

# Marqueurs sanguins de l'hépatite C

## Contexte des données

La cirrhose est une maladie grave du foie qui est le résultat d'hépatite (par exemple du fait d'infection par les virus des hépatites) ou de consommation excessive d'alcool. La cirrhose endommage irréversiblement le foie, il s'agit d'une inflammation chronique qui endommage les cellules et conduit à leur régénération anarchique, sous forme de nodules. Suite à une cirrhose, le foie diminue voire perd ses fonctions vitales ce qui met en danger la vie des patients. Environ 200 000 personnes sont atteintes de cirrhose en France, dont 30% ont atteint le stade sévère de la maladie. On estime à 10 000 à 15 000 le nombre de décès qui lui sont associés chaque année. Le diagnostic survient en moyenne à l'âge de 50 ans. L'identification de biomarqueurs associés à la cirrhose pourrait permettre un diagnostic plus précoce et donc un traitement plus rapide des patient es.

Le jeu de données fourni ici décrit des patients porteurs du virus de l'hépatite C et des patients sains en considérant les différents stades de la maladie : on observe d'abord une hépatite générale, puis une fibrose du foie, rendant le foie moins souple, avant d'observer une cirrhose, dégradation générale de l'état du foie.

## Descriptif des données

Ce jeu de données contient **615 lignes** et **14 colonnes** dont voici la description

Nom de la colonne	Type de variable	Description
ID	Nombre réel	Identifiant du patient
Category	Chaîne de caractères	Catégorie du patient (0=Blood Donor : donneur de sang, 0s=suspect Blood Donor : potentiel donneur de sang, 1=Hepatitis : patient souffrance d'une hépatite C, 2=Fibrosis : patient souffrant d'une fibrose du foie, 3=Cirrhosis : patient souffrant d'une cirrhose du foie)
Age	Nombre entier	Âge du patient
Sex	Chaîne de caractères	Sexe du patient (m: homme, f : femme)
ALB	Nombre réel	Taux d'albumine dans le sang
ALP	Nombre réel	Taux de phosphatase alcaline
ALT	Nombre réel	Taux d'alanine amino-transférase dans le sang
AST	Nombre réel	Taux d'aspartate amino-transférase dans le sang
BIL	Nombre réel	Taux de bilirubine dans le sang
CHE	Nombre réel	Taux de choline estérase dans le sang
CHOL	Nombre réel	Taux de cholestérol dans le sang
CREA	Nombre réel	Taux de créatinine dans le sang
GGT	Nombre réel	Taux de -glutamyl-transférase dans le sang
PROT	Nombre réel	Taux de protéines totales dans le sang

## Explications variable Category

Les patients sains 0=Blood Donor ont été recrutés parmi les donneurs de sang de l'hôpital d'Hanovre, c'est pour cela qu'on les appellent *blood donors*. Pour les patients 0s=suspect Blood donor, il s'agit de patients

qui devraient être aussi des donneurs de sang, mais dont on n'est pas sûr du statut car il n'a pas été enregistré correctement.

## 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.

- On pourra chercher à comprendre l'influence des variables dans la variance de la population en général, c'est-à-dire à identifier les variables pertinentes qui permettent de séparer des groupes.
- On pourra s'intéresser également aux facteurs qui influent sur la catégorie du donneur.
- On pourra enfin chercher à prédire si le donneur est sain ou pas.

## Référence

Référence générale sur la cirrhose : <https://www.inserm.fr/dossier/cirrhose/>

Hoffmann, G.F., Bietenbeck, A., Lichtinghagen, R., & Klawonn, F. (2018). Using machine learning techniques to generate laboratory diagnostic pathways—a case study. *Journal of Laboratory and Precision Medicine*.