

Application du score de propension à haute dimension sur les données du SNDS

Évaluation de la mortalité chez les patients atteints de BPCO sous trithérapie

CONTEXTE

Les études observationnelles, malgré l'utilisation de vastes bases de données comme le SNDS, sont exposées aux biais de confusion, notamment d'indication. Le score de propension (PS) est couramment employé pour atténuer ces biais, mais sa performance dépend du choix des covariables, ce qui peut limiter son efficacité.

Le High-Dimensional Propensity Score (HDPS) a été développé pour surmonter cette limite en identifiant et intégrant automatiquement un grand nombre de covariables issues des données brutes, sans sélection a priori. Cette approche optimise le contrôle des biais de confusion et s'avère particulièrement adaptée aux bases médico-administratives comme le SNDS, où l'exploitation de diagnostics, traitements et hospitalisations permet d'améliorer la robustesse des analyses.

OBJECTIFS

Comparer les résultats obtenus avec la méthode du PS à ceux obtenus avec la méthode du HDPS dans l'évaluation de la mortalité des patients BPCO sous triple thérapie en France.

MÉTHODE

Il s'agit d'une étude observationnelle rétrospective basée sur le SNDS. Les patients inclus avaient un diagnostic de BPCO et ont initié une trithérapie en 2015. Deux méthodes d'appariement (PS et HDPS) ont été appliquées pour comparer les patients avec et sans exacerbations.

Le PS a été construit avec des covariables sélectionnées par des cliniciens experts, tandis que le HDPS a été élaboré en plusieurs étapes : identification des dimensions d'intérêt, sélection des covariables dans chaque dimension, estimation du HDPS et son utilisation dans une régression logistique pour effectuer un appariement avec un ratio (1:1). Une analyse a été menée en variant le nombre de covariables incluses dans le HDPS (K=100, 200, 300, 400, 500).

Les analyses de survie ont été réalisées par la méthode de Kaplan-Meier et modèles de Cox.

Construction du HDPS¹

1/ Sélection des hautes dimensions :
Identification de 6 sources de données sur une période définie :

- Antécédents ALD (3 ans)
- Diagnostics d'hospitalisation MCO (3 ans)
- Diagnostics d'hospitalisation HAD (3 ans)
- Visites médicales et paramédicales (3 ans)
- Délivrance de médicaments (ATC niveau 5, 5 ans)
- Délivrance de dispositifs médicaux (3 ans)
- Ajout des covariables présents dans le score de propension (et non présents dans les dimensions retenues)

2/ Identification des Covariables Potentielles

Sélection des 200 covariables les plus prévalentes dans chaque haute dimension.

3/ Analyse de la Récurrence des Covariables

Création de 3 sous-covariables par covariable ($200 \times 3 = 600$ covariables) en fonction de leur fréquence.

4/ Évaluation du Potentiel d'Ajustement

Calcul du biais multiplicatif introduit par chaque covariable via la formule de Bross.

5/ Sélection des covariables finales

Classement des 3600 covariables par biais potentiel et sélection des k meilleures (K=100, 200, 300, 400, 500).

