

Quel apport des méthodes de Machine Learning pour l'estimation des scores de propension
Illustration sur un cas d'étude

INTRODUCTION ET OBJECTIFS

Estimer les effets d'un traitement dans des études observationnelles nécessite de traiter les facteurs de confusion et de modéliser correctement les mécanismes causaux. L'estimation traditionnelle du score de propension repose sur la régression logistique, qui peut être moins performante lorsque l'attribution du traitement dépend d'interactions complexes ou de relations d'ordre supérieur entre les covariables. Les modèles de machine learning (ML) peuvent en principe mieux capturer ces schémas non linéaires, améliorant potentiellement l'équilibre des covariables.

Nous avons utilisé une méthodologie combinant des approches d'inférence causale et un appariement par score de propension (PS) basé sur l'apprentissage automatique afin d'évaluer l'impact de la voie d'administration - Voie A versus Voie B - sur la persistance au traitement.

MÉTHODE

1. Exploration des facteurs de confusion

Deux groupes de traitement (Voie A et Voie B) ont été appariés à l'aide de scores de propension (PS) dérivés de 17 covariables incluant des variables sociodémographiques et des antécédents médicaux.

Relations entre covariables, traitement et persistance -> Corrélations et graphiques acycliques dirigés (DAGs).

2. Calcul des scores de propension (PS)

Benchmark de différents modèles ML pour calculer les PS : Régression logistique (LR), Arbres de décision (DT), Random Forest (RF), Gradient Boosting (GB), AdaBoost (AB), XGBoost (XGB).

-> Estimation la probabilité de recevoir la Voie A compte tenu des covariables.

3. Appariements

L'appariement par score de propension a été réalisé sans remise par l'algorithme des K plus proches voisins, avec un caliper de 0,2 fois la variance du logit du score. L'équilibre des covariables a été évalué à l'aide de la moyenne absolue standardisée (ASAM) et des différences moyennes standardisées (SMD).

-> Modèle ML retenu : minimise ASAM tout en maximisant le taux d'appariement (MR).

4. Étude de l'effet de la voie d'administration sur la persistance

Modèle de Cox : Calcul de l'effet moyen du traitement sur les traités (ATT).

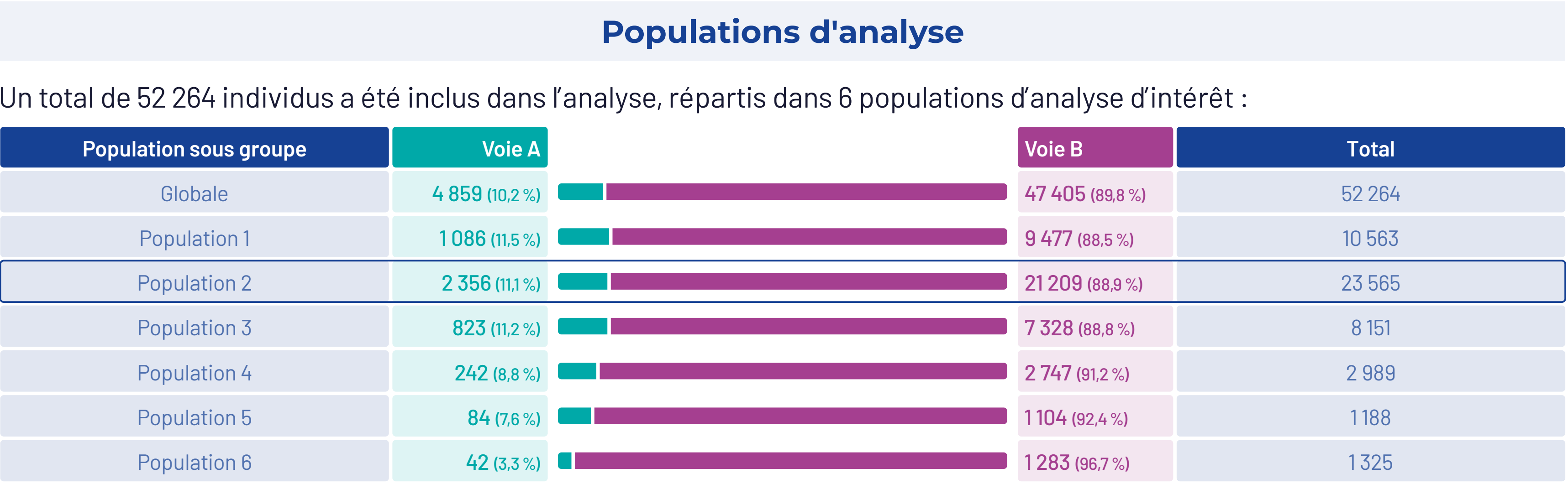
-> Hazard ratio (HR 95 % CI)

