés a posteriori, ces indicateurs permettent d'évaluer les aptitudes pronostiques d'un marqueur lorsque la méthodologie utilisée n'est pas adéquate. Nous illustrons notre approche à partir d'un article publié portant sur le risque de survenue de métastases distantes chez des femmes atteintes d'un cancer du sein. Ainsi, les résultats et les recommandations publiés peuvent être mis en perspective à l'aide d'informations complémentaires à celles couramment fournies.

La sensibilité et la spécificité à un temps de pronostic donné peuvent éventuellement renforcer les conclusions d'un article. Un marqueur significativement associé au risque de survenue d'un événement peut ne pas réussir à discriminer les patients à haut et à bas risque : la sensibilité peut être acceptable au détriment d'une faible spécificité et vice versa. Un avantage de la sensibilité et de la spécificité dépendantes du temps ainsi que des rapports de vraisemblance dépendants du temps est leur indépendance avec la probabilité de connaître l'événement.

Les valeurs prédictives dépendantes du temps sont plus intéressantes pour une interprétation clinique puisqu'elles correspondent aux probabilités de survenue de l'événement ou non selon le groupe de risque

pour un temps de pronostic. Les rapports de probabilités post-tests dépendants du temps complètent cette évaluation en quantifiant combien de fois le risque de connaître l'événement avant le temps de pronostic est plus élevé que celui de ne pas le connaître. La principale limite de ces deux indicateurs est leur dépendance à la fragilité de la population. Des décisions médicales peuvent donc être prises à tort en utilisant le marqueur dans une population où la probabilité de connaître l'événement est différente de celle dans laquelle le marqueur a été développé.

Dans l'exemple présenté, notre approche montre que le marqueur génétique étudié par Mook et al. (2010) permet une sélection précise des patientes à bas risque avec une valeur prédictive négative égale à 95% pour un pronostic à 5 ans. Les patientes considérées à bas risque à partir du marqueur génétique ont 20 fois plus de risque de ne pas voir apparaître de métastases dans les 5 ans que d'en voir avant. En revanche, une chimiothérapie supplémentaire pourrait être proposée aux patientes à fort risque, mais près de 80% des sujets seraient alors surtraités. Les patientes considérées à haut risque ont 4 fois plus de risque de ne pas voir apparaître de métastases avant 5 ans que d'en voir apparaître avant. Toutefois, dans le contexte du cancer du sein, la question du surtraitement à tort reste en suspens car la survenue de métastases conduit invariablement au décès. Une femme pourrait préférer recevoir un traitement par chimiothérapie à tort plutôt que de ne pas être considérée à haut risque alors qu'elle le devrait.

Les indicateurs dépendants du temps présentés devraient permettre aux lecteurs de mieux évaluer les articles de recherche clinique présentant un marqueur comme pronostic. Bien évidemment, ces conclusions doivent être considérées selon la pathologie étudiée. Par ailleurs, il serait préférable d'utiliser systématiquement une méthodologie correcte pour prouver a priori les capacités pronostiques d'un marqueur plutôt que de les réévaluer a posteriori.

## Références

- Steffen JA and Steffen JS (2013). Driving Forces Behind the Past and Future Emergence of Personalized Medicine. *J Pers Med*, 3(1) : p. 14-22.
- Lachenbruch PA, Rosenberg AS, Bonvini E, Cavaillé-Coll MW and Colvin RB (2004). Biomarkers and surrogate endpoints in renal transplantation : present status and considerations for clinical trial design. *American Journal of Transplantation*, 4(4) : p. 451-7.
- Ware JH (2006). The limitations of risk factors as prognostic tools. *N Engl J Med*, 355(25) : p. 2615-7.
- Pepe MS, Janes H, Longton G, Leisenring W and Newcomb P (2004). Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am J Epidemiol*, 159(9) : p. 882-90.
- Heagerty PJ, Lumley T and Pepe MS (2000) Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics*, 56(2) : p. 337-44.
- Kaplan B, Schold J and Meier-Kriesche HU (2003). Poor predictive value of serum creatinine for renal allograft loss. *American Journal of Transplantation*, 3(12) : p. 1560-5.
- Uno H, Cai T, Pencina MJ, D'Agostino RB and Wei LJ (2011). On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat Med*, 30 :1105-17.
- Simel DL, Easter J and Tomlinson G (2013). Likelihood ratios, sensitivity, and specificity values can be back-calculated when the odds ratios are known. *J Clin Epidemiol*, 66 :458-60.
- Dantan E, Combescure C, Lorent M, Ashton-Chess J, Daguin P, Classe JM, Girard M, Foucher Y (2014). An original approach was used to better evaluate the capacity of a prognostic marker using published survival curves *Journal of Clinical Epidemiology*, sous presse.

Mook S, Knauer M, Bueno-de-Mesquita JM, Retel VP, Wesseling J, Linn SC, et al. (2010). Metastatic potential of T1 breast cancer can be predicted by the 70-gene MammaPrint signature. *Ann Surg Oncol*, 17(5) : p. 1406-13.