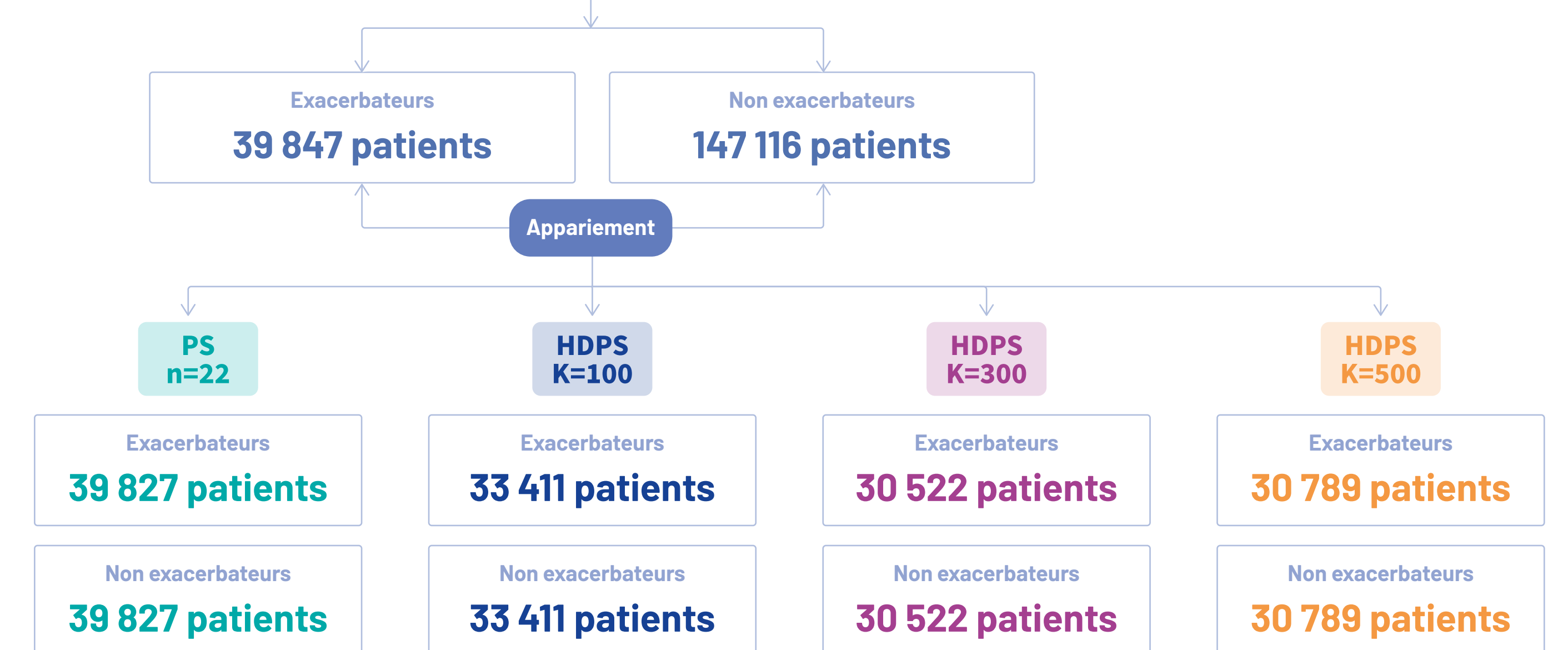
6/ Utilisation du HDPS comme un Score de Propension Classique

Appariement 1:1 (Greedy matching) avec $caliper = 0,2 \times SD(\log[P(A=1 | Xi)])$.

