Heva
Une filiale de DOCAPOSTE

Alice Brouquet, Aurélie Schmidt, Eléonore Herquelot
Heva, Lyon, France

EPICLIN 2025
du 14 au 16 mai 2025 - Bordeaux, France

Utilisation des arbres causaux pour l'exploration de l'hétérogénéité des effets

INTRODUCTION

L'utilisation d'un arbre causal permet d'explorer les relations complexes entre un traitement et ses effets cliniques, en identifiant les éventuels effets hétérogènes du traitement suivant les caractéristiques croisées des patients. Cette méthodologie d'analyse sera détaillée sur la comparaison de l'efficacité d'un dispositif de télésurveillance en fonction des caractéristiques des patients dans les données du SNDS.

MÉTHODES

Nous avons construit un arbre causal basé sur les données du SNDS de patients cible.

Les patients **télésurveillés par un dispositif A** (cas) ont été appariés à des patients **télésurveillés par un dispositif B** (témoins) afin de comparer la survenue d'**évènements indésirables**.

L'arbre a été élaboré sur 3 949 patients en prenant en compte les **caractéristiques à la date d'inclusion** : le sexe, l'âge, l'année d'inclusion, les comorbidités, les antécédents de traitements, l'indice de défavorisation sociale et l'historique d'hospitalisations d'intérêt.

L'impact de plusieurs paramètres sur la structure de l'arbre a été étudié, notamment la **taille des feuilles**. La méthode « **honnête** » a également été testée.

L'effet du traitement au sein de chaque feuille a été évalué par le biais d'un **modèle linéaire généralisé** avec une distribution binomiale et un lien logit. L'ensemble des analyses a été **pondérée** en fonction du poids des cas et des témoins.

Définitions

Effet du traitement : Soit i l'individu, T_i l'indicateur de traitement (1 = traité, 0 = contrôle), Y_i l'outcome, X_i les covariables :

- Contrefactuel sous le traitement a par individu : $Y_i(a)$
- Effet du traitement par individu : $Y_i(1) - Y_i(0)$
- Moyenne de l'effet du traitement : $\mathbb{E}[Y_i(1) - Y_i(0)]$
- Moyenne de l'effet conditionnel du traitement : $\tau(x) = \mathbb{E}[Y_i(1) - Y_i(0) | X_i = x]$

	Arbre causal	
	Méthode adaptative	Méthode honnête
Echantillon d'entrainement de l'arbre	Training set	Train Set et Estimation Set (indépendants)
Critère d'ajustement	$EMSE = \mathbb{E}[(Y_i^* - \hat{\tau}(X_i))^2 - Y_i^{*2}]$	
	$\tau(X_i)$ calculé à partir du Train Set	$\tau(X_i)$ calculé à partir de l'Estimation Set

MSE : Erreur quadratique moyenne ; **EMSE :** Erreur quadratique moyenne attendue des effets du traitement

Étapes de construction[1]

1 Initialisation de l'arbre

L'initialisation de l'arbre se fait à l'aide d'un partitionnement récursif. Pour chaque feuille, l'algorithme évalue **toutes les divisions candidates** de cette feuille en minimisant un critère d'ajustement.