CONCLUSION

L'intégration de l'inférence causale avec l'appariement par score de propension basé sur des modèles de ML offre un cadre robuste pour évaluer les effets des traitements dans les données de vie réelle. Dans le contexte de notre analyse, les modèles ML plus complexes avaient tendance à produire des scores de propension moins stables, probablement en raison d'un surapprentissage, d'une prévalence limitée de la voie de traitement A, et d'un chevauchement réduit des scores de propension, ce qui se traduisait par des valeurs ASAM plus élevées et un nombre de paires appariées réduit dans les cohortes de petite taille. La régression logistique a permis d'obtenir un meilleur équilibre des covariables au sein des différentes populations. Ces résultats soulignent la nécessité d'une sélection rigoureuse des modèles et suggèrent que les stratégies hybrides pourraient améliorer les performances et la fiabilité de l'appariement, bien que leur applicabilité reste dépendante du contexte et doit être confirmée par de futures études à plus grande échelle.

Abréviations
AB : AdaBoost ; ASAM : moyenne absolue standardisée ; DAG : graphique acyclique dirigé ; DT : Arbre de décision ; GB : Gradient Boosting ; HR : Hazard ratio ; LR : Régression logistique ; ML : Machine learning ; MR : taux d'appariement ; PS : score de propension ; RF : Random forest ; SMD : différences moyennes standardisées ; XGB : XGBoost

RÉSULTATS

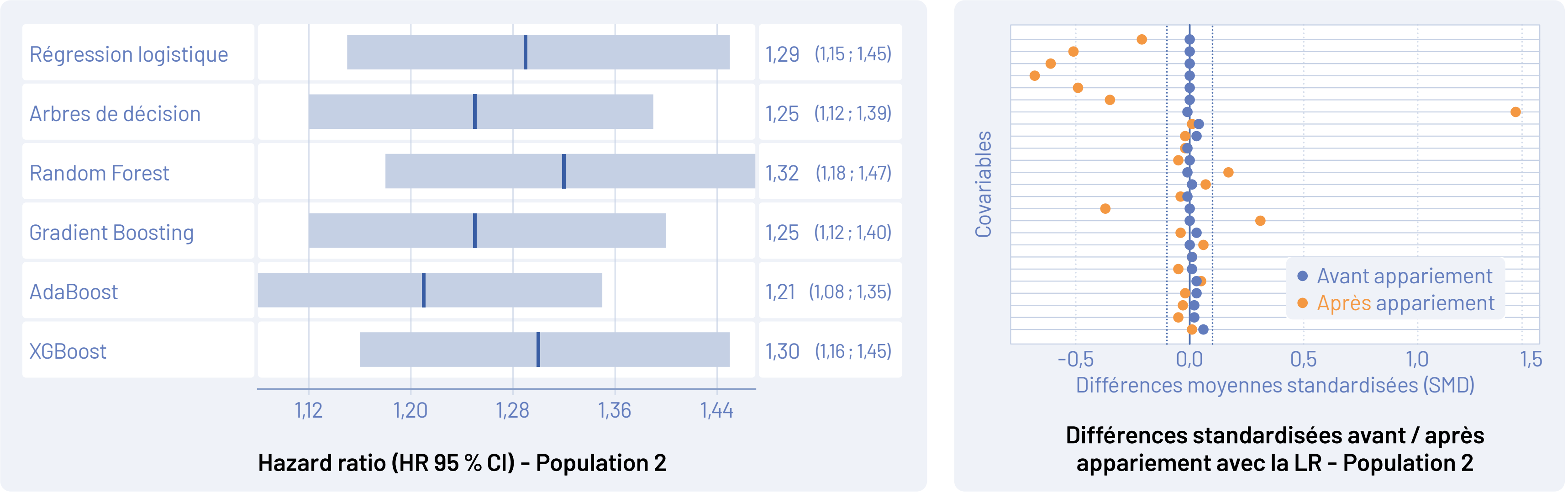
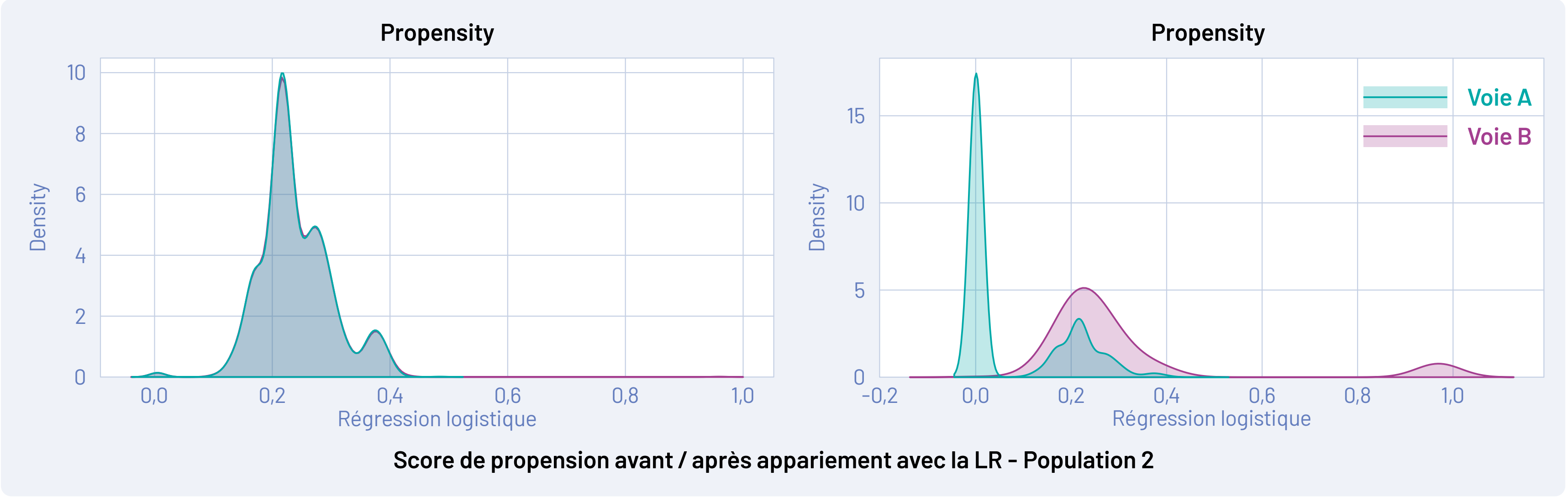