RÉSULTATS

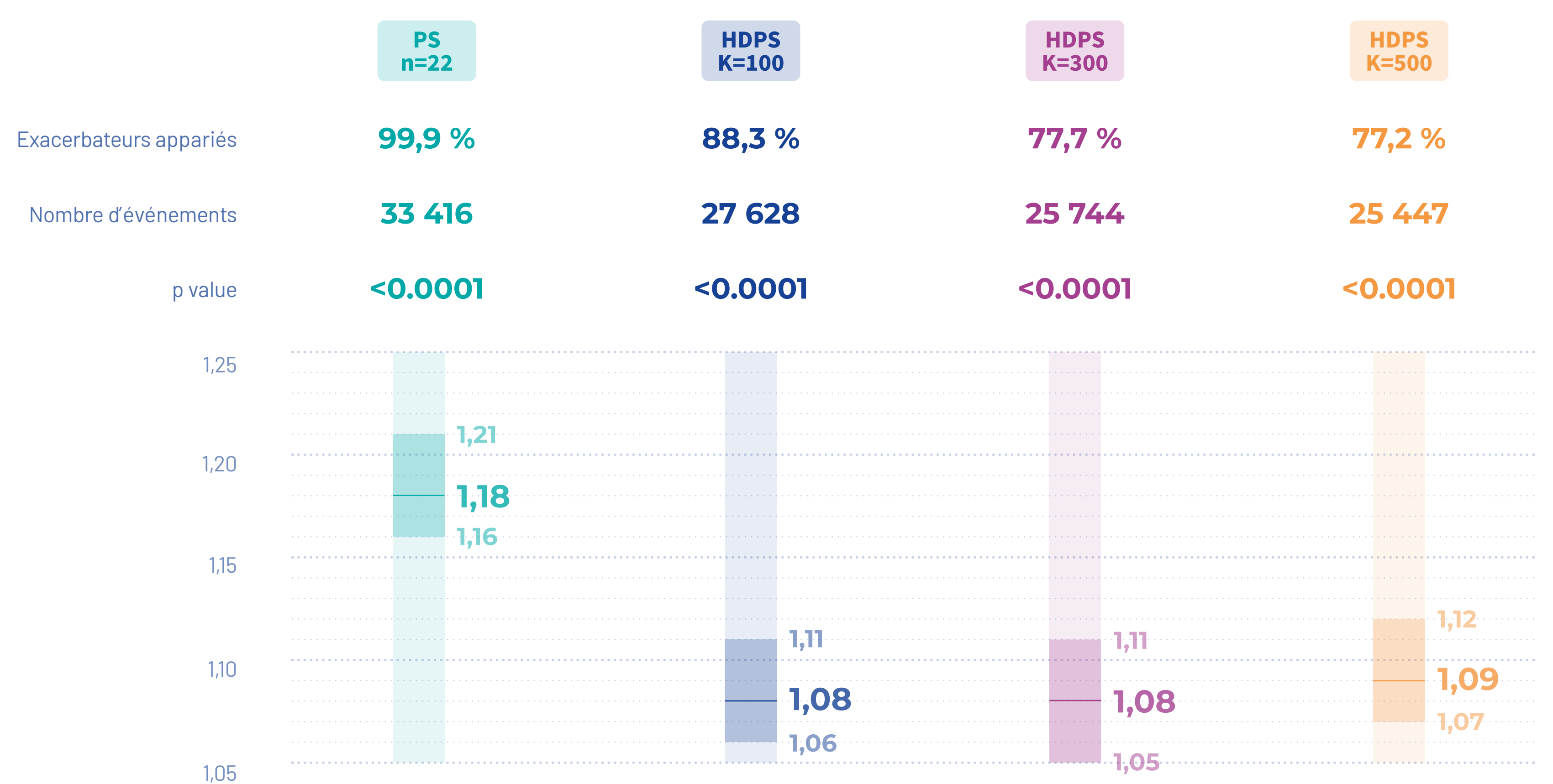
Flowchart en fonction des différents scenarii d'appariement