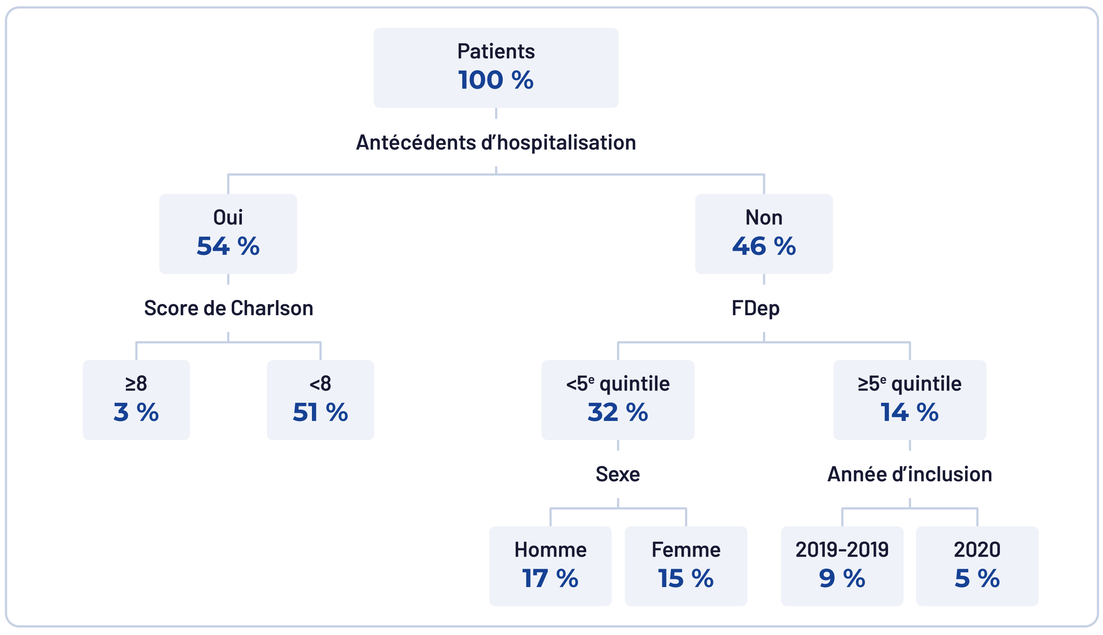
Dans le cas des arbres causaux, ce critère d'ajustement :

- utilise une **variable auxiliaire** de résultat notée Y_i^* dont la valeur attendue pour chaque individu est **l'effet individuel du traitement** :

$$Y_i^* = \frac{Y_i}{D_i \cdot p(X_i) - (1 - D_i) \cdot (1 - p(X_i))}$$

- intègre un terme de **pénalité sur la profondeur de l'arbre**,
- ajoute une **contrainte sur la présence d'individus traités et contrôles** dans les feuilles.

Figure : Arbre causal obtenu à l'initialisation



2 Validation croisée k-fold

Cette méthode de validation est appliquée pour sélectionner un paramètre de complexité utilisé pour **l'élagage**. Un nouvel arbre est construit afin d'obtenir la **profondeur optimale**, c'est-à-dire la profondeur qui minimise le critère d'ajustement.

3 Élagage de l'arbre

L'arbre obtenu à l'entrainement est ensuite élagué à la profondeur obtenue grâce à la validation croisée.

4 Estimation des effets de traitement

Les moyennes conditionnelles sont calculées pour chaque feuille de l'arbre.

5 Mise en avant des effets significatifs

Un modèle linéaire généralisé est appliqué afin d'évaluer l'effet de traitement au sein de chaque feuille.

Mise en pratique sur R

```
formula = as.formula(Y ~ Q5 + sexe + annee + score_Charlson + top_FDiabet_ind + top_comor1 + top_comor2 + top_atcd_trt1 +
                     top_atcd_trt2 + top_atcd_trt3 + top_atcd_trt4 + top_atcd_trt5 + top_sej_1an + classe_age)
```

