

# Cancer du sein

## Contexte des données

Le cancer du sein est la première cause de décès par cancer chez les femmes, en outre, il représente un tiers des nouveaux cas de cancer chez les femmes. Cette maladie se caractérise par une prolifération anarchique de cellules au niveau du tissu mammaire, pouvant se développer à partir des canaux galactophores ou des lobules producteurs de lait. Les facteurs de risque incluent l'âge, les antécédents familiaux, certaines mutations génétiques (notamment BRCA1 et BRCA2), ainsi que des facteurs hormonaux et environnementaux. Le pronostic et les options thérapeutiques varient considérablement selon le stade de diagnostic, allant de la chirurgie conservatrice pour les tumeurs localisées jusqu'à des traitements combinant chirurgie, chimiothérapie, radiothérapie et hormonothérapie pour les cancers plus avancés.

L'atteinte ganglionnaire, c'est-à-dire la présence de cellules cancéreuses dans les ganglions lymphatiques axillaires (*node positive breast cancer*), constitue l'un des facteurs pronostiques les plus importants dans le cancer du sein. Lorsque les cellules tumorales se propagent au-delà de la tumeur primaire vers les ganglions lymphatiques situés sous le bras, cela indique une capacité métastatique accrue et un risque plus élevé de récidive ou de dissémination à distance. Le nombre de ganglions envahis influence directement le stade de la maladie et les décisions thérapeutiques : l'absence d'atteinte ganglionnaire (N0) est associée à un meilleur pronostic, tandis qu'une atteinte extensive (plusieurs ganglions positifs) nécessite généralement des traitements adjuvants plus agressifs. L'évaluation précise du statut ganglionnaire, réalisée par biopsie du ganglion sentinelle ou curage axillaire, permet ainsi de stratifier les patientes selon leur risque et d'adapter les protocoles de traitement pour améliorer les chances de survie sans récidive.

Le *German Breast Cancer Study Group* a conduit un essai clinique randomisé majeur portant sur l'efficacité de l'hormonothérapie chez les patientes atteintes de cancer du sein avec atteinte ganglionnaire. Cette étude, analysée en profondeur par Royston et Altman dans leurs travaux sur la modélisation pronostique (Royston & Altman, 2013), a suivi 686 patientes traitées par hormonothérapie ou non après chirurgie. L'analyse de Royston et Altman a permis de démontrer l'importance de ces facteurs dans la prédiction du pronostic et de développer des modèles statistiques robustes pour l'analyse de survie. Ce jeu de données est depuis devenu une référence incontournable dans la littérature biostatistique, servant de base pour illustrer et valider diverses méthodologies d'analyse de survie, notamment les modèles de régression de Cox et les techniques de sélection de variables pronostiques.

## Descriptif des données

Ce jeu de données comprend 11 colonnes et 686 lignes:

Nom de la colonne	Type de variable	Description
pid	Nombre entier	Identifiant de la patiente
age	Nombre entier	Âge de la patiente en années
meno	Nombre entier	La patiente a-t-elle atteint la ménopause ? (0 = non, elle est pré-ménopause ; 1 = oui, elle est post-ménopause)
size	Nombre entier	Taille de la tumeur en mm
grade	Nombre entier	Grade du cancer base sur une analyse histologique (1 = grade I, 2 = grade II, 3 = grade III, voir explications détaillées plus bas)
nodes	Nombre entier	Nombre de ganglions lymphatiques touchés par la tumeur

Nom de la colonne	Type de variable	Description
pgr	Nombre entier	Concentration en récepteur à la progestérone ( $\text{fmol} \cdot \text{L}^{-1}$ )
er	Nombre entier	Concentration en récepteur aux œstrogènes ( $\text{fmol} \cdot \text{L}^{-1}$ )
hormon	Nombre entier	Traitement par hormonothérapie (0 = non, 1 = oui)
rfstime	Nombre entier	Durée avant récidive, décès ou sortie d'étude (en jours), durée maximum : 7 ans
status	Nombre entier	Statut de la patiente (0 = survie sans récidive, 1 = récidive ou décès)

### Explication de la variable grade

La variable **grade** est un niveau qui décrit le stade histologique de la tumeur. Ils vont de 1 à 3 et s'écrivent en chiffres romains : I, II, et III. Après une étude cellulaire poussée des tissus cancéreux, on calcule un score sur plusieurs critères et on obtient le grade de la tumeur (Bloom & Richardson, 1957). Cela permet de rapidement évaluer la dangerosité (on parle aussi d'agressivité de la tumeur), son risque de récidive, son potentiel métastatique et sa rapidité d'évolution. Une tumeur de grade I aura un développement lent, avec une croissance lente et des tissus qui ressemblent davantage à un tissu mammaire normal. Une tumeur de grade III sera considérée comme se développant plus rapidement, avec des cellules d'aspect très différents aux cellules normales.

Le grade d'une tumeur n'est pas forcément associé à son stade. Car une tumeur peut-être de stade avancé, c'est-à-dire avoir colonisé beaucoup de ganglions lymphatiques, mais avec une évolution lente, et donc un grade faible.

### Exemple de questions

Ce jeu de données est divers, voici des exemples de questions que vous pouvez vous poser. Celles-ci ne sont que des **exemples** et vous pouvez décider de traiter bien d'autres questions.

- Quels sont les liens relatifs entre la taille de la tumeur, son grade, et le nombre de ganglions lymphatiques affectés ?
- L'hormonothérapie permet-elle d'augmenter la probabilité de survie des patientes ?
- Quels sont les effets de l'âge et des concentrations en récepteur des œstrogènes sur la probabilité de survie des patientes ?

### Référence

Référence du jeu de données initial : M. Schumacher, G. Basert, H. Bojar, K. Huebner, M. Olschewski, W. Sauerbrei, C. Schmoor, C. Beyerle, R.L.A. Neumann and H.F. Rauschecker for the German Breast Cancer Study Group (1994). Randomized  $2 \times 2$  trial evaluating hormonal treatment and the duration of chemotherapy in node-positive breast cancer patients. *Journal of Clinical Oncology*, 12, 2086–2093.

Référence de l'article qui a popularisé le jeu de données : Royston, P., Altman, D.G. External validation of a Cox prognostic model: principles and methods. *BMC Med Res Methodol* 13, 33 (2013). <https://doi.org/10.1186/1471-2288-13-33>

Référence sur les grades de tumeurs : Bloom, H. J. G., & Richardson, W. (1957). Histological grading and prognosis in breast cancer: a study of 1409 cases of which 359 have been followed for 15 years. *British journal of cancer*, 11(3), 359.

Lien général sur le cancer du sein et sa prévalence en France : <https://www.santepubliquefrance.fr/maladies-et-traumatismes/cancers/cancer-du-sein>

Lien général sur la compréhension du cancer du sein : <https://www.cancer.fr/personnes-malades/les-cancers/sein/comprendre-les-cancers-du-sein/l-essentiel>