Population de plus de 40 ans initiant une trithérapie en 2015
186 963 patients



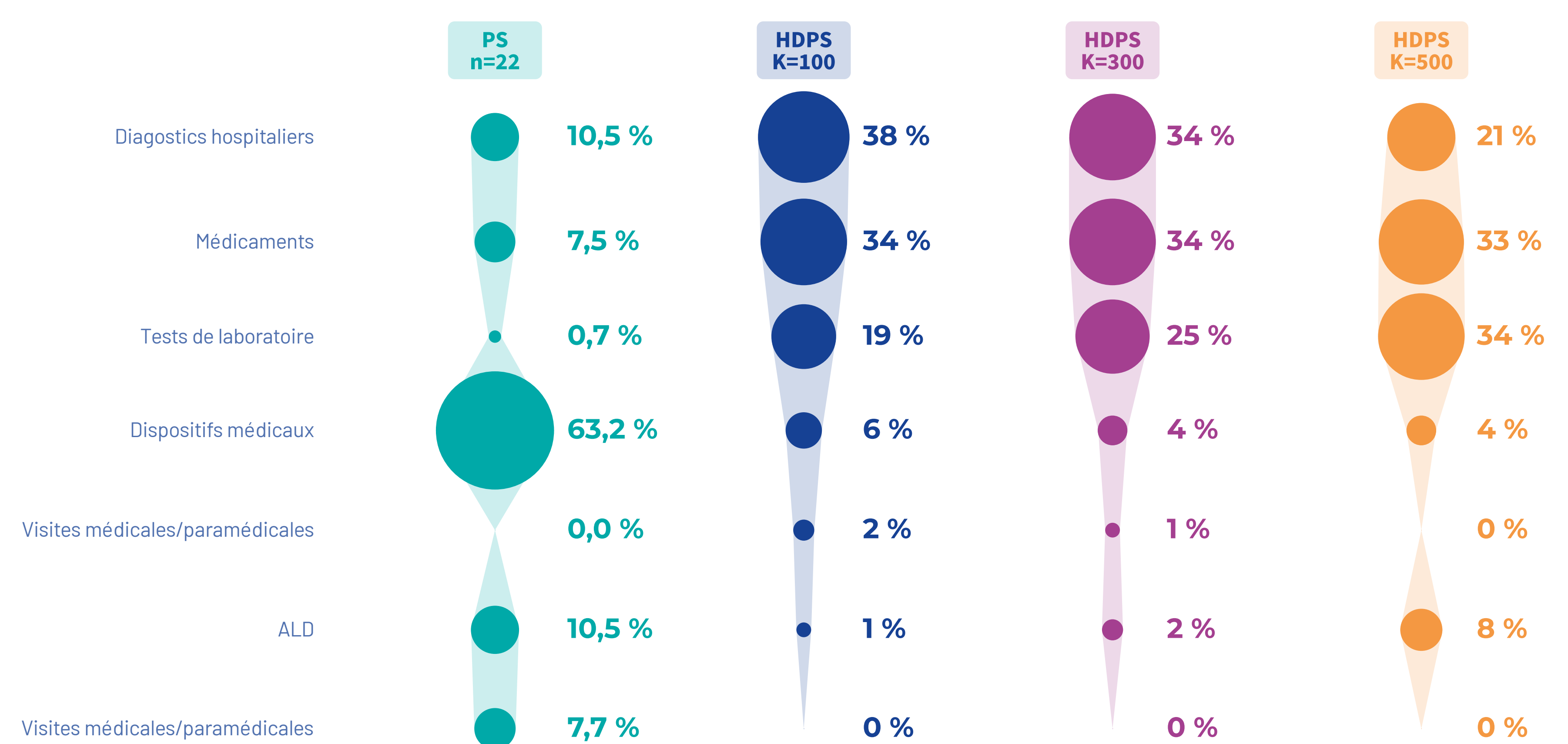
Simulation de l'augmentation du K

Au-delà de 100 covariables, l'ajout de covariables supplémentaires dans le HDPS ne modifie par l'estimation du HR.



Part des covariables sélectionnées

Les dimensions des covariables sélectionnées dans le PS et le HDPS sont très différentes, avec pour le HDPS une part plus importante des diagnostics hospitaliers et des médicaments et une part minoritaire des diagnostics médicaux et des ALD. L'ajout de covariables au HDPS modifie à la marge cette répartition.



CONCLUSION

L'appariement par HDPS a produit des résultats comparables à ceux obtenus par un appariement par PS, malgré une sélection différente des covariables (notamment une importance accrue des tests biologiques). La sélection sans a priori du HDPS a permis un ajustement de qualité sur des covariables non retenues par le comité d'experts, suggérant son utilité pour les analyses de sensibilité et la confirmation des résultats obtenus avec des méthodes classiques.

Références

¹Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology*. 2009 Jul;20(4):512-22.

Abréviations

ALD : Affections de Longue Durée
HDPS : High-Dimensional Propensity Score
PS : Propensity Score
SNDS : Système National des Données de Santé

BPCO : Bronchopneumopathie chronique obstructive
MCO : Médecine Chirurgie Obstétrique
HAD : Hospitalisation à Domicile
ATC : Anatomique, Thérapeutique et Chimique