Calcul de l'arbre causal

Méthode adaptative

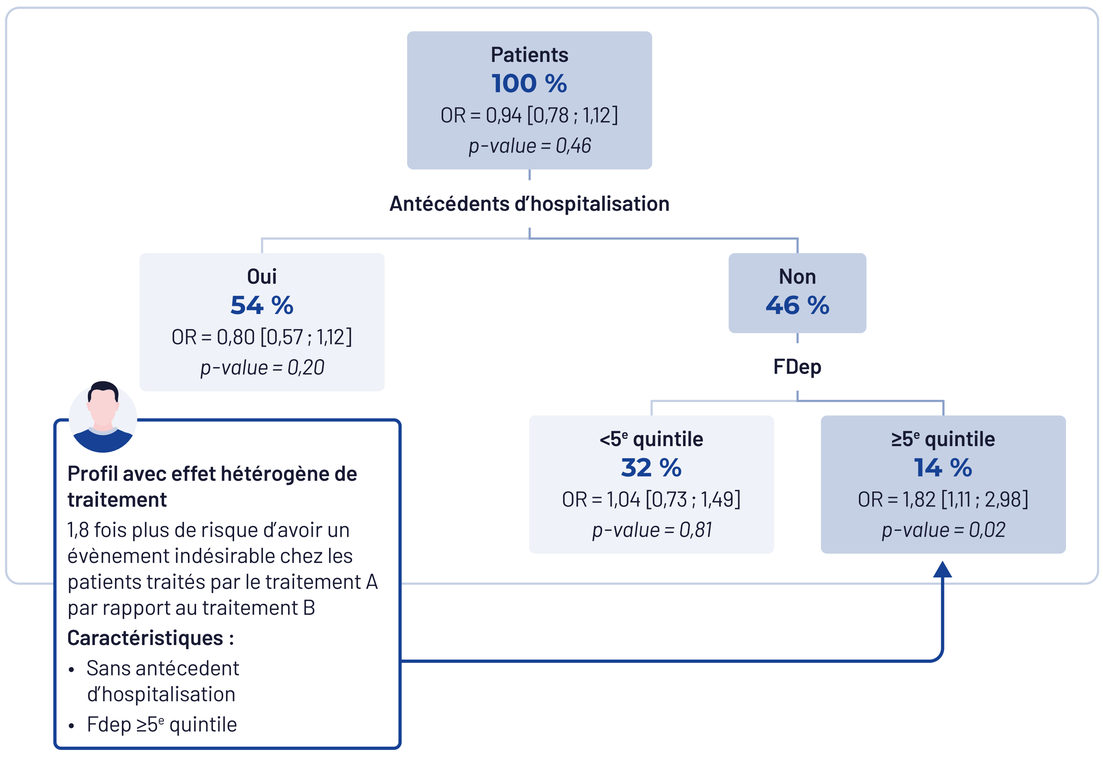
```
tree= causalTree(
    formula,
    data=data.frame(Y=Y, X),
    treatment=T, weights=W,
    #paramètre de division
    split.Rule = "CT", split.Honest = F, split.alpha = 1,
    split.Bucket = T,
    #paramètre de validation croisée
    cv.option = "CT", cv.Honest = F,
    minsize = 80)
```

Méthode honnête

```
tree= honest.causalTree(
    formula,
    #1er échantillon
    data=data.frame(Y=Y1,X1),
    treatment=T1, weights=W1,
    #2e échantillon
    est_data = data.frame(Y=Y2,X2),
    est_treatment= T2, est_weights=W2,
    #paramètre de division
    split.Rule = "CT", split.Honest = T,  split.alpha = 0.5,
    #paramètre de validation croisée
    cv.option = "CT",  cv.Honest = T,  cv.alpha = 0.5,
    split.Bucket = F,
    HonestSampleSize = nrow(est_data),
    minsize = 10)
```

RÉSULTATS

Figure : Arbre causal obtenu par la méthode adaptative



La taille des feuilles avait peu d'impact sur les résultats obtenus, l'algorithme convergeant vers le même arbre à partir d'un certain seuil. Les arbres « honnêtes » présentaient une forte instabilité, lié au faible effectif.

CONCLUSION

Alors que l'arbre de décision classique a pour but de décrire une population en groupes homogènes d'individus, l'arbre causal permet lui de mettre en évidence des groupes d'individus au sein desquels des effets hétérogènes de traitement existent.

	Arbre causal	
	Méthode adaptative	Méthode honnête
Avantages	Fonctionne sur toutes les tailles d'échantillon	2 échantillons indépendants utilisés : Pas de surajustement
Inconvénients	Surajustement : on utilise le même échantillon pour générer l'arbre et calculer les prédictions	Perte de puissance en cas d'échantillon de faible taille : • feuilles plus petites • variance plus grande

Les arbres causaux, sur un effectif limité, ont permis d'explorer des analyses en sous-groupes qui n'auraient pas été possibles manuellement.

Cette méthodologie permet de mettre en évidence des facteurs causaux clés, ouvrant la voie à une meilleure compréhension des mécanismes sous-jacents et à des stratégies thérapeutiques personnalisées.

Références

[1] Athey S, Imbens G. Recursive partitioning for heterogeneous causal effects. Proc Natl Acad Sci U S A. 5 juill 2016;113(27):7353-60.

Remerciements

Les auteurs tiennent à remercier les équipes de la Direction de la stratégie, des études et des statistiques, du Département Accès, Traitement et Analyse de la Donnée et des Demandes Externes de la Caisse nationale de l'assurance maladie (CNAM) pour l'extraction des données.