Les populations d'intérêt comportaient entre 1 188 et 23 565 individus, avec un ratio Voie A / Voie B compris entre 3,3 % et 11,0 %.

Suite à l'analyse d'exploration des facteurs de confusion, les 17 variables (caractéristiques sociodémographiques, comorbidités d'intérêt, historique médical ...) initialement considérées ont été conservées pour calculer les scores de propension.

Résultats principaux

Pour la plus grande population (2), la régression logistique (LR) a obtenu le meilleur compromis entre équilibre des covariables (ASAM = 0,01) et le taux d'appariement (MR = 91 %), suivie par le gradient boosting (ASAM = 0,03 ; MR = 0,91)



Les effets du traitement (ATT) sur la persistance étaient cohérents d'un modèle à l'autre, avec une variation de l'HR entre 1,21 (1,08 ; 1,35) et 1,32 (1,18 ; 1,47).

Dans les autres populations (1, 3, 4 et 5), la régression logistique a obtenu le meilleur équilibre des covariables, mais pas systématiquement le meilleur taux d'appariement (MR < 65 % pour les populations plus petites).

Les autres modèles de ML avaient des performances réduites dans les petites populations, avec des valeurs ASAM supérieures à 0,1.

Les arbres de décision (DT) étaient le modèle le moins performant pour toutes les populations, avec systématiquement les ASAM et MR les plus élevés. Les temps de calcul d'estimation des scores de propension étaient similaires pour tous les modèles (max = 1,5 secondes).

Résultats comparatifs par population

Population	Nombre de patients appariés	Taux d'appariement (MR)	Meilleur modèle ML*	ASAM après appariement	HR (IC95 %)	P-value
1	1 004	92 %	AB	0,02	1,39 (1,19 ; 1,62)	<0,001
2	2 141	91 %	LR	0,01	1,29 (1,15 ; 1,45)	<0,001
3	329	40 %	LR	0,03	0,74 (0,57 ; 0,94)	0,020
4	53	22 %	LR	0,07	1,01 (0,59 ; 1,73)	0,970
5	37	44 %	GB	0,10	0,77 (0,32 ; 1,87)	0,570
6	25	60 %	AB	0,07	0,94 (0,42 ; 2,03)	0,840

*Modèle qui minimise l'ASAM. En cas d'égalité, le modèle avec le plus grand MR a été sélectionné

